QUANTILE REGRESSION FOR INTERVAL CENSORED DATA USING AN ENRICHED LAPLACE DISTRIBUTION

Benjamin Deketelaere, Ingrid Van Keilegom

REPRINT | 2026 / 01

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

Quantile Regression for Interval Censored Data using an Enriched Laplace Distribution

Ingrid Van Keilegom[*,†] **and Benjamin Deketelaere**[*]

ORSTAT, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
ISBA/LIDAM, UCLouvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium
e-mail: ingrid.vankeilegom@kuleuven.be, benjamin.deketelaere@uclouvain.be

Abstract: Consider a linear quantile regression model in which the response is subject to interval censoring. Quantile regression is an alternative to the common mean regression, and has the advantage that it allows to give a broader view of the conditional distribution and it is less sensitive to outliers. We propose novel estimators of the quantile regression coefficients for any quantile level $0 < \tau < 1$ by converting the quantile regression problem into a maximum likelihood problem. However, unlike the case of uncensored data, the quantile regression problem cannot be regarded as a maximum likelihood problem with a Laplace distribution for the error term. Instead, the error distribution will be approximated by a generalization of the Laplace distribution employing Laguerre polynomial expansions. This enriched Laplace distributuion will by construction have $\tau-$th quantile equal to zero and can approximate the true underlying error distribution arbitrarily well. We show the consistency of the proposed estimator and investigate the finite sample performance of the method by means of extensive simulations. The estimator is also applied on data regarding the starting salary of newly graduated students.

Keywords and phrases: Bootstrap, consistency, cross-validation, interval censoring, Laguerre polynomials, quantile regression, survival analysis.

1. Introduction

In quantile regression one is interested in evaluating the quantile curves of a response variable given a set of covariates. An advantage of quantile regression over classical mean regression is that it allows to give a broader view of the conditional distribution and to better explore it by studying different quantiles of interest. Another usual praised feature is that it can be used when some common assumptions (e.g. normality, homoscedasticity) of mean regression are not satisfied, and that it is less sensitive to outliers than mean regression.

Interval censoring occurs when a random variable is known only to lie within an interval instead of being observed exactly. This can happen for numerous

arXiv: 2010.00000

[*]Supported by the Global PhD partnership between UCLouvain and KU Leuven.

[†]Supported by the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme, project ASTeRISK (grant No. 40007517).

reasons. For instance in the survey discussed in Section 6 of this paper, participants were asked to report the range of their starting salary and not the exact value. We then only know that the respondent's salary is within the chosen range. Another common example occurs in medical or health studies that entail periodic follow-up. In this case, the event of interest happens in between the last negative and the first positive visit, but the precise moment is not known.

In this paper, we propose a model to perform quantile regression when the response variable is interval censored. Although quantile regression with fully observed outcomes has been extensively studied in the literature, the corresponding literature for interval censored responses is much less rich. Both Kim et al. (2010) and Shen (2013) proposed estimators for the special case of median regression. The estimator from Kim et al. relies on the Missing Information Principle by Efron (1967). No theoretical guarantees of this estimator are shown. Also, the estimator from Kim et al. can only be computed when the covariates take on a finite number of values. Shen's estimator is also based on Efron's Missing Information Principle. Shen proved the consistency of his estimator again when the covariates take on a finite number of values. Shen proposed also a kernel adaptation of his estimator for continuous covariates but no theoretical properties were shown. For general quantile regression models (i.e. not only the median but any quantile of interest), Bottai and Zhang (2010) proposed to use the asymmetric Laplace from Yu and Zhang (2005) as error distribution in the context of right censored data. However, some criticisms of the estimator of Bottai & Zhang can be found in Koenker (2011). Koenker pointed out some limitations of Bottai & Zhang's estimator and explained that the estimator is inconsistent. Zhou, Feng and Du (2017) proposed an estimator that minimizes a slightly modified check function. However, Zhou assumed that the censoring intervals are short enough to alleviate the bias due to interval censoring and that the intervals become smaller when the sample size increases. Zhou's assumption entails that his estimator is reliable on short intervals only. Another consequence of Zhou's assumption is that his estimator cannot be used when some of the interval censored data are in fact right censored. Frumento (2023) described a two-step estimator that can be applied for left, interval and right censored data. His estimator is implemented in the package ctqr in R. However, the second step of Frumento's estimator relies on a piecewise-constant hazard model that is not well justified. Finally, Kreiss and Van Keilegom (2023) studied the consistency of quantile regression estimators that are somewhat similar to ours, but which allow for right censored instead of interval censored responses.

In this paper, we propose a method that can be used for left, right and interval censored data and for every quantile. The starting point of our estimator is the Laplace distribution that was also used by Bottai and Zhang (2010), but we propose to enrich this distribution in a way that the resulting estimator is consistent. We actually propose to work with a generalization of the asymmetric Laplace distribution that involves Laguerre polynomial expansions. The generalized Laplace distribution is more flexible than the classical Laplace and the estimator obtained with this new family of distributions takes care of censored data in a correct way, contrary to the classical Laplace. Moreover, we do not make any assumptions on the size of the intervals.

The paper is organised as follows. Section 2 introduces the model and explains the interval censoring mechanism that will be studied in this paper. Section 3 presents the Enriched Laplace distribution and our estimation procedure. In Section 4 the consistency of the presented estimator is established. Some simulation results are reported in Section 5. Our estimator is compared with the estimators presented in Kim et al. (2010), Zhou, Feng and Du (2017), Frumento (2023) and also with the classical asymmetric Laplace distribution. Section 6 provides an application of our methodology in a real-life context about new graduates who were asked to enter a range for their first salary. Final conclusions and a discussion are given in Section 7.

2. The model

Let T be a (transformed) survival time defined on $(-\infty, \infty)$, using e.g. a logarithmic transformation. Let $\boldsymbol{X} = (X_0, X_1, \ldots, X_p)^t$ be a vector of covariates of length $p+1$. We will assume for notational convenience that $X_0 = 1$, and that the covariance matrix of $(X_1, \ldots, X_p)^t$ is positive definite. We allow T to be subject to interval censoring of type K, where $K = 1, 2, 3, \ldots$ is random. This means that for given $K = k$, there are k inspection times $Y_{k,1} < \ldots < Y_{k,k}$ that form an array of random variables $\boldsymbol{Y} = \{Y_{k,j} : k = 1, 2, \ldots; j = 1, \ldots, k\}$. We assume throughout that $(K, \boldsymbol{Y})$ and T are independent given $\boldsymbol{X}$ and that $E(K) < \infty$. On the event $\{K = k\}$, let L and R denote the endpoints of the random interval among $(-\infty, Y_{k,1}], (Y_{k,1}, Y_{k,2}], \ldots, (Y_{k,k}, \infty)$ which contains the true survival time T, i.e.

$$
\begin{aligned}
(L, R) = (-\infty, Y_{k,1})\mathbb{1}_{T \le Y_{k,1}} &+ \sum_{j=2}^{k} (Y_{k,j-1}, Y_{k,j})\mathbb{1}_{Y_{k,j-1} < T \le Y_{k,j}} \\
&+ (Y_{k,k}, +\infty)\mathbb{1}_{T > Y_{k,k}}.
\end{aligned}
$$

This means that the observable random vector is $(\boldsymbol{X}, L, R)$ where $L < T \le R$. Finally, let $\{(\boldsymbol{X}_i, L_i, R_i) : i = 1, \ldots, n\}$ be a sample of independent and identically distributed random vectors having the same distribution as $(\boldsymbol{X}, L, R)$.

This model has also been considered by Schick and Yu (2000) in the absence of covariates, among others. It can be viewed as a mixture of various case k models. A common case could be the one where $K - 1$ follows a Poisson distribution, in which case the inter-follow-up times are independent and have a common exponential distribution. The variables $Y_{k,j}$ could e.g. correspond to medical follow-up times, and the event is in that case known to occur after the last visit where the event did not take place and before the first one where it did happen.

Our main goal is to estimate the τ-th conditional quantile of the survival time T given the covariates $\boldsymbol{X}$ for a given quantile level $0 < \tau < 1$, which we suppose to be linear:

$$
T = \boldsymbol{X}^t \boldsymbol{\beta}(\tau) + \sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X})\epsilon(\tau). \tag{1}
$$

Here, $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \ldots, \beta_p(\tau))^t$ is the vector of regression coefficients, we assume that $\epsilon(\tau)$ is independent of $\boldsymbol{X}$, and the τ-th quantile of $\epsilon(\tau)$ equals 0, i.e. $P(\epsilon(\tau) \leq 0) = \tau$. We allow for heteroscedasticity in the model via the scale factor $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X})$, which can be any parametric positive function depending on the vector $\boldsymbol{\gamma}(\tau)$ of length $p+1$. A natural example is $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \exp(\boldsymbol{X}^t \boldsymbol{\gamma}(\tau))$. Note that the τ-th quantile of T given $\boldsymbol{X}$ equals $q_\tau(\boldsymbol{X}) = \boldsymbol{X}^t \boldsymbol{\beta}(\tau) = \beta_0(\tau) + \beta_1(\tau) X_1 + \ldots + \beta_p(\tau) X_p$. We also like to highlight that our model only requires linearity of the τ-th quantile, and not of other quantiles. Some other papers on quantile regression, especially in the literature on right censored data, require that linearity holds for all quantile levels $0 < \bar{\tau} \leq \tau$. Throughout the paper we assume that conditionally on $\boldsymbol{X}$, the variables T, L and R are continuous, and that the vector $\boldsymbol{X}$ is also continuous. While continuity is required for T, we assume continuity of L, R and $\boldsymbol{X}$ for convenience, but this is not necessary for the method to work.

We will use the following notations. The distribution of $\boldsymbol{X}$, $\epsilon(\tau)$ and of T given $\boldsymbol{X} = \boldsymbol{x}$ are denoted by $F_{\boldsymbol{X}}$, $F_{\epsilon(\tau)}$ and $F(\cdot|\boldsymbol{x})$. Lowercase letters are used for the corresponding densities. Then under model (1), we have that

$$f(t|\boldsymbol{x}) = \frac{1}{\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x})} f_{\epsilon(\tau)}\Big(\frac{t - \boldsymbol{x}^t \boldsymbol{\beta}(\tau)}{\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x})}\Big). \tag{2}$$

The fact that $L < T \leq R$ implies that the range of (L, R, T) is $\{(\ell, r, t) : -\infty \leq \ell < t \leq r \leq \infty\}$. Then, for a subject with observed interval $(\ell, r]$ and covariate vector $\boldsymbol{x}$ we have that

$$\begin{aligned} &P(L \in d\ell, R \in dr, \ell < T \leq r, \boldsymbol{X} \in d\boldsymbol{x}) \\ &= f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})[F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})] f_{\boldsymbol{X}}(\boldsymbol{x}), \end{aligned} \tag{3}$$

where $d\ell$ is a short-hand notation for an infinitesimal small interval $[\ell, \ell + d\ell)$ and similarly for dr and $d\boldsymbol{x}$, and

$$f_{L,R|X}(\ell, r|x) = \frac{\partial^2}{\partial \ell \partial r} P(L \leq \ell, R \leq r | L < T \leq R, \boldsymbol{X} = \boldsymbol{x}).$$

Throughout the paper we will work under the so-called non-informative censoring assumption, which says that for all $t \in \mathbb{R}$ the conditional density of T given L, R and $\boldsymbol{X}$ equals

$$f_{T|L,R,\boldsymbol{X}}(t|\ell, r, \boldsymbol{x}) = \frac{f(t|\boldsymbol{x})}{P(\ell < T \leq r | \boldsymbol{X} = \boldsymbol{x})} \mathbb{1}_{\ell < t \leq r}, \tag{4}$$

and we will also assume that the parameters in model (1) are not present in the model for $f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})$ and $f_{\boldsymbol{X}}(\boldsymbol{x})$. Condition (4) means that once L and R are observed, the distribution of T no longer depends on these bounds. These two assumptions will be crucial to replace the so-called full likelihood, which will be based on (3), by the simplified likelihood, which will be solely based on $F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})$. We will use the latter to estimate the distribution F, as we will explain in the next section.

3. The estimation procedure

3.1. The uncensored case

In the uncensored case, Koenker and Bassett (1978) proposed to estimate $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(T_i - \boldsymbol{X}_i^t \boldsymbol{\beta}(\tau)), \tag{5}$$

where ρ_τ is the quantile check function given by $\rho_\tau(a) = a(\tau - \mathbb{1}(a < 0))$.

The estimator $\hat{\boldsymbol{\beta}}(\tau)$ admits a representation in the form of *maximum likelihood estimation* (MLE). Indeed, if T follows an *Asymmetric Laplace Distribution* (ALD), its conditional density function is

$$f_{\boldsymbol{\eta},\tau}(y|\boldsymbol{x}) = \frac{\tau(1-\tau)}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \exp\Big[- \rho_\tau\Big(\frac{y - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big], \tag{6}$$

where $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})^t$, and $0 < \tau < 1$, $\boldsymbol{x}^t\boldsymbol{\beta}$ and $\sigma(\boldsymbol{\gamma}, \boldsymbol{x})$ are respectively a skewness, location and scale parameter. Note also that for $\tau = 0.5$, the density is symmetric and the ALD reduces to the "double exponential" Laplace distribution. If T has an ALD distribution, which we denote by $T|\boldsymbol{X} = \boldsymbol{x} \sim \text{ALD}(\boldsymbol{x}^t\boldsymbol{\beta}, \sigma(\boldsymbol{\gamma}, \boldsymbol{x}), \tau)$, then $P(T \le \boldsymbol{x}^t\boldsymbol{\beta}|\boldsymbol{X} = \boldsymbol{x}) = \tau$, i.e. $\boldsymbol{x}^t\boldsymbol{\beta}$ is the conditional τ-th quantile of T. This property allows us to express the estimator $\hat{\boldsymbol{\beta}}(\tau)$ defined in (5) as a maximum likelihood estimator. Indeed, it can be easily seen that if $(T_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$, is an i.i.d. sample of size n having the same distribution as $(T, \boldsymbol{X})$, then for any $\boldsymbol{\gamma}$ fixed,

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \prod_{i=1}^{n} f_{\boldsymbol{\eta},\tau}(T_i|\boldsymbol{X}_i). \tag{7}$$

Hence, the maximization of the ALD likelihood leads to a correct quantile regression estimator independently of what the true underlying distribution of T is. In that sense the ALD distribution is just a working model, which does not need to be correct to do valid inference. More details on the equivalence between the maximization of the ALD likelihood and the minimization of the check function criterion can e.g. be found in Koenker and Machado (1999). We also like to refer to the book by Koenker (2005) and the review paper by Koenker (2017) for more details on the methodology of quantile regression, and to Kotz, Kozubowski and Podgorski (2012) for a detailed account of the properties of the ALD distribution.

3.2. The Enriched Laplace family

The possibility to express the quantile regression estimation problem as a MLE problem in the uncensored case is very attractive, and makes it tempting to also use MLE in the case where data are censored. In their paper, Bottai and Zhang (2010) adapted the estimator defined in (7) to the case of right censored responses by replacing the density $f_{\boldsymbol{\eta},\tau}(T_i|\boldsymbol{X}_i)$ by the corresponding survival function whenever T_i is right censored. Although this simple adaptation of the quantile estimator seems natural, it does not lead to consistent or even correct estimators, as was explained by Koenker (2011), who also gave some simulation evidence of the inconsistency for right censored data of the estimator of Bottai and Zhang (2010). The intuitive reason for that is that the mass of a censored observation is redistributed to the right in a way determined by the survival function, and so if the true distribution is not an ALD, the ALD likelihood does not redistribute this mass in a correct way. The ALD likelihood leads to consistent estimators if the true underlying distribution is ALD, whereas in the uncensored case this model assumption is not required. Although the paper by Bottai and Zhang (2010) and the reaction by Koenker (2011) were limited to the case of right censored data, the same applies to the case of interval censoring, where the maximization of the likelihood (appropriately adapted to the case of interval censoring) based on the ALD distribution would not lead to correct estimators of the quantile function.

We propose to alleviate this inconsistency problem by replacing the ALD by a generalization based on Laguerre polynomial expansions. By doing this, we will create a flexible and rich family of distributions that we call the *Enriched Laplace family*. This family has also been used by Kreiss and Van Keilegom (2023) in the context of quantile regression with right censored data, and we will construct it keeping two crucial requirements in mind:

(1) The τ-th quantile of any distribution in this new family should by construction be equal to $\boldsymbol{x}^t\boldsymbol{\beta}$, like for the ALD
(2) The new family should be sufficiently rich in the sense that any continuous density could be approximated arbitrarily well by a member of this family.

The idea is to start from the asymmetric Laplace density given in (6), and to rewrite it as

$$f_{\boldsymbol{\eta},\tau}(y|\boldsymbol{x}) = \frac{1}{\sigma(\boldsymbol{\gamma}(\tau),\boldsymbol{x})} f_{\epsilon,\tau}\Big(\frac{y-\boldsymbol{x}^t\boldsymbol{\beta}(\tau)}{\sigma(\boldsymbol{\gamma}(\tau),\boldsymbol{x})}\Big), \tag{8}$$

where

$$f_{\epsilon,\tau}(t) = \tau(1-\tau)\Big(e^{-t\tau}\mathbb{1}_{t\geq 0} + e^{-t(\tau-1)}\mathbb{1}_{t<0}\Big),$$

and so $\epsilon \sim \mathrm{ALD}(0,1,\tau)$. We now add Laguerre polynomials to make the model more flexible. For natural numbers k and m and for $t \geq 0$, let

$$v_k(t) = \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^j}{j!} t^j, \quad P_{m,\boldsymbol{\theta}}(t) = \sum_{k=0}^{m} \theta_k v_k(t),$$

be a Laguerre polynomial of order k and a linear combination of Laguerre polynomials depending on a vector $\boldsymbol{\theta} = (\theta_0,\ldots,\theta_m)^t$ of length $m+1$. Then, we define the Enriched Laplace density by

$$f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t) = \tau(1-\tau)\Big(e^{-t\tau}P^2_{m,\boldsymbol{\theta}}(t\tau)\mathbb{1}_{t\geq 0} + e^{-t(\tau-1)}P^2_{\tilde{m},\tilde{\boldsymbol{\theta}}}(t(\tau-1))\mathbb{1}_{t<0}\Big), \tag{9}$$

where $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ have Euclidean norm equal to one, $\theta_0, \tilde{\theta}_0 > 0$, and $m, \tilde{m} \geq 0$ are natural numbers. In the case where $m = \tilde{m} = 0$ this density reduces to the asymmetric Laplace density, whereas for strictly positive values of m and $\tilde{m}$, the extra parameters in the model enhance its flexibility.

Note that it can be easily seen (see Appendix A.1 for the proof) that $f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(\cdot)$ is a density function, thanks to the orthonormality of Laguerre polynomials with respect to the exponential density, i.e. $\int_0^\infty e^{-t} v_k(t) v_{k'}(t) dt = \mathbb{1}_{k=k'}$ (see Szego (1939)). We write $\epsilon \sim \mathrm{ELD}(0, 1, \tau, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ if the density of ϵ is given by (9). For the corresponding survival time T we write

$$T|\boldsymbol{X} = \boldsymbol{x} \sim \mathrm{ELD}(\boldsymbol{x}^t \boldsymbol{\beta}, \sigma(\boldsymbol{\gamma}, \boldsymbol{x}), \tau, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}),$$

and its conditional density is denoted by

$$f_{\boldsymbol{\alpha},\tau}(y|\boldsymbol{x}) = \frac{1}{\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x})} f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}\Big(\frac{y - \boldsymbol{x}^t \boldsymbol{\beta}(\tau)}{\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x})}\Big), \tag{10}$$

with $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})^t$.

Laguerre polynomials have been used in many other contexts, where the goal is each time to construct a flexible yet simple family of functions satisfying certain properties. For example, Florens, Simar and Van Keilegom (2020) used Laguerre expansions to estimate the boundary of a random variable, Kreiss and Van Keilegom (2022) estimated incubation and generation times based on Laguerre polynomials, and Geerdens, Claeskens and Janssen (2013) constructed a new class of frailty models that extend the Gamma frailty model by using Laguerre polynomials. An important property of Laguerre polynomials is their ability to approximate any continuous function in the following sense: if φ is a continuous function defined on the positive real line, satisfying $\int_0^\infty \varphi^2(t) e^{-t} dt < \infty$, then

$$\lim_{m\to\infty} \int_0^\infty \big[\varphi(t) - P_{m,\boldsymbol{\theta}}(t)\big]^2 e^{-t} dt = 0, \quad \text{with} \quad \boldsymbol{\theta} = \Big(\int_0^\infty \varphi(t) v_k(t) e^{-t} dt\Big)_{k=0}^m$$

(see Nikiforov and Uvarov (1988) for more details).

The following proposition shows that the two crucial properties that we required for our new family of distributions are satisfied for the ELD family. The first one follows easily from the orthonormality of Laguerre polynomials, whereas the second one is based on the above key property of Laguerre polynomials. The proof can be found in Appendix A.3.

Proposition 3.1. *Suppose that* $(T, \boldsymbol{X})$ *satisfies* $T = \boldsymbol{X}^t \boldsymbol{\beta} + \sigma(\boldsymbol{\gamma}, \boldsymbol{X}) \epsilon(\tau)$ *with* $\epsilon(\tau)$ *and* $\boldsymbol{X}$ *independent.*

(i) Suppose that $\epsilon(\tau) \sim ELD(0, 1, \tau, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$. *Then,* $P(\epsilon(\tau) \leq 0) = \tau$, *i.e.* $\boldsymbol{x}^t \boldsymbol{\beta}$ *is the* τ*-th quantile of* T *conditional on* $\boldsymbol{X} = \boldsymbol{x}$.

(ii) Suppose that $\epsilon(\tau)$ *has an arbitrary continuous density satisfying* $P(\epsilon(\tau) \leq 0) = \tau$, *and let* A *be a subset of* $\mathbb{R}$ *for which* $\inf_{t\in A} f_{\epsilon(\tau)}(t) > 0$. *Then,*

$$\lim_{m,\tilde{m}\to\infty} \int_A \Big(\sqrt{f_{\epsilon(\tau)}(t)} - \sqrt{f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t)}\Big)^2 dt = 0,$$

where

$$\boldsymbol{\theta} = \Big(\frac{1}{\sqrt{\tau(1-\tau)}} \int_0^\infty \sqrt{f_{\epsilon(\tau)}\Big(\frac{t}{\tau}\Big) e^{-t}}\, v_k(t)\, dt \Big)_{k=0}^{m},$$

and

$$\tilde{\boldsymbol{\theta}} = \Big(\frac{1}{\sqrt{\tau(1-\tau)}} \int_0^\infty \sqrt{f_{\epsilon(\tau)}\Big(-\frac{t}{1-\tau}\Big) e^{-t}}\, v_k(t)\, dt \Big)_{k=0}^{\tilde{m}}.$$

Part (ii) of Proposition 3.1 shows that if the density of T given $\boldsymbol{X} = \boldsymbol{x}$ is continuous, it can be approximated by a member of the ELD family in the sense that the Hellinger distance (restricted to the set A) between the true density and the density of the ELD family goes to zero if the number of parameters in the ELD family tends to infinity.

3.3. Maximum likelihood estimation

We now return to our original problem, which was the estimation of the conditional quantiles of T in the presence of interval censoring. As already mentioned earlier, we will do this by using MLE, but instead of naively using the ALD, we will use the ELD, which should allow us, thanks to Proposition 3.1, to estimate the quantiles in a consistent way, contrary to those based on the ALD. However, we did not discuss yet the construction of the likelihood in the presence of interval censoring.

For the construction of the likelihood we start from the density of $(L, R, \boldsymbol{X})$ given in (3), which yields the following full likelihood:

$$L_{\text{full}} = \prod_{i=1}^{n} f_{L,R|\boldsymbol{X}}(L_i, R_i|\boldsymbol{X}_i) \big[F(R_i|\boldsymbol{X}_i) - F(L_i|\boldsymbol{X}_i)\big] f_{\boldsymbol{X}}(\boldsymbol{X}_i)$$

for an independent sample $(L_1, R_1, \boldsymbol{X}_i), \dots, (L_n, R_n, \boldsymbol{X}_n)$ from the same distribution as $(L, R, \boldsymbol{X})$. However, it has been shown by Oller, Gomez and Calle (2004) that under the non-informative censoring assumption given in (4), the maximization of the full likelihood is equivalent to the maximization of the following expression, called the simplified likelihood:

$$L_{\text{simplified}} = \prod_{i=1}^{n} \big[F(R_i|\boldsymbol{X}_i) - F(L_i|\boldsymbol{X}_i)\big],$$

which we will use to estimate the distribution of F. Oller, Gomez and Calle (2004) showed in their Proposition 1 that the non-informative censoring assumption implies the so-called constant-sum condition. We do not give the definition of the latter condition for reasons of brevity, but it is useful to note that the latter condition is sufficient for the equivalence of the full and the simplified likelihood (see Theorem 2 in Oller, Gomez and Calle (2004)).

If we now maximize the simplified likelihood over the ELD family, we can define our maximum likelihood estimators as

$$\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\tilde{\boldsymbol{\theta}}})^t = \underset{\boldsymbol{\alpha} \in A}{\operatorname{argmax}}\, L(\boldsymbol{\alpha}|\tau), \tag{11}$$

where

$$L(\boldsymbol{\alpha}|\tau) = \prod_{i=1}^{n} \big[F_{\boldsymbol{\alpha},\tau}(R_i|\boldsymbol{X}_i) - F_{\boldsymbol{\alpha},\tau}(L_i|\boldsymbol{X}_i)\big], \tag{12}$$

with $F_{\boldsymbol{\alpha},\tau}$ the distribution corresponding to the Enriched Laplace density $f_{\boldsymbol{\alpha},\tau}$ and $A = \{(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta} \in \boldsymbol{S}_m, \tilde{\boldsymbol{\theta}} \in \boldsymbol{S}_{\tilde{m}}\}$, where B and Γ are compact subsets of $\mathbb{R}^{p+1}$ and $\boldsymbol{S}_m$ and $\boldsymbol{S}_{\tilde{m}}$ are m and $\tilde{m}$-dimensional semi-spheres of radius 1. To compute this likelihood, a closed-form expression of the distribution $F_{\boldsymbol{\alpha},\tau}$ can be used, which is available in Appendix A.2.

A final remark is regarding the estimation of the scale function $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x})$. Since we use MLE, we do not only estimate the coefficients $\boldsymbol{\beta}$, but also all other parameters in the model including $\boldsymbol{\gamma}(\tau)$. Since the τ-th quantile of $\epsilon(\tau)$ is equal to zero, but the scale of $\epsilon(\tau)$ is not fixed, this might lead to identification issues. For example, when $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{x}) = \exp(\boldsymbol{x}^t\boldsymbol{\gamma}(\tau))$, the intercept γ_0 and the scale of $\epsilon(\tau)$ might not be identifiable. However, this is not an issue if we are only interested in estimating the τ-th quantile $\boldsymbol{x}^t\boldsymbol{\beta}$, which is usually the case.

4. Asymptotic theory

4.1. Notations and definitions

In the previous section m and $\tilde{m}$ were considered fixed, but in order to show the consistency of the proposed estimators, we will work with increasing sequences $(m_n)_{n\in\mathbb{N}}$ and $(\tilde{m}_n)_{n\in\mathbb{N}}$ that tend to infinity as n tends to infinity. Let $\mathcal{F}_{m_n,\tilde{m}_n}$ be the sieve-space of densities defined by

$$\mathcal{F}_{m_n,\tilde{m}_n} = \Big\{(y, \boldsymbol{x}) \mapsto f_{\boldsymbol{\alpha},\tau}(y|\boldsymbol{x}) : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta} \in \boldsymbol{S}_{m_n}, \tilde{\boldsymbol{\theta}} \in \boldsymbol{S}_{\tilde{m}_n}\Big\}, \tag{13}$$

where $f_{\boldsymbol{\alpha},\tau}$ is given in (10). The following notations are introduced:

$$h(\boldsymbol{x}, \ell, r) = f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})\,[F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})]\,f_{\boldsymbol{X}}(\boldsymbol{x}), \tag{14}$$

$$p_+(t) = \sqrt{\frac{f_{\epsilon(\tau)}\big(\frac{t}{\tau}\big)}{\tau(1-\tau)}e^t}, \qquad p_-(t) = \sqrt{\frac{f_{\epsilon(\tau)}\big(-\frac{t}{1-\tau}\big)}{\tau(1-\tau)}e^t},$$

$$p_{+,n}(t) = \sum_{k=0}^{m_n} \theta_k v_k(t), \qquad p_{-,n}(t) = \sum_{k=0}^{\tilde{m}_n} \tilde{\theta}_k v_k(t).$$

Then, similarly as in Kreiss and Van Keilegom (2023), any function $f_n \in \mathcal{F}_{m_n,\tilde{m}_n}$ can be written as

$$\begin{aligned} f_n(y|\boldsymbol{x}) =& \frac{1-\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_\tau\Big(\frac{y - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big) p_{+,n}^2\Big(\tau\frac{y - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big) \\ &+ \frac{\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_{1-\tau}\Big(-\frac{y - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big) p_{-,n}^2\Big(-(1-\tau)\frac{y - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big), \end{aligned} \tag{15}$$

where $f_\tau(t) = \tau \exp(-\tau t) \mathbb{1}(t > 0)$. We can decompose f in a similar way:

$$f(y|\boldsymbol{x}) = \frac{1-\tau}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})} f_\tau\Big(\frac{y-\boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})}\Big) p_+^2\Big(\tau\frac{y-\boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})}\Big) + \frac{\tau}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})} f_{1-\tau}\Big(-\frac{y-\boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})}\Big) p_-^2\Big(-(1-\tau)\frac{y-\boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma},\boldsymbol{x})}\Big). \tag{16}$$

For $f_n \in \mathcal{F}_{m_n,\tilde{m}_n}$ we denote the corresponding distribution function by F_n. Similarly to (14), we define

$$h_n(\boldsymbol{x},\ell,r|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) = f_{L,R|\boldsymbol{X}}(\ell,r|\boldsymbol{x}) \left[F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x})\right] f_{\boldsymbol{X}}(\boldsymbol{x}). \tag{17}$$

In addition to the spaces $\mathcal{F}_{m_n,\tilde{m}_n}$ the following sequence of spaces is defined:

$$\mathcal{F}_n = \Big\{(\boldsymbol{x},\ell,r) \mapsto h_n(\boldsymbol{x},\ell,r|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta} \in \boldsymbol{S}_{m_n}, \tilde{\boldsymbol{\theta}} \in \boldsymbol{S}_{\tilde{m}_n}\Big\}. \tag{18}$$

We are now ready to define

$$L_n(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) = \prod_{i=1}^n h_n(\boldsymbol{X}_i, L_i, R_i|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}). \tag{19}$$

Then, the estimator $\hat{\boldsymbol{\alpha}}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\gamma}}_n, \hat{\boldsymbol{\theta}}_n, \hat{\tilde{\boldsymbol{\theta}}}_n)$ that maximizes $L(\boldsymbol{\alpha}|\tau)$ in (12) over the space $A_n = \{(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) : \boldsymbol{\beta} \in B, \boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta} \in \boldsymbol{S}_{m_n}, \tilde{\boldsymbol{\theta}} \in \boldsymbol{S}_{\tilde{m}_n}\}$, also maximizes the likelihood $L_n(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta},\tilde{\boldsymbol{\theta}})$ in (19) over A_n. Finally, let $\hat{h}_n(\boldsymbol{x},\ell,r) = h_n(\boldsymbol{x},\ell,r|\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\gamma}}_n, \hat{\boldsymbol{\theta}}_n, \hat{\tilde{\boldsymbol{\theta}}}_n)$.

4.2. Consistency of the quantile regression estimator

To show the consistency of our estimator, we define the following for $\alpha \neq 0$ and for densities p_1 and p_2:

$$\mathcal{H}_\alpha(p_1,p_2) = \frac{1}{\alpha}\int_{\mathbb{R}^{p+1}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_1(\boldsymbol{x},l,r)\Big[\Big(\frac{p_1(\boldsymbol{x},l,r)}{p_2(\boldsymbol{x},l,r)}\Big)^\alpha - 1\Big] dl\, dr\, d\boldsymbol{x}. \tag{20}$$

Note that $\mathcal{H}_{-1/2}$ equals twice the squared Hellinger distance, which we denote by H^2:

$$H^2(p_1,p_2) = \int_{\mathbb{R}^{p+1}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(\sqrt{p_1(\boldsymbol{x},l,r)} - \sqrt{p_2(\boldsymbol{x},l,r)}\Big)^2 dl\, dr\, d\boldsymbol{x}$$

(see the proof in Appendix B.1). For $\alpha \in (0,1]$, let

$$\delta_n(\alpha) = \inf_{h_n \in \mathcal{F}_n} \mathcal{H}_\alpha(h,h_n).$$

We are now ready to state the first asymptotic result, which says that the Hellinger distance between $\hat{h}_n$ and the true h is $o_P(r_n)$ for some $r_n \to 0$. The proof is given in Appendix B.4, and relies on an adaptation of Theorem 4 in Wong and Shen (1995), which we give in Appendix B.1, and which was also used in Kreiss and Van Keilegom (2023).

Theorem 4.1. *Suppose that*

1. *There is an $\alpha \in (0, 1/2)$ such that $\delta_n(\alpha) \to 0$ as n tends to infinity.*
2. *There exists a sequence $\varphi_n = o(n)$ such that*

$$\frac{m_n + \tilde{m}_n}{\varphi_n} \log\left(\frac{n(m_n + \tilde{m}_n)}{\varphi_n}\right) = O(1). \tag{21}$$

Then,

$$H(h, \hat{h}_n) = o_P(r_n), \quad \textit{with} \quad r_n^2 = \max\left(\frac{\varphi_n}{n}, \delta_n(\alpha)\right).$$

This result will allow us to show the consistency of the quantile regression estimator. But before stating that result, let us first look closer into the two conditions required in Theorem 4.1.

Remark 4.1. The two conditions in Theorem 4.1 ask for some discussion.

1. Lemma B.1 in Appendix B.1 gives an upper bound for $\mathcal{H}_\alpha(h_n, h)$, from which it follows that $\delta_n(\alpha) \to 0$ provided that

$$\sup_t |p_{+,n}(t) - p_+(t)| \to 0 \quad \text{and} \quad \sup_t |p_{-,n}(t) - p_-(t)| \to 0.$$

   Theorem 1 in Chapter II.8 in Nikiforov and Uvarov (1988) shows that, under a weak condition on the tails of the distribution of $\epsilon(\tau)$, $p_{+,n}$ and $p_{-,n}$ converge uniformly on compact intervals to p_+ and p_-. A more complete discussion on this is given in Kreiss and Van Keilegom (2023).
2. The sequence φ_n depends on the speed at which m_n and $\tilde{m}_n$ tend to infinity. An example of sequences φ_n, m_n and $\tilde{m}_n$ for which (21) holds, is $\varphi_n = n/\log n$ and $m_n = \tilde{m}_n = \varphi_n/\log n$. Indeed, the left hand side of (21) then equals

$$\frac{1}{\log n} \log\left(\frac{n}{\log n}\right),$$

   and this is $O(1)$.

We conclude this section with a corollary regarding the consistency of $\hat{\boldsymbol{\beta}}_n$. The proof is given in Appendix B.5, and requires the following assumptions, where $\mathcal{X}$ is the support of $\boldsymbol{X}$ and $q(\tau|\boldsymbol{x}) = \inf\{t : F(t|\boldsymbol{x}) \geq \tau\}$ is the true τ-th quantile of T given $\boldsymbol{X} = \boldsymbol{x}$, for any $0 < \tau < 1$.

(A1) The density $f_{\boldsymbol{X}}$ is continuous on its support $\mathcal{X}$, and $\mathcal{X}$ is compact.

(A2) There exists a $\xi > 0$ such that

$$\max\left\{\inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{|q(\tau|\boldsymbol{x})-r|<\xi} f_{R|\boldsymbol{X}}(r|\boldsymbol{x}), \inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{|q(\tau|\boldsymbol{x})-\ell|<\xi} f_{L|\boldsymbol{X}}(\ell|\boldsymbol{x})\right\} > 0,$$

and there exists a $\delta > 0$ such that $\inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{|q(\tau|\boldsymbol{x})-t|<\delta} f(t|\boldsymbol{x}) > 0$.

(A3) The support of K is $\{1, 2, \ldots, k_{\max}\}$ for some $k_{\max} < \infty$.

(A4) There exists a constant $c_L > 0$ such that for all $k = 1, 2, \ldots, k_{\max}$ and for all $j = 1, \ldots, k$, $P(\inf_{\boldsymbol{x}\in\mathcal{X}}(F(Y_{k,j}|\boldsymbol{x}) - F(Y_{k,j-1}|\boldsymbol{x})) \geq c_L) \to 1$ as n tends to infinity (where $Y_{k,0} = -\infty$ for all k).

Note that (A1) is a standard assumption and implies that $\inf_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x}) > 0$. If $f_{R|\boldsymbol{X}}(r|\boldsymbol{x})$ is uniformly continuous in $(r, \boldsymbol{x})$ and $f_{R|\boldsymbol{X}}(q(\tau|\boldsymbol{x})|\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{X}$, then $\inf_{\boldsymbol{x}\in\mathcal{X}} f_{R|\boldsymbol{X}}(r|\boldsymbol{x}) > 0$ for all r in a neighborhood of $q(\tau|\boldsymbol{x})$, and similarly for $f_{L|\boldsymbol{X}}(\ell|\boldsymbol{x})$. Hence, assumption (A2) is only slightly stronger than assuming that $f_{R|\boldsymbol{X}}(q(\tau|\boldsymbol{x})|\boldsymbol{x}) + f_{L|\boldsymbol{X}}(q(\tau|\boldsymbol{x})|\boldsymbol{x}) > 0$, which was also imposed in e.g. Geskus and Groeneboom (1999) (condition (D1) p. 634).

Assumptions (A3) and (A4) are related to the variable K. Consider the common case where $Y_{k,j} = \xi_0 + \sum_{\ell=1}^{j} \xi_\ell$ and $k = \sup\{k \geq 1 : \sum_{\ell=1}^{k} \xi_\ell \leq d\}$, where d is the length of the study, ξ_0 is a fixed time point (usually chosen sufficiently far in the left tail of F), ξ_ℓ ($\ell \geq 1$) is the time between the $(\ell-1)$-th and the ℓ-th visit, and assume that all ξ_ℓ are i.i.d. and have support $[a, b]$ for some $0 < a < b < \infty$. Then, k will be at most equal to d/a and so (A3) is satisfied. Also, $F(Y_{k,j}|\boldsymbol{x}) - F(Y_{k,j-1}|\boldsymbol{x}) = f(\zeta|\boldsymbol{x})(Y_{k,j} - Y_{k,j-1})$ for some $Y_{k,j-1} < \zeta < Y_{k,j}$, and hence (A4) is satisfied provided $c_L = \inf_{\boldsymbol{x}\in\mathcal{X}} \min\{F(\xi_0|\boldsymbol{x}), a \inf_{\xi_0 \leq y \leq \xi_0 + d} f(y|\boldsymbol{x})\} > 0$.

Corollary 4.1. *Suppose that (A1)-(A4) hold and suppose the conditions of Theorem 4.1 hold. Then,*

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \xrightarrow{P} 0. \tag{22}$$

5. Simulations

In this section the performance of our estimator is evaluated by means of Monte Carlo simulations. We start by describing in Subsection 5.1 the practical implementation of the method and we introduce a cross-validation procedure that allows to select the order of the Laguerre polynomials. Next, in Subsection 5.2 we define the models that will be used in the simulations. Our estimator is compared with other estimators in the literature in Subsection 5.3, while Subsection 5.4 presents the performance of bootstrap confidence intervals. Finally, in Subsection 5.5 we report the results for a heteroscedastic model.

5.1. Implementation and selection of polynomial orders

For practical reasons related to the implementation in R, the constraints $\|\boldsymbol{\theta}\| = 1$ and $\theta_0 > 0$ are replaced by the constraint $\theta_0 = 1$ in the simulations; the same applies to $\tilde{\boldsymbol{\theta}}$. Since there is a one-to-one relation between the open half-sphere $\boldsymbol{S}^m$ and the hyperplane $\{(\theta_0, \ldots, \theta_m) \in \mathbb{R}^{m+1} : \theta_0 = 1\}$, the identifiability of our model is maintained. We also impose $m = \tilde{m}$, since only one regularization parameter needs to be chosen in that case and since preliminary simulations have shown that the performance of our method is not influenced a lot by this restriction. To enhance the quality of the estimated density, we also force the ELD to be smooth at the origin, which is satisfied if the following constraint holds:

$$\frac{\left(1+\sum_{i=1}^{m}\theta_i\right)^2}{1+\sum_{i=1}^{m}\theta_i^2}=\frac{\left(1+\sum_{i=1}^{m}\tilde{\theta}_i\right)^2}{1+\sum_{i=1}^{m}\tilde{\theta}_i^2}. \tag{23}$$

In the simulations the logarithm of (12) will be maximised under the constraint (23) using the COBYLA algorithm (Powell (1994)) from the *nloptr()* function in R (Johnson (2007)). Due to non-convexity of the log-likelihood, the optimization is repeated using multiple starting values (20 in our case) and we then select the maximizer that corresponds to the highest likelihood value. Note that (12) is convex when m equals zero (and we therefore take only one starting value), and that the continuity constraint (23) is automatically satisfied in that case.

We now propose a method to select the regularization parameter m, that is an extension of the classical K-fold cross-validation to our context of quantile regression with interval censored data. We consider here the case where data are never left or right censored. Let $K \in \mathbb{N}_0$ denote the number of folds that will be used, and partition the sample randomly into K equal sized subsamples. For $k = 1, \dots, K$ and $m \in \mathbb{N}$, the prediction error is approximated by

$$\text{Pred.error}_{k,m}=\frac{1}{n_k}\sum_{i=1}^{n_k} w_i\ \rho_\tau\Big(\frac{L_i+R_i}{2}-\boldsymbol{X}_i^t\hat{\beta}_{(-k),m}\Big), \tag{24}$$

where n_k is the number of subjects in group k, $\hat{\beta}_{(-k),m}$ is the vector or regression coefficients estimated from the model with Laguerre polynomials of order m based on all observations except those in fold k, and $w_i = 1/(R_i - L_i)$. Note that the check function ρ_τ is used rather than the Euclidean distance, and that we use the middle of the interval $[L_i, R_i]$ as a substitute for the unknown T_i. The factor w_i gives more weight to short intervals than to long ones, as short intervals contain more accurate information. Next, the following cross-validation criterion is computed by averaging the prediction error over the K groups:

$$K\text{-Fold-cv}_m=\frac{1}{K}\sum_{k=1}^{K}\text{Pred.error}_{k,m}. \tag{25}$$

Finally, the polynomial order m that minimizes this criterion is selected. The procedure described above will be applied with $K = 5$ folds.

5.2. *Simulation design*

We will set up simulations according to the following data generating process:

$$T=\beta_0+\beta_1X_1+\beta_2X_2+\sigma(\boldsymbol{\gamma},\boldsymbol{X})\varepsilon, \tag{26}$$

where X_1, X_2 and ε are mutually independent, X_1 is either generated from an uniform distribution on $[-1, 1]$ or from a Bernoulli distribution with success probability 0.5, and X_2 always follows a Bernoulli(0.5) distribution. For ε we either take a standard extreme value distribution, i.e. $P(\varepsilon > t) = \exp(-(e^t/\sigma)^a)$ with $\sigma = 2$ and $a = 1.5$ (DGP A), or a standard normal distribution (DGP B). The function $\sigma(\boldsymbol{\gamma}, \boldsymbol{X})$ will be chosen later (both homoscedastic and heteroscedastic designs will be considered). Also note that we will sometimes consider the above model without the covariate X_2. Finally, for given τ, we shift the error ε in such a way that the $\tau-$th quantile of ε equals 0.

We next describe how the intervals $[L, R]$ are generated for a given $(T, \boldsymbol{X})$ generated from model (26). First, a set of examination times $Y_{k,j}$, $k = 1, 2, \ldots, j = 1, \ldots, k$, is given for each observation. Let ξ_0 (resp. d) be a number such that the probability of simulating an observation smaller than ξ_0 (resp. larger than $\xi_0 + d$) is very small. In the following simulations ξ_0 is fixed to -10 and $\xi_0 + d$ to 10, which basically means that the data are (almost) never left or right censored. The inspection times are constructed as follows: $Y_{k,j} = \xi_0 + \sum_{\ell=1}^{j} \xi_\ell$, where all ξ_ℓ are independent random variables that have an uniform distribution on $[0.1, l_{\max}]$ for some $l_{\max} > 0.1$. Finally, the interval is constructed as the last inspection time before the event of interest and the first inspection after the occurrence of the event: $L = \max_j\{Y_{k,j} : Y_{k,j} < T\}$ and $R = \min_j\{Y_{k,j} : Y_{k,j} \geq T\}$. The number of inspections is $k = \sup\{k \geq 1 : \sum_{\ell=1}^{k} \xi_\ell \leq d\}$. With this procedure, inspection times are generated independently of T. For this reason, the non-informative censoring assumption (4) is satisfied. An average length for the interval censored observations can be targeted by adjusting the value of $l_{\max}$. Note that $\xi_{k,j}$, $k = 1, 2, \ldots, j = 1, \ldots, k$ are generated from an uniform distribution with minimum 0.1, which implies that assumption (A4) is satisfied.

5.3. Comparison with other estimators

We start by comparing the proposed estimator with other estimators found in the literature. The comparison is done with the methods presented in Kim et al. (2010), Zhou, Feng and Du (2017) and Frumento (2023). The results are also compared with the Laplace distribution used in Yu and Zhang (2005). The estimator of Kim et al. (2010) is based on the Missing Information Principle of Efron (1967), it only works when the covariates are discrete and is only available for the special case of the median. The estimator of Zhou, Feng and Du (2017) is defined by

$$\arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{1+p}} \frac{1}{n} \sum_{i=1}^{n} \Big\{ (1-\tau)|R_i - \max(R_i, \boldsymbol{X}_i^t \boldsymbol{\beta})| + \tau |L_i - \min(L_i, \boldsymbol{X}_i^t \boldsymbol{\beta})| \Big\}$$

and applying the direct bias correction proposed in Section 3.2 of Zhou, Feng and Du (2017). The length of the intervals must be finite for this estimator to be applied, so left or right censoring are excluded. Frumento (2023)'s estimator is a two-step estimator that uses a similar methodology to the one presented in Frumento and Bottai (2017), and that is obtained through the package ctqr in R. Finally, the Laplace column in the tables below corresponds to the results obtained by maximizing (12) for $m = \tilde{m} = 0$. In this subsection no heteroscedasticity is considered and we apply estimator (12) with $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \gamma_0(\tau)$. Tables 1 and 2 are based on 500 Monte Carlo simulations, and we consider in both tables DGP A with $\tau = 0.25$ and DGP B with $\tau = 0.1$. The polynomial orders are selected by K-fold cross-validation among all models with $m \leq 5$.

For Table 1 we do not consider covariate X_2 in the model, and take $\beta_0 = 0, \beta_1 = 1$ and $\gamma_0 = 1$. Covariate X_1 has a uniform distribution on $[-1, 1]$ and hence we cannot compare with Kim et al. (2010)'s estimator, since it only works for discrete covariates. Table 1 shows that for the slope β_1 the smallest MSE is (almost) always obtained by the proposed estimator, independently of the sample size, the average length of the intervals, the error distribution and the value of τ. For the intercept β_0 the smallest MSE is usually attained by the Laplace estimator or the estimator by Frumento (2023). Note that for the estimation of β_0, the Laplace and Zhou, Feng and Du (2017)'s estimator are systematically biased, whereas for β_1 the estimator by Frumento (2023) is often biased. On the other hand, the bias of the proposed estimator is always quite small.

Model	l_c	n		Laplace β_0	Laplace β_1	ELD β_0	ELD β_1	Zhou et al β_0	Zhou et al β_1	Frumento β_0	Frumento β_1
1	1	100	Bias	-0.037	-0.030	-0.011	-0.027	0.107	-0.004	0.014	-0.085
			MSE	0.021	0.041	0.023	**0.030**	0.029	0.047	**0.020**	0.069
		200	Bias	-0.039	0.007	-0.023	-0.005	0.101	0.004	0.002	-0.046
			MSE	**0.009**	0.023	0.015	**0.015**	0.018	0.024	0.010	0.027
		400	Bias	-0.035	-0.003	-0.006	-0.003	0.103	0.003	0.006	-0.040
			MSE	**0.005**	0.008	0.009	**0.007**	0.014	0.012	**0.005**	0.014
	2	100	Bias	-0.078	-0.024	0.014	-0.018	0.062	-0.006	0.039	-0.145
			MSE	0.029	0.048	0.029	**0.041**	**0.026**	0.064	0.030	0.104
		200	Bias	-0.096	-0.005	-0.001	-0.009	0.084	-0.014	-0.005	-0.057
			MSE	0.018	0.028	0.022	**0.018**	0.016	0.029	**0.014**	0.038
		400	Bias	-0.093	-0.011	0.004	-0.023	0.078	-0.023	-0.009	-0.036
			MSE	0.014	0.011	0.013	**0.010**	0.011	0.015	**0.007**	0.018
2	1	100	Bias	0.068	-0.032	0.032	-0.031	0.283	-0.004	0.013	0.033
			MSE	0.032	0.071	0.041	**0.055**	0.109	0.088	**0.029**	0.101
		200	Bias	0.051	-0.008	0.020	-0.016	0.263	0.001	0.021	-0.049
			MSE	**0.016**	0.038	0.024	**0.024**	0.083	0.041	**0.016**	0.061
		400	Bias	0.050	-0.031	0.021	-0.017	0.270	-0.013	0.027	-0.048
			MSE	0.009	0.024	0.015	**0.014**	0.080	0.022	**0.008**	0.032
	2	100	Bias	0.114	-0.014	0.109	-0.003	0.344	-0.061	0.062	-0.071
			MSE	**0.049**	**0.056**	0.069	0.059	0.148	0.093	0.051	0.152
		200	Bias	0.087	0.001	0.065	-0.018	0.362	-0.024	0.050	-0.175
			MSE	0.024	0.036	0.033	**0.026**	0.147	0.047	**0.023**	0.108
		400	Bias	0.082	-0.026	0.034	-0.022	0.358	-0.023	0.066	-0.210
			MSE	0.015	0.023	0.025	**0.016**	0.136	0.025	**0.014**	0.080

TABLE 1

Bias and MSE of the estimators based on the ALD (Laplace), the proposed estimator based on Laguerre polynomials (ELD), and the estimators of Zhou, Feng and Du (2017) and Frumento (2023), for different sample sizes n, values of l_c (the average interval length) and models (model 1 corresponds to DGP A with $\tau = 0.25$, whereas model 2 is DGP B with $\tau = 0.10$). The numbers in bold correspond to the smallest MSE among the 4 estimators.

For Table 2, we now consider a Bernoulli distribution with success probability 0.5 for X_1, and we let X_2 in the model. This change is motivated by the fact that we like to compare our estimator with the one by Kim et al. (2010), which is only defined for discrete covariates. To accommodate the additional restriction that this estimator is only designed for median regression and not for a general quantile level τ, we replace the term 0.5 by τ in equation 3 of Kim's paper. The values of the coefficients in model (26) are $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$ and $\gamma_0 = 1$. Table 2 shows that the proposed estimator outperforms all its competitors in terms of the MSE, independently of the coefficient (β_0, β_1 or β_2), the model, the average length l_c and the sample size n. In fact, in 33 of the 36 MSE's the smallest value is obtained by the proposed estimator.

5.4. Confidence intervals

In this subsection, we construct confidence intervals for the regression coefficients, and compute their coverage probability and average length as a way to verify their performance in practice. The confidence intervals are constructed based on a naive bootstrap procedure. For a given data set we generate B resamples by drawing the observations $(\boldsymbol{X}_i, L_i, R_i)$ $(i = 1, \dots, n)$ with replacement. Next, for each resample and for each $j = 0, 1, \dots, p$ we estimate the regression coefficient β_j, and we then compute the variance $\hat{\sigma}_j^2$ of these B bootstrap coefficients. Finally a $(1 - \alpha) \times 100\%$ confidence interval for β_j is given by $[\hat{\beta}_j - z_{1-\alpha/2}\hat{\sigma}_j, \hat{\beta}_j + z_{1-\alpha/2}\hat{\sigma}_j]$, where $z_{1-\alpha/2}$ is the quantile of order $1 - \alpha/2$ of the standard normal distribution.

To test the performance of these confidence intervals we generate data from model (26) using the extreme value error distribution (DGP A) and using a Bernoulli(0.5) distribution for X_1. The scale function equals $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \gamma_0(\tau) = 1$, $\tau = 0.25$, $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = 1$. The average interval length is $l_c = 0, 5, 1$ or 2. Table 3 shows the results based on 500 Monte Carlo simulations in which 100 bootstrap replications were carried out with $\alpha = 0.05$. For each Monte Carlo simulation, $m \in \{0, 1, \dots, 5\}$ was selected by K-fold cross-validation with $K = 5$. We see that the coverage probabilities of the intervals for β_0 are somewhat lower than the nominal value, whereas for β_1 and β_2 the coverage is close to 95%. The average length decreases consistently when n increases or l_c decreases, as expected.

		β_0		β_1		β_2	
l_c	n	CP	Length	CP	Length	CP	Length
0.5	100	0.906	0.639	0.968	0.820	0.956	0.828
	200	0.906	0.481	0.970	0.610	0.972	0.617
	400	0.912	0.367	0.968	0.451	0.974	0.454
1	100	0.866	0.687	0.950	0.846	0.946	0.862
	200	0.878	0.525	0.960	0.656	0.962	0.652
	400	0.894	0.405	0.958	0.504	0.954	0.492
2	100	0.826	0.785	0.942	0.931	0.934	0.937
	200	0.862	0.606	0.928	0.686	0.936	0.689
	400	0.868	0.493	0.938	0.515	0.948	0.513

TABLE 3

Coverage probability (CP) and average length of bootstrap confidence intervals for the regression coefficients β_0, β_1 and β_2 under DGP A.

Model	l_c	n		Laplace			ELD			Zhou et al			Kim et al			Frumento		
				β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
1	1	100	Bias	-0.027	-0.021	0.009	-0.005	-0.024	-0.006	0.112	-0.001	0.000	-0.127	-0.026	0.002	0.005	0.001	0.011
			MSE	0.050	0.059	0.060	**0.032**	**0.033**	**0.030**	0.056	0.062	0.060	0.073	0.073	0.069	0.080	0.093	0.088
		200	Bias	-0.045	0.010	-0.001	-0.014	-0.022	-0.017	0.105	-0.001	-0.000	-0.155	0.012	0.016	0.071	-0.041	-0.049
			MSE	0.026	0.032	0.028	**0.017**	**0.019**	**0.019**	0.036	0.031	0.030	0.054	0.037	0.034	0.032	0.048	0.048
		400	Bias	-0.028	0.005	-0.013	-0.001	-0.010	-0.018	0.109	-0.003	-0.005	-0.141	0.008	-0.014	0.070	-0.043	-0.054
			MSE	0.012	0.012	0.015	**0.011**	**0.010**	**0.012**	0.022	0.016	0.014	0.033	0.014	0.015	0.017	0.028	0.026
	2	100	Bias	-0.075	-0.015	0.005	0.065	-0.005	-0.041	0.055	0.046	0.028	-0.440	-0.017	-0.011	0.173	-0.104	-0.094
			MSE	0.056	0.057	0.072	0.060	**0.052**	**0.053**	**0.058**	0.087	0.082	0.270	0.082	0.112	0.161	0.138	0.147
		200	Bias	-0.114	0.033	0.022	0.027	-0.001	0.016	0.056	0.036	0.038	-0.537	0.019	0.060	0.087	-0.072	-0.054
			MSE	0.042	0.040	0.032	0.042	**0.025**	**0.023**	**0.034**	0.041	0.043	0.335	0.045	0.061	0.045	0.083	0.075
		400	Bias	-0.089	0.005	-0.002	0.023	-0.006	-0.007	0.075	0.004	0.007	-0.515	-0.006	0.015	0.050	-0.027	-0.035
			MSE	0.024	0.016	0.017	**0.017**	**0.011**	**0.013**	0.020	0.021	0.020	0.287	0.022	0.025	0.018	0.041	0.042
2	1	100	Bias	0.098	-0.031	-0.008	0.037	-0.034	-0.033	0.291	-0.001	-0.005	-0.102	-0.003	0.002	-0.036	0.036	0.040
			MSE	0.097	0.094	0.083	**0.045**	**0.057**	**0.064**	0.166	0.105	0.101	0.108	0.110	0.120	0.088	0.143	0.146
		200	Bias	0.055	-0.006	0.012	0.034	-0.025	0.008	0.263	0.012	0.003	-0.122	-0.002	0.012	-0.005	0.026	0.013
			MSE	0.044	0.049	0.049	**0.032**	**0.041**	**0.035**	0.117	0.064	0.058	0.058	0.050	0.060	0.059	0.076	0.071
		400	Bias	0.069	-0.021	-0.013	0.039	-0.039	-0.021	0.270	0.010	-0.003	-0.110	-0.024	-0.004	-0.036	0.033	0.040
			MSE	**0.020**	0.026	0.023	0.022	**0.019**	**0.015**	0.094	0.027	0.028	0.032	0.027	0.031	0.031	0.039	0.038
	2	100	Bias	0.141	-0.044	-0.017	0.100	-0.047	-0.021	0.331	0.035	0.034	-0.462	-0.004	0.018	0.008	0.010	0.003
			MSE	0.109	0.110	0.096	**0.083**	**0.069**	**0.067**	0.210	0.153	0.145	0.314	0.132	0.133	0.149	0.225	0.195
		200	Bias	0.092	-0.015	0.004	0.042	-0.032	-0.012	0.364	-0.009	-0.012	-0.503	0.002	0.019	0.159	-0.093	-0.088
			MSE	0.058	0.055	0.056	**0.042**	**0.029**	**0.038**	0.182	0.076	0.073	0.312	0.079	0.085	0.132	0.137	0.134
		400	Bias	0.090	-0.016	-0.002	0.028	-0.024	-0.014	0.354	-0.005	-0.001	-0.515	-0.001	-0.011	0.118	-0.074	-0.062
			MSE	0.025	0.026	0.026	**0.023**	**0.018**	**0.018**	0.148	0.033	0.032	0.294	0.050	0.048	0.088	0.086	0.088

Table 2: Bias and MSE of the estimators based on the ALD (Laplace), the proposed estimator based on Laguerre polynomials (ELD), and the estimators of Zhou, Feng and Du (2017), Kim et al. (2010) and Frumento (2023), for different sample sizes n, values of l_c (the average interval length) and models (model 1 corresponds to DGP A with $\tau = 0.25$, whereas model 2 is DGP B with $\tau = 0.1$). The numbers in bold correspond to the smallest MSE among the 5 estimators.

5.5. Heteroscedastic models

In this subsection we investigate the performance of our estimator when the model is heteroscedastic, and the scale function $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X})$ is possibly misspecified. Consider model (26) with DGP A for the error distribution and with a uniform[-1,1] distribution for X_1, and consider $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = 1 + X_1$, so the true scale varies between 0 and 2, which is a substantial amount of heteroscedasticity. We omit covariate X_2 from the model. The true values of the coefficients are $\beta_0 = 0$ and $\beta_1 = 1$. For each dataset, three models were fitted for $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X})$, namely $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \gamma_0$ (model (I)), $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \gamma_0 + \gamma_1 X_1$ (model (II)), and $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \exp(\gamma_0 + \gamma_1 X_1)$ (model (III)). Note that models (I) and (III) are misspecified, whereas model (II) is correctly specified. As before, the polynomial orders are selected by K-fold cross-validation among all models with $m \leq 5$. The results presented in Table 4 are obtained based on 500 Monte Carlo replications. The table shows that the performance of the estimators of β_0 and β_1 deteriorates when the model for the scale function is misspecified, as can be expected. Both the bias and the variance of the estimators increases, especially for the homoscedastic model fit, which gives a poor fit to the data.

				I		II		III	
τ	l_c	n		β_0	β_1	β_0	β_1	β_0	β_1
0.5	0.5	100	Bias	0.049	0.179	0.086	0.091	0.115	0.154
			SD	0.135	0.238	0.118	0.141	0.137	0.174
			MSE	0.021	0.089	0.021	0.028	0.032	0.054
		400	Bias	0.046	0.157	0.100	0.101	0.098	0.135
			SD	0.071	0.126	0.082	0.092	0.078	0.094
			MSE	0.007	0.041	0.017	0.019	0.016	0.027
	2	100	Bias	0.028	0.098	0.077	0.073	0.092	0.105
			SD	0.177	0.279	0.146	0.185	0.166	0.209
			MSE	0.032	0.088	0.027	0.039	0.036	0.055
		400	Bias	0.018	0.102	0.098	0.091	0.085	0.097
			SD	0.087	0.144	0.094	0.106	0.093	0.112
			MSE	0.008	0.031	0.018	0.020	0.016	0.022
0.75	0.5	100	Bias	-0.027	-0.096	0.018	0.014	0.033	0.001
			SD	0.120	0.214	0.081	0.094	0.091	0.120
			MSE	0.015	0.055	0.007	0.009	0.009	0.014
		400	Bias	-0.012	-0.071	0.020	0.019	0.040	0.004
			SD	0.063	0.105	0.050	0.053	0.046	0.061
			MSE	0.004	0.016	0.003	0.003	0.004	0.004
	2	100	Bias	-0.072	-0.251	-0.032	-0.058	0.006	-0.075
			SD	0.142	0.244	0.112	0.137	0.110	0.157
			MSE	0.025	0.123	0.014	0.022	0.012	0.030
		400	Bias	-0.056	-0.246	0.012	-0.003	0.028	-0.045
			SD	0.102	0.135	0.061	0.071	0.055	0.086
			MSE	0.013	0.079	0.004	0.005	0.004	0.009

TABLE 4

Bias, standard deviation (SD) and mean squared error (MSE) of the estimators of β_0 and β_1 when the model is heteroscedastic and the data are fitted using three models for the variance function. Model (II) is correctly specified, whereas models (I) and (III) are misspecified.

6. Real data analysis

In 2019, a survey was conducted on the first employment of 813 university graduates from a certain European university. Each respondent provided information about their age, gender, diploma and their starting salary at their first job. Since participants were asked to enter the range of their starting salary and not the exact value, this variable is subject to interval censoring. We are interested in understanding which factors have an influence on the starting salary, and whether these factors are different for low and high salaries. We will use our methodology based on the ELD to answer this question.

In total 325 men and 488 women responded to the survey. In an attempt to keep a parsimonious model, the different diplomas have been grouped into 3 different sectors: Human Sciences Sector (HSS), Health Sciences Sector (MED) and Science and Technology (SST). It is suspected that the effect of age on starting salary works differently from one sector to another. For this reason, a first-order interaction effect between the variables age and sector is included in the model. We propose to use the model

$$T = \beta_0(\tau) + \sum_{j=1}^{4} \beta_j(\tau) X_j + \beta_5(\tau) X_2 X_3 + \beta_6(\tau) X_2 X_4 + \sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X})\varepsilon,$$

where T is the logarithm of the gross starting salary in € divided by 1000, X_1 is gender (1 for men and 0 for women), X_2 is the age centered to the mean which is equal to 26.43 (minimum, first quartile, median, third quartile and maximum are given respectively by 23, 24, 26, 28 and 63), and X_3 (resp. X_4) equals 1 if the respondent has a diploma from MED (resp. SST) and 0 otherwise. Finally, $\sigma(\boldsymbol{\gamma}(\tau), \boldsymbol{X}) = \exp(\gamma_0(\tau) + \sum_{j=1}^{4} \gamma_j(\tau) X_j + \gamma_5(\tau) X_2 X_3 + \gamma_6(\tau) X_2 X_4)$. The model is fitted for a range of values of τ starting from 0.1 to 0.85 with a step size of 0.05. The K-fold cross-validation procedure outlined in Subsection 5.1 is used to select the polynomial order $m = \tilde{m}$. The only difference is that whenever $R_i = \infty$ (69 right censored observation) we replace $(L_i + R_i)/2$ by L_i in (24) and (25). Similarly for the 7 left censored observations, $(L_i + R_i)/2$ is replaced by R_i. The selection is performed among all models with $m \leq 5$ and the respondents are divided randomly in $K = 5$ folds. Similarly as in the simulations, we maximize the log-likelihood under the constraint (23), which ensures that the estimated density is continuous at the origin.

Figure 1 shows the estimated quantile coefficient curves $\hat{\beta}_0(\tau), \ldots, \hat{\beta}_6(\tau)$ for $\tau = 0.1, 0.15, \ldots, 0.85$. The shaded area represents 95% pointwise bootstrap confidence intervals following the method presented in Subsection 5.4. The confidence intervals are computed based on $B = 200$ resamples. A number of interesting findings can be seen in this figure. First of all, men receive a larger starting salary than women. In particular $\hat{\beta}_1(0.5)$ is 0.098. This suggests that the median starting salary for a man is $\exp(0.098) = 1.103$ times larger than the starting salary of a woman (keeping all other variables fixed). An age effect on the starting salary is also visible. Even if the effect does not seem present in the lower tail, we can see a significant increase for larger quantiles. For example,

$\hat{\beta}_2(0.5)$ is 0.028, so our model suggests that the median starting salary increases with 2.8% for every extra year of age. Regarding the effect of the different sectors, it can be observed that the medical sector provides the highest starting wage. We can also see that the interaction between age and the medical sector is significant. On the other hand, for the science and technology sector, no significant effect is seen compared to the humanities sector, both for the main and the interaction effect with age.

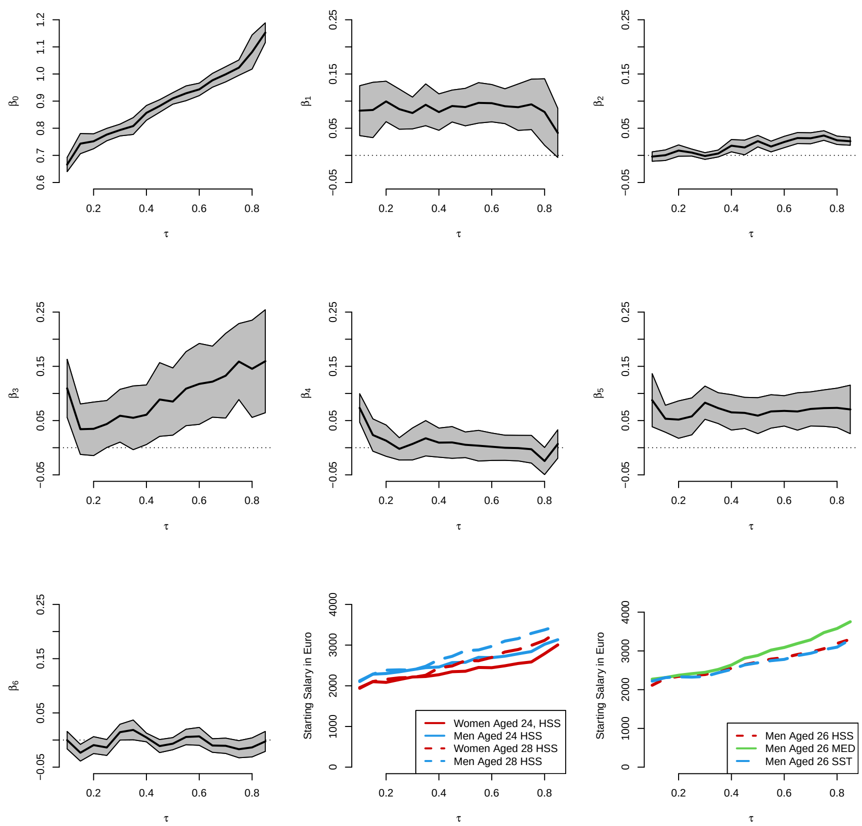

Fig 1: Estimated quantile coefficient curves $\hat{\beta}_0(\tau)$ (top-left), $\ldots, \hat{\beta}_6(\tau)$ bottom-left) for $\tau = 0.1, 0.15, \ldots, 0.85$ for the starting salary data, together with their 95% confidence intervals. The bottom-center figure shows the estimated quantile function for a man and a woman aged 24 and 28 from the humanities sector, while the bottom-right figure depicts the quantiles of a man aged 26 for each sector.

7. Conclusion

In this paper we proposed a new approach to perform quantile regression when the response variable is interval censored. The proposed estimator is based on the so-called Enriched Laplace distribution. This distribution extends the usual Laplace distribution by multiplying it by squared linear combinations of Laguerre polynomials, in such a way that the quantile of order τ is by construction equal to zero (for given $0 < \tau < 1$). This property is very useful in the context of quantile regression, as it can be used to approximate the distribution of the error term in a linear quantile regression model. The consistency of the proposed estimator is shown and its practical performance is studied via thorough simulations.

In the future it would be interesting to study how the Enriched Laplace distribution can be used when the data are imperfect for other reasons than interval censoring in the response. This includes, among others, the case of right or interval censoring in the covariates, measurement errors, missing data, etc. In fact, since our approach is likelihood based, it can be used in any context where inference can be done by maximizing a likelihood.

Appendix A: Some properties regarding the Enriched Laplace distribution

A.1. Density function

To show that $f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}$ is a density, we will show that $\int_{-\infty}^{0} f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t)\,dt = \tau$. In a similar way it can be shown that $\int_{0}^{\infty} f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t)\,dt = 1-\tau$, from which the statement follows. Write,

$$\begin{aligned}
\int_{-\infty}^{0} f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t)\,dt &= \tau(1-\tau)\int_{-\infty}^{0} e^{t(1-\tau)}\Big[\sum_{k=0}^{\tilde{m}}\tilde{\theta}_k v_k(-t(1-\tau))\Big]^2 dt \\
&= \tau\int_{0}^{\infty} e^{-s}\Big[\sum_{k=0}^{\tilde{m}}\tilde{\theta}_k v_k(s)\Big]^2 ds \\
&= \tau\sum_{k_1=0}^{\tilde{m}}\sum_{k_2=0}^{\tilde{m}}\tilde{\theta}_{k_1}\tilde{\theta}_{k_2}\int_{0}^{\infty} e^{-s} v_{k_1}(s) v_{k_2}(s)\,ds \\
&= \tau\sum_{k=0}^{\tilde{m}}\tilde{\theta}_k^2 = \tau,
\end{aligned}$$

which follows from the orthonormality of Laguerre polynomials and the fact that the Euclidean norm of $\tilde{\boldsymbol{\theta}}$ equals 1.

A.2. Cumulative distribution function

Write $v_k(t) = \sum_{j=0}^{k} A_{j,k} t^j$, where

$$A_{j,k} = \binom{k}{j} \frac{(-1)^j}{j!}.$$

Then, for $t \le 0$,

$$\begin{aligned}
&F_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t) \\
&= \tau(1-\tau) \int_{-\infty}^{t} e^{s(1-\tau)} \Big[\sum_{k=0}^{\tilde{m}} \tilde{\theta}_k v_k(-s(1-\tau)) \Big]^2 \, ds \\
&= \tau \int_{-t(1-\tau)}^{\infty} e^{-s} \Big[\sum_{k=0}^{\tilde{m}} \tilde{\theta}_k v_k(s) \Big]^2 \, ds \\
&= \tau \sum_{k_1=0}^{\tilde{m}} \sum_{k_2=0}^{\tilde{m}} \tilde{\theta}_{k_1} \tilde{\theta}_{k_2} \int_{-t(1-\tau)}^{\infty} e^{-s} v_{k_1}(s) v_{k_2}(s) \, ds \\
&= \tau \sum_{k_1=0}^{\tilde{m}} \sum_{k_2=0}^{\tilde{m}} \tilde{\theta}_{k_1} \tilde{\theta}_{k_2} \int_{-t(1-\tau)}^{\infty} e^{-s} \Big(\sum_{j_1=0}^{k_1} A_{j_1,k_1} s^{j_1} \Big) \Big(\sum_{j_2=0}^{k_2} A_{j_2,k_2} s^{j_2} \Big) \, ds \\
&= \tau \sum_{k_1=0}^{\tilde{m}} \sum_{k_2=0}^{\tilde{m}} \tilde{\theta}_{k_1} \tilde{\theta}_{k_2} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} A_{j_1,k_1} A_{j_2,k_2} \int_{-t(1-\tau)}^{\infty} e^{-s} s^{j_1+j_2} \, ds \\
&= -\tau \sum_{k_1=0}^{\tilde{m}} \sum_{k_2=0}^{\tilde{m}} \tilde{\theta}_{k_1} \tilde{\theta}_{k_2} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} A_{j_1,k_1} A_{j_2,k_2} \sum_{\ell=0}^{j_1+j_2} \frac{(j_1+j_2)!}{\ell!} s^{\ell} e^{-s} \Big|_{-t(1-\tau)}^{\infty} \\
&= \tau \sum_{k_1=0}^{\tilde{m}} \sum_{k_2=0}^{\tilde{m}} \tilde{\theta}_{k_1} \tilde{\theta}_{k_2} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} A_{j_1,k_1} A_{j_2,k_2} \sum_{\ell=0}^{j_1+j_2} \frac{(j_1+j_2)!}{\ell!} (-t(1-\tau))^{\ell} e^{t(1-\tau)},
\end{aligned}$$

since it can be easily seen that for any $y \ge 0$ and for any $k \in \mathbb{N}$,

$$\int_{y}^{\infty} e^{-s} s^k ds = - \sum_{\ell=0}^{k} \frac{k!}{\ell!} s^{\ell} e^{-s} \Big|_{y}^{\infty}.$$

Note that $F_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(0) = \tau$ since it can be easily shown that

$$\sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} A_{j_1,k_1} A_{j_2,k_2} (j_1+j_2)! = \mathbb{1}_{k_1=k_2}.$$

Similarly, for $t \ge 0$,

$$\begin{aligned}
&F_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t) \\
&= 1 - (1-\tau) \sum_{k_1=0}^{m} \sum_{k_2=0}^{m} \theta_{k_1} \theta_{k_2} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} A_{j_1,k_1} A_{j_2,k_2} \sum_{\ell=0}^{j_1+j_2} \frac{(j_1+j_2)!}{\ell!} (t\tau)^{\ell} e^{-t\tau}.
\end{aligned}$$

A.3. Proof of Proposition 3.1

Part (i) follows directly from the calculations done in Appendix A.1. For part (ii), note that in Section 4.1 of Florens, Simar and Van Keilegom (2020) it has been shown that if Z is a random variable defined on the positive real line and

$$f_m(t|\boldsymbol{\theta}) = e^{-t} P^2_{m,\boldsymbol{\theta}}(t) \mathbb{1}_{t\geq 0},$$

with $\boldsymbol{\theta}$ in the m-dimensional semi-sphere of radius 1, and if A is a subset of $\mathbb{R}^+$ for which $\inf_{t\in A} f_Z(t) > 0$, then

$$\lim_{m\to\infty} \int_A \Big(\sqrt{f_Z(t)} - \sqrt{f_m(t|\boldsymbol{\theta})}\Big)^2 dt = 0, \tag{27}$$

with

$$\boldsymbol{\theta} = \Big(\int_0^\infty \varphi(t) v_k(t) e^{-t} dt\Big)_{k=0}^m \quad \text{and} \quad \varphi(z) = \sqrt{f_Z(z) e^z}.$$

Now write $f_{\epsilon(\tau)}(t) = \tau(1-\tau)\big[f_Z(t\tau)\mathbb{1}_{t\geq 0} + f_{\tilde{Z}}(-t(1-\tau))\mathbb{1}_{t<0}\big]$, where $\tilde{Z}$ is another random variable defined on the positive real line, and where

$$f_Z(z) = \frac{1}{\tau(1-\tau)} f_{\epsilon(\tau)}\Big(\frac{z}{\tau}\Big)\mathbb{1}_{z\geq 0} \quad \text{and} \quad f_{\tilde{Z}}(z) = \frac{1}{\tau(1-\tau)} f_{\epsilon(\tau)}\Big(-\frac{z}{1-\tau}\Big)\mathbb{1}_{z\geq 0}.$$

Then, with $A^+ = A \cap [0,\infty)$ and $A^- = A \cap (-\infty, 0)$, we have

$$\begin{aligned}
&\int_A \Big(\sqrt{f_{\epsilon(\tau)}(t)} - \sqrt{f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}(t)}\Big)^2 dt \\
&= \tau(1-\tau) \int_{A^+} \Big(\sqrt{f_Z(t\tau)} - \sqrt{f_m(t\tau|\boldsymbol{\theta})}\Big)^2 dt \\
&\quad + \tau(1-\tau) \int_{A^-} \Big(\sqrt{f_{\tilde{Z}}(-t(1-\tau))} - \sqrt{f_{\tilde{m}}(-t(1-\tau)|\tilde{\boldsymbol{\theta}})}\Big)^2 dt,
\end{aligned}$$

and this tends to zero as m and $\tilde{m}$ tend to infinity, thanks to (27). Hence, part (ii) follows provided

$$\boldsymbol{\theta} = \Big(\frac{1}{\sqrt{\tau(1-\tau)}} \int_0^\infty \sqrt{f_{\epsilon(\tau)}\Big(\frac{t}{\tau}\Big) e^{-t}}\, v_k(t)\, dt\Big)_{k=0}^m,$$

and similarly for $\tilde{\boldsymbol{\theta}}$.

A.4. Plots of the Enriched Laplace density

Plots of the Enriched Laplace density for $m = \tilde{m} = 2$ and for several combinations of the parameters are given in Figure 2. Note that the plot for $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}} = (1,0,0)$ corresponds to the asymmetric Laplace density. The figure illustrates how flexible the family is with only $m = 2$. Larger values of m lead to more parameters, and so even more flexibility.

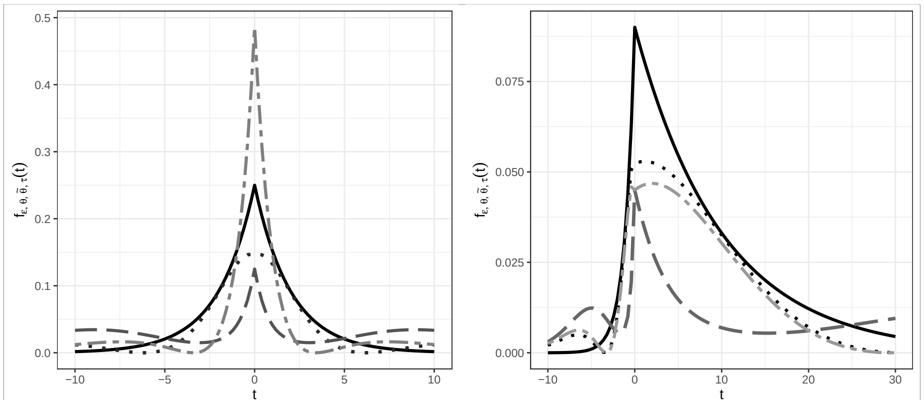

Fig 2: Plot of $f_{\epsilon,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}},\tau}$ for several values of $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, for $m = 2$, and for $\tau = 0.5$ (left) and $\tau = 0.1$ (right): $\boldsymbol{\theta} = (1, 0, 0)$ (black solid), $\boldsymbol{\theta} = (1/\sqrt{2}, 1/2, -1/2)$ (dark grey dashed), $\boldsymbol{\theta} = (1/\sqrt{2}, 1/2, 1/2)$ (light grey dash-dotted) and $\boldsymbol{\theta} = (1/\sqrt{2}, -1/2, 1/2)$ (dotted).

Appendix B: Proofs of Section 4

This Appendix is organized as follows. In Subsection B.1 we give some definitions and auxiliary results, which will be needed in the proofs of the main results. Subsections B.2 and B.3 are devoted to the proofs of these auxiliary results, whereas the proofs of Theorem 4.1 and Corollary 4.1 are given in Subsections B.4 and B.5, respectively.

B.1. Definitions and auxiliary results

We start by showing that for $\alpha = -1/2$, $\mathcal{H}_\alpha$, defined in (20), reduces to twice the squared Hellinger distance:

$$
\begin{aligned}
&\mathcal{H}_{-1/2}(p_1, p_2) \\
&= -2 \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\sqrt{p_1(\boldsymbol{x}, l, r) p_2(\boldsymbol{x}, l, r)} - p_1(\boldsymbol{x}, l, r) \right) dl\, dr\, d\boldsymbol{x} \\
&= 2 - 2 \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{p_1(\boldsymbol{x}, l, r) p_2(\boldsymbol{x}, l, r)}\, dl\, dr\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left((p_1 + p_2)(\boldsymbol{x}, l, r) - 2\sqrt{p_1(\boldsymbol{x}, l, r) p_2(\boldsymbol{x}, l, r)} \right) dl\, dr\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\sqrt{p_1(\boldsymbol{x}, l, r)} - \sqrt{p_2(\boldsymbol{x}, l, r)} \right)^2 dl\, dr\, d\boldsymbol{x}
\end{aligned}
$$

Next, we give the definition of the bracketing number of a subset $(\mathcal{F}, \|\cdot\|)$ of a normed space of real functions $f : \mathcal{X} \to \mathbb{R}$ on a set $\mathcal{X}$. Given two functions $\ell(\cdot)$ and $u(\cdot)$, the bracket $[\ell, u]$ is the set of all functions $f \in \mathcal{F}$ with $\ell(x) \leq f(x) \leq u(x)$, for all $x \in \mathcal{X}$. An ε-bracket is a bracket $[\ell, u]$ with $\|u - \ell\| < \varepsilon$. The bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimum number of ε-brackets needed to cover $\mathcal{F}$.

The next theorem can be found in Kreiss and Van Keilegom (2023), and is an adapted version of Theorem 4 in Wong and Shen (1995). To show our Theorem 4.1 we will verify the conditions of this theorem.

Theorem B.1. *Suppose that there are constants $c_1, c_2 > 0$ and a sequence $\varepsilon_n > 0$ such that*

$$\int_{\varepsilon_n^2/2^8}^{\sqrt{2}\varepsilon_n} \sqrt{\log N_{[\,]}\Big(\frac{u}{c_1}, \mathcal{F}_n, H\Big)}du \leq c_2\sqrt{n}\varepsilon_n^2. \tag{28}$$

Then, for any $\alpha \in (0, 1]$, if $\delta_n(\alpha) < 1/\alpha$, we have that

$$P\Big(H(h, \hat{h}_n) \geq \varepsilon_n^*(\alpha)\Big) \leq 5\exp\Big(-c_2 n\varepsilon_n^*(\alpha)^2\Big) + \exp\Big(-\frac{1}{4}n\alpha c_1\varepsilon_n^*(\alpha)^2\Big), \tag{29}$$

where

$$\varepsilon_n^*(\alpha) = \max\Big(\varepsilon_n, \sqrt{\frac{4\delta_n(\alpha)}{c_1}}\Big).$$

We finish this subsection with two lemmas, which were adapted from Kreiss and Van Keilegom (2023). Their proofs are given in Subsections B.2 and B.3. For the first lemma, we need to introduce the following notations. For any $f_n \in \mathcal{F}_{m_n,\tilde{m}_n}$, let F_n be the distribution function of f_n, and let $\mathcal{X}$ be the support of $\boldsymbol{X}$. For any $\alpha < 1$, we denote:

$$\begin{aligned}
C_n(\alpha) = \sup_{\boldsymbol{x}\in\mathcal{X}}\Big(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}&\Big(\frac{F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})}{F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x})}\Big)^{\frac{\alpha}{1-\alpha}} \\
&\times (F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}))f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})\,d\ell\,dr\Big)^{1-\alpha},
\end{aligned}$$

$$\delta_{+,n}(\boldsymbol{x}) = \frac{1-\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\int_0^{\infty} f_\tau\Big(\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big)\Big|(p_+ - p_{+,n})\Big(\tau\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big)\Big|^2 du,$$

$$\delta_{-,n}(\boldsymbol{x}) = \frac{\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\int_{-\infty}^{0} f_{1-\tau}\Big(-\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big)\Big|(p_- - p_{-,n})\Big(-(1-\tau)\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})}\Big)\Big|^2 du.$$

Lemma B.1. *For $\alpha \in (0, 1/2)$,*

$$\delta_n(\alpha) \leq \frac{1}{\alpha}C_n(\alpha)\int_{\mathcal{X}}\Big(\delta_{+,n}(\boldsymbol{x})^\alpha + \delta_{-,n}(\boldsymbol{x})^\alpha\Big)f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}.$$

In the following lemma, a bound on the bracketing number is constructed.

Lemma B.2. *There exist constants $0 < C_1, C_2 < \infty$ such that for $\varepsilon > 0$,*

$$\log N_{[\,]}(\varepsilon, \mathcal{F}_n, H) \leq C_1(m_n + \tilde{m}_n)\log\Big(C_2\frac{m_n + \tilde{m}_n}{\varepsilon^2}\Big).$$

B.2. Proof of Lemma B.1

The proof presented below is an adaptation from right censoring to the case of interval censoring of the proof of Lemma 4.1 in Kreiss and Van Keilegom (2023)). First, we compute

$$\mathcal{H}_\alpha(h, h_n) = \frac{1}{\alpha} \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})) f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) \times \Big[\Big(\frac{F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})}{F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x})}\Big)^\alpha - 1\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\, d\ell\, dr\, d\boldsymbol{x},$$

which after putting everything on the same denominator, equals

$$\begin{aligned}
&\frac{1}{\alpha} \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})) f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})}{(F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^\alpha} \\
&\qquad \times \Big[(F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}))^\alpha - (F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^\alpha\Big] f_{\boldsymbol{X}}(\boldsymbol{x})\, d\ell\, dr\, d\boldsymbol{x} \\
&= \frac{1}{\alpha} \int_{\mathbb{R}^{p+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Big[\frac{F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x})}{(F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^\alpha} f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})^{1-\alpha}\Big] \\
&\qquad \times f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})^\alpha \Big[(F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}))^\alpha - (F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^\alpha\Big] \\
&\qquad \times f_{\boldsymbol{X}}(\boldsymbol{x})\, d\ell\, dr\, d\boldsymbol{x}.
\end{aligned}$$

Next, Hölder's inequality is applied with $p = 1/(1-\alpha)$ and $q = 1/\alpha$. It follows that

$$\begin{aligned}
&\mathcal{H}_\alpha(h, h_n) \\
&\le \frac{1}{\alpha} \int_{\mathbb{R}^{p+1}} \Big[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}))^{\frac{1}{1-\alpha}}}{(F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^{\frac{\alpha}{1-\alpha}}} f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x})\, d\ell\, dr\Big]^{1-\alpha} \\
&\qquad \times \Big[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) \\
&\qquad\qquad \times \Big|(F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}))^\alpha - (F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x}))^\alpha\Big|^{\frac{1}{\alpha}} d\ell\, dr\Big]^\alpha \\
&\qquad \times f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}. \qquad (30)
\end{aligned}$$

Now, write $F(r|\boldsymbol{x}) - F(\ell|\boldsymbol{x}) = \int_\ell^r f(u|\boldsymbol{x})\, du$, and similarly for $F_n(r|\boldsymbol{x}) - F_n(\ell|\boldsymbol{x})$. We can show that

$$\|g\|_{\frac{1}{\alpha},\ell,r} = \Big(\int_\ell^r g(u|\boldsymbol{x})^{\frac{1}{\alpha}}\, du\Big)^\alpha$$

is a norm, for any $\alpha > 0$, for $\ell < r$ and for any positive function g. By applying the reverse triangular inequality, we obtain

$$\begin{aligned}
\Big|\Big(\int_\ell^r f(u|\boldsymbol{x})\, du\Big)^\alpha - \Big(\int_\ell^r f_n(u|\boldsymbol{x})\, du\Big)^\alpha\Big| &= \Big|\|f^\alpha\|_{\frac{1}{\alpha},\ell,r} - \|f_n^\alpha\|_{\frac{1}{\alpha},\ell,r}\Big| \\
&\le \|f^\alpha - f_n^\alpha\|_{\frac{1}{\alpha},\ell,r}
\end{aligned}$$

$$= \Big(\int_\ell^r |f(u|\boldsymbol{x})^\alpha - f_n(u|\boldsymbol{x})^\alpha|^{\frac{1}{\alpha}}\, du \Big)^\alpha.$$

It now follows that (30) is bounded by

$$\frac{1}{\alpha} C_n(\alpha) \int_{\mathbb{R}^{p+1}} \Big(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) \times \int_\ell^r \Big| f(u|\boldsymbol{x})^\alpha - f_n(u|\boldsymbol{x})^\alpha \Big|^{\frac{1}{\alpha}} du\, d\ell\, dr \Big)^\alpha f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{31}$$

Next, bounding the integral from ℓ to r by the integral over the whole real line, we obtain that

$$\begin{aligned}
&\mathcal{H}_\alpha(h, h_n) \\
&\le \frac{1}{\alpha} C_n(\alpha) \int_{\mathbb{R}^{p+1}} \Big(\int_{\mathbb{R}} \Big| f(u|\boldsymbol{x})^\alpha - f_n(u|\boldsymbol{x})^\alpha \Big|^{\frac{1}{\alpha}} du \Big)^\alpha f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\
&\le \frac{1}{\alpha} C_n(\alpha) \int_{\mathbb{R}^{p+1}} \Big(\int_0^\infty \frac{1-\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_\tau\Big(\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big| (p_+^{2\alpha} - p_{+,n}^{2\alpha}) \Big(\tau \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big|^{\frac{1}{\alpha}} du \\
&\quad + \int_{-\infty}^0 \frac{\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_{1-\tau}\Big(- \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big| (p_-^{2\alpha} - p_{-,n}^{2\alpha}) \Big(-(1-\tau) \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big|^{\frac{1}{\alpha}} du \Big)^\alpha \\
&\quad \times f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\
&\le \frac{1}{\alpha} C_n(\alpha) \int_{\mathbb{R}^{p+1}} \Big(\int_0^\infty \frac{1-\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_\tau\Big(\frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big| (p_+ - p_{+,n}) \Big(\tau \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big|^2 du \\
&\quad + \int_{-\infty}^0 \frac{\tau}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} f_{1-\tau}\Big(- \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big| (p_- - p_{-,n}) \Big(-(1-\tau) \frac{u - \boldsymbol{x}^t\boldsymbol{\beta}}{\sigma(\boldsymbol{\gamma}, \boldsymbol{x})} \Big) \Big|^2 du \Big)^\alpha \\
&\quad \times f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} \\
&\le \frac{1}{\alpha} C_n(\alpha) \int_{\mathbb{R}^{p+1}} (\delta_{+,n}(\boldsymbol{x})^\alpha + \delta_{-,n}(\boldsymbol{x})^\alpha) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x},
\end{aligned}$$

where the third inequality follows from the relation $|a^{2\alpha} - b^{2\alpha}| \le |a - b|^{2\alpha}$ for $a, b > 0$ and $\alpha \in (0, 1/2)$, and similarly, the last inequalty is justified by using $(a + b)^\alpha \le a^\alpha + b^\alpha$ for $a, b > 0$ and $\alpha \in (0, 1)$.

B.3. Proof of Lemma B.2

We recall that $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})^t$, where $\boldsymbol{\theta} = (\theta_0, \dots, \theta_{m_n}) \in \boldsymbol{S}_{m_n}$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_0, \dots, \tilde{\theta}_{\tilde{m}_n}) \in \boldsymbol{S}_{\tilde{m}_n}$, and that $\boldsymbol{\alpha} \in A_n = B \times \Gamma \times \boldsymbol{S}_{m_n} \times \boldsymbol{S}_{\tilde{m}_n}$. Before giving the proof of Lemma B.2, we start with some definitions. For $\varepsilon > 0$, let $B_\varepsilon(\boldsymbol{\alpha})$ denote the ball of radius ε (with respect to the Euclidean norm $\|\cdot\|$) centered around $\boldsymbol{\alpha}$. We say that $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N\}$ is a ε-covering of the space A_n if the balls $B_\varepsilon(\boldsymbol{\alpha}_1), \dots, B_\varepsilon(\boldsymbol{\alpha}_N)$ cover the space A_n, i.e. for all $\boldsymbol{\alpha}$ in A_n, there is a $k = 1, \dots, N$ such that $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\| \le \varepsilon$. For given $\varepsilon > 0$, the covering number $N(\varepsilon, A_n, \|\cdot\|)$ is the size of the smallest ε-covering of A_n.

Let $u > 0$ and let $\{B_u(\boldsymbol{\alpha}_k) : k = 1, \ldots, N(u)\}$ be a u-covering of the parameter space A_n, with $N(u) = N(u, A_n, \|\cdot\|)$. Since B and Γ are compact subsets of $\mathbb{R}^{p+1}$ and $\boldsymbol{S}_{m_n}$ and $\boldsymbol{S}_{\tilde{m}_n}$ are semi-spheres in $\mathbb{R}^{m_n+1}$ and $\mathbb{R}^{\tilde{m}_n+1}$ respectively, it is easily seen that $N(u) \leq (C/u)^{m_n+\tilde{m}_n+2p+2}$ for some $0 < C < \infty$. We define the brackets as follows:

$$\underline{M}_k(\boldsymbol{x}, \ell, r|u) = \inf_{\boldsymbol{\alpha} \in B_u(\boldsymbol{\alpha}_k)} h(\boldsymbol{x}, \ell, r|\boldsymbol{\alpha})$$
$$\overline{M}_k(\boldsymbol{x}, \ell, r|u) = \sup_{\boldsymbol{\alpha} \in B_u(\boldsymbol{\alpha}_k)} h(\boldsymbol{x}, \ell, r|\boldsymbol{\alpha}).$$

Then, for any $h(\boldsymbol{x}, \ell, r|\boldsymbol{\alpha}) \in \mathcal{F}_n$, there exists a $k = 1, \ldots, N(u)$ such that $\boldsymbol{\alpha} \in B_u(\boldsymbol{\alpha}_k)$ and hence $\underline{M}_k(\boldsymbol{x}, \ell, r|u) \leq h(\boldsymbol{x}, \ell, r|\boldsymbol{\alpha}) \leq \overline{M}_k(\boldsymbol{x}, \ell, r|u)$. Similarly as in the proof of Lemma S.1.1 in Kreiss and Van Keilegom (2023), it can now be shown that there is a $0 < \tilde{C} < \infty$ such that

$$H(\underline{M}_k, \overline{M}_k) \leq \tilde{C}(m_n + \tilde{m}_n)^{1/2} u^{1/2}.$$

Let $\varepsilon > 0$. Then, if $u = \varepsilon^2/(\tilde{C}^2(m_n + \tilde{m}_n))$, we have that $H(\underline{M}_k, \overline{M}_k) \leq \varepsilon$. Finally,

$$\begin{aligned} N_{[\,]}(\varepsilon, \mathcal{F}_n, H) &\leq N\Big(\frac{\varepsilon^2}{\tilde{C}^2(m_n + \tilde{m}_n)}, A_n, \|\cdot\|\Big) \\ &\leq \Big(\frac{C\tilde{C}^2(m_n + \tilde{m}_n)}{\varepsilon^2}\Big)^{m_n+\tilde{m}_n+2p+2}, \end{aligned}$$

and hence,

$$\log N_{[\,]}(\varepsilon, \mathcal{F}_n, H) \leq C_1(m_n + \tilde{m}_n) \log\Big(\frac{C_2(m_n + \tilde{m}_n)}{\varepsilon^2}\Big)$$

for some $0 < C_1, C_2 < \infty$.

B.4. Proof of Theorem 4.1

In order to prove this theorem, we will check the conditions of Theorem B.1. By assumption, there is an $\alpha \in (0, 1/2)$ such that $\delta_n(\alpha) \to 0$. So $\delta_n(\alpha) < 1/\alpha$ for n sufficiently large. We show now that condition (28) is satisfied. Let $b_n = 2C_1(m_n + \tilde{m}_n)$ and $c_n = \sqrt{C_2(m_n + \tilde{m}_n)}$. Also, let $\varepsilon_n = \sqrt{\varphi_n/n}$, $\ell_n = \varepsilon_n^2/2^8$ and $u_n = \sqrt{2}\varepsilon_n$. Then, using Lemma B.2, we get

$$\begin{aligned} &\frac{1}{\sqrt{n}\varepsilon_n^2} \int_{\ell_n}^{u_n} \sqrt{\log N_{[\,]}\Big(\frac{u}{c_1}, \mathcal{F}_n, H\Big)} du \\ &\leq \frac{1}{\sqrt{n}\varepsilon_n^2} \int_{\ell_n}^{u_n} \sqrt{b_n \log\Big(\frac{c_n}{u}\Big)} du \\ &\leq \frac{1}{\sqrt{n}\varepsilon_n^2} \Big(u_n \sqrt{b_n \log\Big(\frac{c_n}{u_n}\Big)} + \sqrt{b_n} c_n \int_{\sqrt{\log(c_n/u_n)}}^{\sqrt{\log(c_n/\ell_n)}} e^{-t^2} dt\Big) \end{aligned}$$

$$\leq \frac{1}{\sqrt{n}\varepsilon_n^2}\left(u_n\sqrt{b_n\log\left(\frac{c_n}{u_n}\right)} + \sqrt{b_n}c_n\int_{\sqrt{\log(c_n/u_n)}}^{\infty} e^{-t^2}dt\right)$$

$$\leq \frac{1}{\sqrt{n}\varepsilon_n^2}\left(u_n\sqrt{b_n\log\left(\frac{c_n}{u_n}\right)} + \sqrt{\frac{\pi}{2}}\sqrt{b_n}\varepsilon_n\right)$$

$$= \sqrt{\frac{b_n}{\varphi_n}\log\left(\frac{nc_n^2}{2\varphi_n}\right)} + \sqrt{\frac{\pi b_n}{2\varphi_n}}, \tag{32}$$

where for the second inequality we used twice that

$$\int_0^v \sqrt{b\log\left(\frac{c}{u}\right)}du = v\sqrt{b\log\left(\frac{c}{v}\right)} - \sqrt{b}c\int_0^{\sqrt{\log(c/v)}} e^{-t^2}dt,$$

with $v = u_n$ and $v = \ell_n$, and the last inequality follows since for any x, $\int_x^\infty e^{-t^2}dt \leq \frac{\sqrt{\pi}}{2}e^{-x^2}$. It now follows from (21) and from the fact that $nc_n^2/\varphi_n \to \infty$ (which is guaranteed since $c_n \to \infty$ and $\varphi_n/n \to 0$), that (32) is bounded. Hence the conditions of Theorem B.1 are satisfied, from which the result follows.

B.5. Proof of Corollary 4.1

Let $\hat{F}_n(\cdot|\boldsymbol{x})$ be the distribution function corresponding to the density in (15), but with $\boldsymbol{\alpha}_n = \hat{\boldsymbol{\alpha}}_n$, and let the estimated τ-th quantile of T given $\boldsymbol{X} = \boldsymbol{x}$ be denoted by $\hat{q}_n(\tau|\boldsymbol{x}) = \inf\{t : \hat{F}_n(t|\boldsymbol{x}) \geq \tau\} = \hat{\boldsymbol{\beta}}_n^t\boldsymbol{x}$.

To prove that $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = o_P(1)$, we will show the following two facts:

(P1) The function $B^2 \to \mathbb{R}^+ : (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \to d(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sup_{\boldsymbol{x}\in\mathcal{X}} |(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)^t\boldsymbol{x}|$ is a metric

(P2) $\sup_{\boldsymbol{x}\in\mathcal{X}} |\hat{q}_n(\tau|\boldsymbol{x}) - q(\tau|\boldsymbol{x})| = o_P(1)$

Then, to show that $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = o_P(1)$ it suffices by (P1) to show that $d(\hat{\boldsymbol{\beta}}_n, \boldsymbol{\beta}) = o_P(1)$, and this holds thanks to (P2).

We start by showing (P1). The function d is clearly symmetric and it satisfies the triangle inequality. So it suffices to show that $d(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = 0$ if and only if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Indeed, if $d(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = 0$, then $0 = \mathrm{Var}((\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)^t\boldsymbol{X}) = (\tilde{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_2)^t\mathrm{Var}(\tilde{\boldsymbol{X}})(\tilde{\boldsymbol{\beta}}_1 - \tilde{\boldsymbol{\beta}}_2)$, where $\boldsymbol{X} = (1, \tilde{\boldsymbol{X}})$ and $\boldsymbol{\beta} = (\beta_0, \tilde{\boldsymbol{\beta}})$ for any $\boldsymbol{\beta} \in B$. Since $\mathrm{Var}(\tilde{\boldsymbol{X}})$ is positive definite, it follows that $\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_2$, and hence $0 = d(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = |\beta_{1,0} - \beta_{2,0}|$, which shows that $\beta_{1,0} = \beta_{2,0}$.

To show property (P2), we start by showing that there exists a small $\xi > 0$ such that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{|q(\tau|\boldsymbol{x})-t|<\xi} |\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})| = o_P(1).$$

Let $\xi_R = \inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{|q(\tau|\boldsymbol{x})-r|<\xi} f_{R|\boldsymbol{X}}(r|\boldsymbol{x})$ and similarly for ξ_L. We know from assumption (A2) that $\max(\xi_R, \xi_L) > 0$, and assume without loss of generality that $\xi_R > 0$. Then, since $\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})$ is uniformly continuous in t, we have for ξ sufficiently small and for all $\boldsymbol{x} \in \mathcal{X}$,

$$\sup_{|q(\tau|\boldsymbol{x})-t|<\xi} |\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})|$$

$$
\begin{aligned}
&\le 2 \int_{q(\tau|\boldsymbol{x})-\xi}^{q(\tau|\boldsymbol{x})+\xi} |\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})| dt \\
&\le \frac{2}{\xi_R} \int_{q(\tau|\boldsymbol{x})-\xi}^{q(\tau|\boldsymbol{x})+\xi} |\hat{F}_n(r|\boldsymbol{x}) - F(r|\boldsymbol{x})| f_{R|\boldsymbol{X}}(r|\boldsymbol{x}) dr \\
&\le \frac{2}{\xi_R} \int\int |\hat{F}_n(r|\boldsymbol{x}) - F(r|\boldsymbol{x})| f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) d\ell dr \\
&= \frac{2}{\xi_R} E_{L,R|\boldsymbol{X}}\big(\big|\hat{F}_n(R|\boldsymbol{x}) - F(R|\boldsymbol{x})\big|\big) \\
&= \frac{2}{\xi_R} \sum_{k=1}^{k_{\max}} P(K=k) E\Big\{\sum_{j=1}^{k} \big|\hat{F}_n(Y_{k,j}|\boldsymbol{x}) - F(Y_{k,j}|\boldsymbol{x})\big|\big(F(Y_{k,j}|\boldsymbol{x}) - F(Y_{k,j-1}|\boldsymbol{x})\big)\Big\} \\
&\le \frac{2}{\xi_R} \sum_{k=1}^{k_{\max}} P(K=k) E\Big\{\sum_{j=1}^{k}\sum_{\ell=1}^{j} \Big|\hat{F}_n(Y_{k,\ell}|\boldsymbol{x}) - F(Y_{k,\ell}|\boldsymbol{x}) \\
&\quad - \hat{F}_n(Y_{k,\ell-1}|\boldsymbol{x}) + F(Y_{k,\ell-1}|\boldsymbol{x})\Big|\Big(F(Y_{k,j}|\boldsymbol{x}) - F(Y_{k,j-1}|\boldsymbol{x})\Big)\Big\} \\
&\le \frac{2k_{\max}}{\xi_R c_L} \sum_{k=1}^{k_{\max}} P(K=k) E\Big\{\sum_{\ell=1}^{k} \Big|\hat{F}_n(Y_{k,\ell}|\boldsymbol{x}) - F(Y_{k,\ell}|\boldsymbol{x}) \\
&\quad - \hat{F}_n(Y_{k,\ell-1}|\boldsymbol{x}) + F(Y_{k,\ell-1}|\boldsymbol{x})\Big|\Big(F(Y_{k,\ell}|\boldsymbol{x}) - F(Y_{k,\ell-1}|\boldsymbol{x})\Big)\Big\} + o(1) \\
&= \frac{2k_{\max}}{\xi_R c_L} E\big|\hat{F}_n(R|\boldsymbol{x}) - F(R|\boldsymbol{x}) - \hat{F}_n(L|\boldsymbol{x}) + F(L|\boldsymbol{x})\big| + o(1)
\end{aligned}
$$

thanks to assumption (A4). The latter expression is bounded by (where $\xi > 0$ is chosen small enough)

$$
\begin{aligned}
&\frac{4k_{\max}}{\xi_R c_L} \int_{\boldsymbol{x}-\xi}^{\boldsymbol{x}+\xi} \int\int \big|\hat{F}_n(r|\boldsymbol{x}) - F(r|\boldsymbol{x}) - \hat{F}_n(\ell|\boldsymbol{x}) + F(\ell|\boldsymbol{x})\big| f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) d\ell dr dx + o(1) \\
&\le \frac{4k_{\max}}{\xi_R c_L} \int_{\mathcal{X}} \int\int \big|\hat{F}_n(r|\boldsymbol{x}) - F(r|\boldsymbol{x}) - \hat{F}_n(\ell|\boldsymbol{x}) + F(\ell|\boldsymbol{x})\big| f_{L,R|\boldsymbol{X}}(\ell, r|\boldsymbol{x}) d\ell dr dx + o(1) \\
&\le \frac{4k_{\max}}{\xi_R c_L \inf_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x})} \int\int\int \Big(\sqrt{\hat{h}_n(\boldsymbol{x}, \ell, r)} - \sqrt{h(\boldsymbol{x}, \ell, r)}\Big)^2 d\ell dr dx + o(1) \\
&= \frac{4k_{\max}}{\xi_R c_L \inf_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x})} H^2(\hat{h}_n, h) + o(1) \\
&= o_P(1),
\end{aligned}
$$

thanks to Theorem 4.1. Since the upper bound does not depend on $\boldsymbol{x}$, it follows that $\sup_{|q(\tau|\boldsymbol{x})-t|<\xi} |\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})| = o_P(1)$ uniformly in $\boldsymbol{x}$. Finally, to show that property (P2) holds, note that it can be easily seen that

$$
\sup_{\boldsymbol{x}\in\mathcal{X}} |\hat{q}_n(\tau|\boldsymbol{x}) - q(\tau|\boldsymbol{x})| \le \frac{1}{\delta_T} \sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{|q(\tau|\boldsymbol{x})-t|<\delta} |\hat{F}_n(t|\boldsymbol{x}) - F(t|\boldsymbol{x})|,
$$

where $\delta_T = \inf_{\boldsymbol{x}\in\mathcal{X}} \inf_{|q(\tau|\boldsymbol{x})-t|<\delta} f(t|\boldsymbol{x})$, and this is $o_P(1)$, thanks to the above calculations and assumption (A2).

Acknowledgments

The authors like to thank Anouar El Ghouch for very useful discussions and interactions that improved the paper substantially. Computational resources have been provided by the supercomputing facilities of UCLouvain (CISM/UCL) and the 'Consortium des Equipements de Calcul Intensif en Fédération Wallonie Bruxelles' (CECI) funded by the F.R.S.-FNRS under convention 2.5020.11 and by the Walloon Region.

References

Bottai, M. and Zhang, J. (2010). Laplace regression with censored data. *Biometrical Journal* **52** 487-503.

Chen, D.-G. D., Sun, J. and Peace, K. E. (2012). *Interval-censored time-to-event data: methods and applications.* CRC Press.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **4** 831–853.

Florens, J.-P., Simar, L. and Van Keilegom, I. (2020). Estimation of the Boundary of a Variable Observed With Symmetric Error. *Journal of the American Statistical Association* **115** 425-441.

Frumento, P. (2023). A quantile regression estimator for interval-censored data. *The International Journal of Biostatistics* **19** 81–96.

Frumento, P. and Bottai, M. (2017). Parametric modeling of quantile regression coefficient functions with censored and truncated data. *Biometrics* **73** 1179-1188.

Geerdens, C., Claeskens, G. and Janssen, P. (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics* **14** 433–446.

Geskus, R. and Groeneboom, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *The Annals of Statistics* **27** 627–674.

Johnson, S. G. (2007). The NLopt nonlinear-optimization package. https://github.com/stevengj/nlopt.

Kim, Y.-J., Cho, H., Kim, J. and Jhun, M. (2010). Median Regression Model with Interval Censored Data. *Biometrical Journal* **52** 201-208.

Koenker, R. (2005). *Quantile Regression. Econometric Society Monographs.* Cambridge University Press.

Koenker, R. (2011). A note on Laplace regression with censored data. *Biometric Journal* **53** 855–860.

Koenker, R. (2017). Quantile Regression: 40 Years On. *Annual Review of Economics* **9** 155-176.

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

Koenker, R. and Machado, J. A. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* **94** 1296-1310.

Kotz, S., Kozubowski, T. and Podgorski, K. (2012). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance.* Springer Science & Business Media.

Kreiss, A. and Van Keilegom, I. (2022). Semi-parametric estimation of incubation and generation times by means of Laguerre polynomials. *Journal of Nonparametric Statistics* **34** 570-606.

Kreiss, A. and Van Keilegom, I. (2023). Efficient quantile regression under censoring using Laguerre polynomials. *Bernoulli (under revision).*

Nikiforov, A. F. and Uvarov, V. B. (1988). *Special functions of mathematical physics* **205**. Springer.

Oller, R., Gomez, G. and Calle, M. L. (2004). Interval censoring: Model characterizations for the validity of the simplified likelihood. *Canadian Journal of Statistics* **32** 315-326.

Peng, L. (2021). Quantile Regression for Survival Data. *Annual Review of Statistics and Its Application* **8** 413-437.

Powell, M. J. D. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis*, (S. Gomez and J. P. Hennart, eds.). *Mathematics and Its Applications* **275** 51–67. Springer.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with Mixed Case Interval-Censored Data. *Scandinavian Journal of Statistics* **27** 45-55.

Shen, P.-S. (2013). Median regression model with left truncated and interval-censored data. *Journal of the Korean Statistical Society* **42** 469-479.

Sun, J. (2003). The Statistical Analysis of Interval-Censored Failure Time Data. *Handbook of Statistics* **23** 105-122.

Szego, G. (1939). *Orthogonal Polynomials. American Math. Soc: Colloquium publ* **vol. 23**. American Mathematical Society.

Wong, W. H. and Shen, X. (1995). Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve MLES. *The Annals of Statistics* **23** 339–362.

Yu, K. and Zhang, J. (2005). A Three-Parameter Asymmetric Laplace Distribution and Its Extension. *Communications in Statistics - Theory and Methods* **34** 1867-1879.

Zhou, X., Feng, Y. and Du, X. (2017). Quantile regression for interval censored data. *Communications in Statistics - Theory and Methods* **46** 3848-3863.