Bernoulli **19**(3), 2013, 748–779
DOI: 10.3150/12-BEJ462

Nonparametric quantile regression for twice censored data

STANISLAV VOLGUSHEV[1,2,*] and HOLGER DETTE[2,**]

[1] *Department of Statistics, University of Illinois at Urbana-Champaign, USA*
[2] *Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.*
E-mail: [*] *stanislav.volgushev@ruhr-uni-bochum.de;* [**] *holger.dette@ruhr-uni-bochum.de*

We consider the problem of nonparametric quantile regression for twice censored data. Two new estimates are presented, which are constructed by applying concepts of monotone rearrangements to estimates of the conditional distribution function. The proposed methods avoid the problem of crossing quantile curves. Weak uniform consistency and weak convergence is established for both estimates and their finite sample properties are investigated by means of a simulation study. As a by-product, we obtain a new result regarding the weak convergence of the Beran estimator for right censored data on the maximal possible domain, which is of its own interest.

Keywords: Beran estimator; censored data; crossing quantile curves; monotone rearrangements; quantile regression; survival analysis

1. Introduction

Quantile regression offers great flexibility in assessing covariate effects on event times. The method was introduced by [28] as a supplement to least squares methods focussing on the estimation of the conditional mean function and since this seminal work it has found numerous applications in different fields (see [26]). Recently, [29] have proposed quantile regression techniques as an alternative to the classical Cox model for analyzing survival times. These authors argued that quantile regression methods offer an interesting alternative, in particular if there is heteroscedasticity in the data or inhomogeneity in the population, which is a common phenomenon in survival analysis (see [39]). Unfortunately, the "classical" quantile regression techniques cannot be directly extended to survival analysis, because for the estimation of a quantile one has to estimate the censoring distribution for each observation. As a consequence, rather stringent assumptions are required in censored regression settings. Early work by [40,41], requires that the censoring times are always observed. Moreover, even under this rather restrictive and – in many cases – not realistic assumption the objective function is not convex, which results in some computational problems (see, e.g., [19]). Even worse, recent research indicates that using the information contained in the observed censored data actually reduces the estimation accuracy (see [27]).

Because in most survival settings the information regarding the censoring times is incomplete several authors have tried to address this problem by making restrictive assumptions on the censoring mechanism. For example, [48] assumed that the responses and censoring times are independent, which is stronger than the usual assumption of conditional independence. Reference [47] proposed a method for median regression under the assumption of i.i.d. errors, which

1350-7265

is computationally difficult to evaluate and cannot be directly generalized to the heteroscedastic case. Recently, [39] suggested a recursively re-weighted quantile regression estimate under the assumption that the censoring times and responses are independent conditionally on the predictor. This estimate adopts the principle of self consistency for the Kaplan–Meier statistic (see [17]) and can be considered as a direct generalization of this classical estimate in survival analysis. Reference [37] pointed out that the large sample properties of this recursively defined estimate are still not completely understood and proposed an alternative approach, which is based on martingale estimating equations. In particular, they proved consistency and asymptotic normality of their estimate.

While all of the cited literature considers the classical linear quantile regression model with right censoring, less results are available for quantile regression in a nonparametric context. Some results on nonparametric quantile regression when no censoring is present can be found in [8] and [49,50]. References [9] and [16] pointed out that many of the commonly proposed parametric or nonparametric estimates lead to possibly crossing quantile curves and modified some of these estimates to avoid this problem. Results regarding the estimation of the conditional distribution function from right censored data can be found in [10,11] or [31]. The estimation of conditional quantile functions in the same setting is briefly stressed in [10] and further elaborated in [12], while [18] proposed a quantile regression procedure for right censored and dependent data. On the other hand, the problem of nonparametric quantile regression for censored data where the observations can be censored from either left or right does not seem to have been considered in the literature.

This gap can partially be explained by the difficulties arising in the estimation of the conditional distribution function with two-sided censored data. The problem of estimating the (unconditional) distribution function for data that may be censored from above and below has been considered by several authors. For an early reference, see [43]. More recent references are [6, 7,24] and [36]. On the other hand- to their best knowledge- the authors are not aware of literature on nonparametric conditional quantile regression, or estimation of a conditional distribution function, for left and right censored data when the censoring is not always observed and only the conditional independence of censoring and lifetime variables is assumed.

In the present paper, we consider the problem of nonparametric quantile regression for twice censored data. We consider a censoring mechanism introduced by [36] and propose an estimate of the conditional distribution function in several steps. On the basis of this estimate and the preliminary statistics which are used for its definition, we construct two quantile regression estimates using the concept of simultaneous inversion and isotonization (see [14]) and monotone rearrangements (see [15], [9] or [3] among others). In Section 2, we introduce the model and the two estimates, while Section 3 contains our main results. In particular, we prove uniform consistency and weak convergence of the estimates of the conditional distribution function and its quantile function. As a by-product, we obtain a new result on the weak convergence of the Beran estimator on the maximal possible interval, which is of independent interest. In Section 4, we illustrate the finite sample properties of the proposed estimates by means of a simulation study. Finally, all proofs and technical details are deferred to the Appendix.

2. Model and estimates

We consider independent identically distributed random vectors (T_i, L_i, R_i, X_i), $i = 1, \ldots, n$, where T_i are the variables of interest, L_i and R_i are left and right censoring variables, respectively, and the $\mathbb{R}^d$-valued random variables X_i denote the covariates. We assume that the distributions of the random variables L_i, R_i and T_i depend on X_i and denote by $F_L(t|x) := P(L \leq t|X = x)$ the conditional distribution function of L given $X = x$. The conditional distribution functions $F_R(\cdot|x)$ and $F_T(\cdot|x)$ are defined analogously.

Additionally, we assume that the random variables T_i, L_i, R_i are almost surely nonnegative and independent conditionally on the covariate X_i. Our aim is to estimate the conditional quantile function $F_T^{-1}(\cdot|x)$. However, due to the censoring, we can only observe the triples (Y_i, X_i, δ_i) where $Y_i = \max(\min(T_i, R_i), L_i)$ and the indicator variables δ_i are defined by

$$\delta_i := \begin{cases} 0, & L_i < T_i \leq R_i, \\ 1, & L_i < R_i < T_i, \\ 2, & T_i \leq L_i < R_i \text{ or } R_i \leq L_i. \end{cases} \tag{2.1}$$

Remark 2.1. An unconditional version of this censoring mechanism was introduced by [36]. Examples of situations where this kinds of data occur can, for example, be found in chapter 15 of [34]. This model also is closely related to the double censoring model, see [43] for the case without covariates. In that setting, the assumption of independence between the random variables L, R, T is replaced by the assumption that T is independent of the pair (R, L) and additionally $P(L < R) = 1$. Note that none of the two assumptions is strictly more or less restrictive than the other. Rather the two models describe different situations. Moreover, since L, T, R are never observed simultaneously, it is not possible to test based on the data which of the models is most appropriate. Instead, an understanding of the underlying data generation process is crucial to identify the right model. A more detailed comparison of the two models can be found in [35] and [36] for the case without covariates.

Roughly speaking, the construction of an estimate for the conditional quantile function of T can be accomplished in three steps. First, we define the variables $S_i := \min(T_i, R_i)$ and consider the model $Y_i = \max(S_i, L_i)$, which is a classical right censoring model. In this model, we estimate the conditional distribution $F_L(\cdot|x)$ of L. In a second step, we use this information to reconstruct the conditional distribution of T (see Section 2.1). Finally, the concept of simultaneous isotonization and inversion (see [14]) and the monotone rearrangements, which was recently introduced by [15] in the context of monotone estimation of a regression function, are used to obtain two estimates of the conditional quantile function (see Section 2.2).

2.1. Estimation of the conditional distribution function

To be more precise, let H denote the conditional distribution of Y. We introduce the notation $H_k(A|x) = \mathrm{P}(A \cap \{\delta = k\}|X = x)$ and obtain the decomposition $H = H_0 + H_1 + H_2$ for the conditional distribution of Y_i. The sub-distribution functions H_k $(k = 0, 1, 2)$ can be represented as follows

$$H_0(\mathrm{d}t|x) = F_L(t-|x)\big(1 - F_R(t-|x)\big)F_T(\mathrm{d}t|x), \tag{2.2}$$

$$H_1(\mathrm{d}t|x) = F_L(t-|x)\big(1 - F_T(t|x)\big)F_R(\mathrm{d}t|x), \tag{2.3}$$

$$H_2(\mathrm{d}t|x) = \big\{1 - \big(1 - F_T(t|x)\big)\big(1 - F_R(t|x)\big)\big\}F_L(\mathrm{d}t|x) = F_S(t|x)F_L(\mathrm{d}t|x). \tag{2.4}$$

Note that the conditional (sub-)distribution functions H_k and H can easily be estimated from the observed data by

$$H_{k,n}(t|x) := \sum_{i=1}^{n} W_i(x) I_{\{Y_i \le t, \delta_i = k\}}, \qquad H_n(t|x) := \sum_{i=1}^{n} W_i(x) I_{\{Y_i \le t\}}, \tag{2.5}$$

where the quantities $W_i(x)$ denote local weights depending on the covariates $X_1, \dots, X_n$, which will be specified below. We will use the representations (2.2)–(2.4) to obtain an expression for F_T in terms of the functions H, H_k and then replace the distribution functions H, H_k by their empirical counterparts $H_n, H_{k,n}$, respectively. We begin with the reconstruction of F_L. First, note that

$$M_2^-(\mathrm{d}t|x) := \frac{H_2(\mathrm{d}t|x)}{H(t|x)} = \frac{F_S(t|x)F_L(\mathrm{d}t|x)}{F_L(t|x)F_S(t|x)} = \frac{F_L(\mathrm{d}t|x)}{F_L(t|x)} \tag{2.6}$$

is the predictable reverse hazard measure corresponding to F_L and hence we can reconstruct F_L using the product-limit representation

$$F_L(t|x) = \prod_{(t,\infty]} \big(1 - M_2^-(\mathrm{d}s|x)\big) \tag{2.7}$$

(see, e.g., [36]). Now having a representation for the conditional distribution function F_L we can define in a second step

$$\begin{aligned}
\Lambda_T^-(\mathrm{d}t|x) &:= \frac{H_0(\mathrm{d}t|x)}{F_L(t-|x) - H(t-|x)} = \frac{H_0(\mathrm{d}t|x)}{F_L(t-|x)(1 - F_S(t-|x))} \\
&= \frac{H_0(\mathrm{d}t|x)}{F_L(t-|x)(1 - F_R(t-|x))(1 - F_T(t-|x))} \\
&= \frac{F_L(t-|x)(1 - F_R(t-|x))F_T(\mathrm{d}t|x)}{F_L(t-|x)(1 - F_R(t-|x))(1 - F_T(t-|x))} = \frac{F_T(\mathrm{d}t|x)}{1 - F_T(t-|x)},
\end{aligned} \tag{2.8}$$

which yields an expression for the predictable hazard measure of F_T. Finally, F_T can be reconstructed by using the product-limit representation

$$1 - F_T(t|x) = \prod_{[0,t]} \big(1 - \Lambda_T^-(\mathrm{d}s|x)\big) \tag{2.9}$$

(see, e.g., [23]). Note that formula (2.9) yields an explicit representation of the conditional distribution function $F_T(\cdot|x)$ in terms of the quantities H_0, H_1, H_2, H, which can be estimated from the data [see equation (2.5)]. The estimate of the conditional distribution function is now defined as follows. First, we use the representation (2.7) to obtain an estimate of $F_L(\cdot|x)$, that is,

$$F_{L,n}(t|x) = \prod_{(t,\infty]} \bigl(1 - M_{2,n}^{-}(\mathrm{d}s|x)\bigr), \tag{2.10}$$

where

$$M_{2,n}^{-}(\mathrm{d}s|x) = \frac{H_{2,n}(\mathrm{d}s|x)}{H_n(s|x)}. \tag{2.11}$$

Second, after observing (2.8) and (2.9), we define

$$F_{T,n}(t|x) = 1 - \prod_{[0,t]} \bigl(1 - \Lambda_{T,n}^{-}(\mathrm{d}s|x)\bigr), \tag{2.12}$$

where

$$\Lambda_{T,n}^{-}(\mathrm{d}s|x) = \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)}. \tag{2.13}$$

In Section 3, we will analyse the asymptotic properties of these estimates, while in the following Section 2.2 these estimates are used to construct nonparametric and noncrossing quantile curve estimates.

Remark 2.2. Throughout this paper, we will adopt the convention '$0/0 = 0$.' This means that if, for example, $H_{0,n}(\mathrm{d}t|x) = 0$ and $F_{L,n}(t-|x) - H_n(t-|x) = 0$, the contribution of

$$\frac{H_{0,n}(\mathrm{d}t|x)}{F_{L,n}(t-|x) - H_n(t-|x)}$$

in (2.13) will be interpreted as zero.

2.2. Non-crossing quantile estimates by monotone rearrangements

In practice, nonparametric estimators of a conditional distribution function $F(\cdot|x)$ are not necessarily increasing for finite sample sizes (see, e.g., [50]). Although this problem often vanishes asymptotically, it still is of great practical relevance, because in a concrete application it is not completely obvious how to invert a non-increasing function. Trying to naively invert such estimators may lead to the well-known problem of quantile crossing (see [26] or [50]) which poses some difficulties in the interpretation of the results. In this paper, we will discuss the following two possibilities to deal with this problem:

1. Use a procedure developed by [16] which is based on a simultaneous isotononization and inversion of a nonincreasing distribution function. As a by-product, this method yields non-

crossing quantile estimates. To be precise, we consider the operator

$$\Psi : \begin{cases} L^\infty(J) \to L^\infty(\mathbb{R}), \\ f \mapsto \left(y \mapsto \int_J I_{\{f(u) \le y\}} \, \mathrm{d}u \right), \end{cases} \tag{2.14}$$

where $L^\infty(I)$ denotes the set of bounded, measurable functions on the set I and J denotes a bounded interval. Note that for a strictly increasing function f this operator yields the right continuous inverse of f, that is, $\Psi(f) = f^{-1}$ [here and in what follows, f^{-1} will denote the generalized inverse, i.e., $f^{-1}(t) := \sup\{s : f(s) \le t\}$]. On the other hand, $\Psi(f)$ is always isotone, even in the case where f does not have this property. Consequently, if $\hat{f}$ is a not necessarily isotone estimate of an isotone function f, the function $\Psi(\hat{f})$ could be regarded as an isotone estimate of the function f^{-1}. Therefore, the first idea to construct an estimate of the conditional quantile function consists in the application of the operator Ψ to the estimate $F_{T,n}$ defined in (2.12), that is,

$$\hat{q}(\tau|x) = \Psi\big(F_{T,n}(\cdot|x)\big)(\tau). \tag{2.15}$$

However, note that formally the mapping Ψ operates on functions defined on bounded intervals. More care is necessary if the operator has to be applied to a function with an unbounded support. A detailed discussion and a solution of this problem can be found in [16]. In the present paper, we use different approach which is a slightly modified version of the ideas from [3]. To be precise, note that estimators of the conditional distribution function $F(\cdot|x)$ [in particular those of the form (2.5), which will be used later] often are constant outside of the compact interval $J := [j_1, j_2] = [\min_i Y_i, \max_i Y_i]$. Now the structure of the estimator $F_{T,n}(\cdot|x)$ implies that $F_{T,n}(\cdot|x)$ will also be constant outside of J. We thus propose to consider the modified operator $\tilde{\Psi}_J$ defined as

$$\tilde{\Psi}_J : \begin{cases} L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R}), \\ f \mapsto \left(y \mapsto j_1 + \int_J I_{\{f(u) \le y\}} \, \mathrm{d}u \right). \end{cases} \tag{2.16}$$

Consequently, the first estimator of the conditional quantile function is given by

$$\hat{q}(\tau|x) = \tilde{\Psi}_J\big(F_{T,n}(\cdot|x)\big)(\tau). \tag{2.17}$$

2. Use the concept of increasing rearrangements (see [15] and [9] for details) to construct an increasing estimate of the conditional distribution function, which is then inverted in a second step. More precisely, we define the operator

$$\Phi : \begin{cases} L^\infty(J) \to L^\infty(\mathbb{R}), \\ f \mapsto \big(y \mapsto \big(\Psi f(\cdot)\big)^{-1}(y)\big), \end{cases} \tag{2.18}$$

where Ψ is introduced in (2.14). Note that for a strictly increasing right continuous function f this operator reproduces f, that is, $\Phi(f) = f$. On the other hand, if f is not isotone,

$\Phi(f)$ is an isotone function and the operator preserves the L^p-norm, that is,

$$\int_J \left|\Phi\big(f(u)\big)\right|^p \mathrm{d}u = \int_J \left|f(u)\right|^p \mathrm{d}u.$$

Moreover, the operator also defines a contraction, that is,

$$\int_J \left|\Phi(f_1)(u) - \Phi(f_2)(u)\right|^p \mathrm{d}u \le \int_J |f_1 - f_2|^2 \,\mathrm{d}u \qquad \forall p \ge 1$$

(see [25] or [32]). This means if $\hat{f}(=f_1)$ is a not necessarily isotone estimate of the isotone function $f(=f_2)$, then the isotonized estimate $\Phi(\hat{f})$ is a better approximation of the isotone function f than the original estimate $\hat{f}$ with respect to any L^p-norm [note that $\Phi(f) = f$ because f is assumed to be isotone]. For a general discussion of monotone rearrangements and the operators (2.14) and (2.18), we refer to [4], while some statistical applications can be found in [15] and [9].

The idea is now to use rearranged estimators of $H_i(\cdot|x)$ and $H(\cdot|x)$ in the representations (2.6)–(2.9). For this purpose, we need to modify the operator Φ so that it can be applied to functions of unbounded support. We propose to proceed as follows:

- Define the operator $\tilde{\Phi}_J$ indexed by the compact interval $J = [j_1, j_2]$ as

$$\tilde{\Phi}_J : \begin{cases} L^\infty(\mathbb{R}) \to L^\infty(\mathbb{R}), \\ f \mapsto \big(y \mapsto I_{\{y<j_1\}} f(j_1-) \\ \qquad + \big(\tilde{\Psi}_J f(\cdot)\big)^{-1}(y) I_{\{j_1 \le y \le j_2\}} + I_{\{y>j_2\}} f(j_2)\big). \end{cases} \tag{2.19}$$

- Truncate the estimator $H_n(\cdot|x)$ for values outside of the interval [0, 1], that is,

$$\tilde{H}_n(t|x) := H_n(t|x) I_{\{H_n(t|x)\in[0,1]\}} + I_{\{H_n(t|x)>1\}}$$

  [note that in general estimators of the form (2.5) do not necessarily have values in the interval [0, 1] since the weights $W_i(x)$ might be negative].
- Use the statistic $H_n^{\mathrm{IP}}(t|x) := \tilde{\Phi}_{J_Y}(\tilde{H}_n(\cdot|x))(t)$ as estimator for $H(t|x)$.
- Observe that the estimator $H_n^{\mathrm{IP}}(t|x)$ is by construction an increasing step function which can only jump in the points $t = Y_i$, that is, it admits the representation

$$H_n^{\mathrm{IP}}(t|x) = \sum_i W_i^{\mathrm{IP}}(x) I_{\{Y_i \le t\}} \tag{2.20}$$

  with weights $W_i^{\mathrm{IP}}(x) \ge 0$. Based on this statistic, we define estimators $H_{k,n}^{\mathrm{IP}}$ of the subdistribution functions H_k as follows

$$H_{k,n}^{\mathrm{IP}}(t|x) = \sum_i W_i^{\mathrm{IP}}(x) I_{\{Y_i \le t\}} I_{\{\delta_i = k\}}, \qquad k = 0, 1, 2. \tag{2.21}$$

  In particular, such a definition ensures that $H^{\mathrm{IP}}(t|x) = H_{0,n}^{\mathrm{IP}}(t|x) + H_{1,n}^{\mathrm{IP}}(t|x) + H_{2,n}^{\mathrm{IP}}(t|x)$.

So far we have obtained increasing estimators of the quantities H and H_i. The next step in our construction is to plug these estimates in representation (2.6) to obtain:

$$\tilde{M}_{2,n}^{-}(\mathrm{d}t|x) = \frac{H_{2,n}^{\mathrm{IP}}(\mathrm{d}t|x)}{H_n^{\mathrm{IP}}(t|x)}, \tag{2.22}$$

which defines an increasing function with jumps of size less or equal to one. This implies that $\tilde{F}_{L,n}(t|x) = \prod_{(t,\infty]}(1 - \tilde{M}_{2,n}^{-}(\mathrm{d}s|x))$ is also increasing. For the rest of the construction, observe the following lemma which will be proved at the end of this section.

Lemma 2.3. *Assume that $Y_i \neq Y_j$ for $i \neq j$. Then the function*

$$\tilde{\Lambda}_{T,n}^{-}(\mathrm{d}t|x) := \frac{H_{0,n}^{\mathrm{IP}}(\mathrm{d}t|x)}{\tilde{F}_{L,n}(t-|x) - H_n^{\mathrm{IP}}(t-|x)} \tag{2.23}$$

is nonnegative, increasing and has jumps of size less or equal to one.

This in turn yields the estimate

$$F_{T,n}^{\mathrm{IP}}(t|x) = 1 - \prod_{[0,t]}\bigl(1 - \tilde{\Lambda}_{T,n}^{-}(\mathrm{d}s|x)\bigr). \tag{2.24}$$

In the final step, we now simply invert the resulting estimate of the conditional distribution function $F_{T,n}^{\mathrm{IP}}$ since it is increasing by construction. We denote this estimator of the conditional quantile function by

$$\hat{q}^{\mathrm{IP}}(t|x) := \sup\bigl\{s : F_{T,n}^{\mathrm{IP}}(s|x) \le t\bigr\}. \tag{2.25}$$

In the next section, we will discuss asymptotic properties of the two proposed estimates $\hat{q}$ and $\hat{q}^{\mathrm{IP}}$ of the conditional quantile curve.

Remark 2.4. In the classical right censoring case, there is no uniformly good way to define the Kaplan–Meier estimator beyond the largest uncensored observation (see, e.g., [20], page 105). Typical approaches include setting it to unity, to the value at the largest uncensored observation, or to consider it unobservable within certain bounds (for more details, see the discussion in [20], page 105 and [2], page 260). When censoring is light, the first of the above mentioned approaches seems to yield the best results (see [2], page 260).

When the data can be censored from either left or right, the situation becomes even more complicated since now we also have to find a reasonable definition below the smallest uncensored observation. From definitions (2.6)–(2.9), it is easy to see that $F_{T,n}$ equals zero below the smallest uncensored observation with non-vanishing weight and is constant at the largest uncensored observation and above. In practice, the latter implies that the estimators $\hat{q}(\tau|x)$ and $\hat{q}^{\mathrm{IP}}(\tau|x)$ are not defined as soon as $\sup_t F_{T,n}(t|x) < \tau$ or $\sup_t F_{T,n}^{\mathrm{IP}}(t|x) < \tau$, respectively. A simple ad-hoc solution to this problem is to define the estimator $F_{T,n}$ or $F_{T,n}^{\mathrm{IP}}$ as 1 beyond the last observation with non-vanishing weight or to locally increase the bandwidth. A detailed investigation of this problem is postponed to future research.

We conclude this section with the proof of Lemma 2.3.

Proof of Lemma 2.3. In order to see that $\tilde{\Lambda}^{-}_{T,n}(\mathrm{d}t|x)$ is increasing, we note that

$$
\begin{aligned}
H_n^{\mathrm{IP}}(t-|x) &= \prod_{[t,\infty)}\left(1-\frac{H_n^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}\right) \\
&= \prod_{[t,\infty)}\left(1-\frac{H_{2,n}^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}-\frac{H_{0,n}^{\mathrm{IP}}(\mathrm{d}s|x)+H_{1,n}^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}\right) \\
&\le \prod_{[t,\infty)}\left(1-\frac{H_{2,n}^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}\right)=\tilde{F}_{L,n}(t-|x).
\end{aligned}
$$

Thus, $\tilde{F}_{L,n}(t-|x)-H_n^{\mathrm{IP}}(t-|x)\ge 0$ and the nonnegativity of $\tilde{\Lambda}^{-}_{T,n}(\mathrm{d}t|x)$ is established. In order to prove the inequality $\tilde{\Lambda}^{-}_{T,n}(\mathrm{d}t|x)\le 1$, we assume without loss of generality that $Y_1<Y_2<\cdots<Y_n$. Observe that as soon as $\delta_k=0$ we have for $k\ge 2$

$$
\begin{aligned}
&\tilde{F}_{L,n}(Y_k-|x)-H_n^{\mathrm{IP}}(Y_k-|x) \\
&\quad=\left[1-\prod_{[Y_k,\infty)}\left(1-\frac{H_{0,n}^{\mathrm{IP}}(\mathrm{d}s|x)+H_{1,n}^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}\right)\right]\prod_{[Y_k,\infty)}\left(1-\frac{H_{2,n}^{\mathrm{IP}}(\mathrm{d}s|x)}{H_n^{\mathrm{IP}}(s|x)}\right) \\
&\quad\overset{(*)}{=}\left[1-\prod_{j\ge k,\delta_j\ne 2}\left(1-\frac{\Delta H_{0,n}^{\mathrm{IP}}(Y_j|x)+\Delta H_{1,n}^{\mathrm{IP}}(Y_j|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right)\right]\prod_{j\ge k+1,\delta_j=2}\left(1-\frac{\Delta H_{2,n}^{\mathrm{IP}}(Y_j|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right) \\
&\quad=\left[1-\prod_{j\ge k,\delta_j\ne 2}\left(\frac{H_n^{\mathrm{IP}}(Y_{j-1}|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right)\right]\prod_{j\ge k+1,\delta_j=2}\left(\frac{H_n^{\mathrm{IP}}(Y_{j-1}|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right) \\
&\quad\overset{(**)}{=}\left[1-\frac{H_n^{\mathrm{IP}}(Y_{k-1}|x)}{H_n^{\mathrm{IP}}(Y_k|x)}\prod_{j\ge k+1,\delta_j\ne 2}\left(\frac{H_n^{\mathrm{IP}}(Y_{j-1}|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right)\right]\prod_{j\ge k+1,\delta_j=2}\left(\frac{H_n^{\mathrm{IP}}(Y_{j-1}|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right) \\
&\quad\ge\left[1-\frac{H_n^{\mathrm{IP}}(Y_{k-1}|x)}{H_n^{\mathrm{IP}}(Y_k|x)}\right]\prod_{j\ge k+1}\left(\frac{H_n^{\mathrm{IP}}(Y_{j-1}|x)}{H_n^{\mathrm{IP}}(Y_j|x)}\right) \\
&\quad=\left[\frac{H_n^{\mathrm{IP}}(Y_k|x)-H_n^{\mathrm{IP}}(Y_{k-1}|x)}{H_n^{\mathrm{IP}}(Y_k|x)}\right]\frac{H_n^{\mathrm{IP}}(Y_k|x)}{H_n^{\mathrm{IP}}(Y_n|x)} \\
&\quad=\Delta H_n^{\mathrm{IP}}(Y_k|x),
\end{aligned}
$$

where the equalities $(*)$ and $(**)$ follow from $\delta_k=0$. An analogous result for $k=1$ follows by simple algebra. Hence, we have established that for $\delta_k=0$ we have $\Delta\tilde{\Lambda}^{-}_{T,n}(Y_k|x)\le 1$, and all the other cases need not be considered since we adopted the convention '$0/0=0$.' Thus, the proof is complete. □

3. Main results

The results stated in this section describe the asymptotic properties of the proposed estimators. In particular, we investigate weak convergence of the processes $\{H_{k,n}(t|x)\}_t$, $\{F_{T,n}(t|x)\}_t$, etc., where the predictor x is fixed. Our main results deal with the weak uniform consistency and the weak convergence of the process $\{F_{T,n}(t|x) - F_T(t|x)\}_t$ and the corresponding quantile processes obtained in Section 2. In order to derive the process convergence, we will assume that it holds for the initial estimates H_n, $H_{k,n}$ and give sufficient conditions for this property in Lemma 3.3. In a next step, we apply the delta method (see [22]) to the map $(H, H_2) \mapsto M_2^-$ defined in (2.6) and the product-limit maps defined in (2.7) and (2.9). Note that the product limit maps are Hadamard differentiable on the set of cadlag functions with total variation bounded by a constant (see Lemma A.1 on page 42 in [35]), and hence the process convergence of $M_{2,n}^-$ and $\Lambda_{T,n}^-$ will directly entail the weak convergence results for $F_{L,n}$ and $F_{T,n}$, respectively. However, the Hadamard differentiability of the map $(H_2, H) \mapsto M_2^-$ only holds on domains where $H(t) > \varepsilon > 0$, and hence more work is necessary to obtain the corresponding weak convergence results on the interval $[t_{00}, \infty]$ if $H(t_{00}|x) = 0$, where

$$t_{00} := \inf\bigl\{t : H_0(t|x) > 0\bigr\}. \tag{3.1}$$

This situation occurs, for example, if $F_R(t_{00}|x) = 0$, which is quite natural in the context considered in this paper because R is the right censoring variable.

For the sake of a clear representation and for later reference, we present all required technical conditions for the asymptotic results at the beginning of this section. We assume that the estimators of the conditional subdistribution functions are of the form (2.5) with weights $W_j(x)$ depending on the covariates $X_1, \dots, X_n$ but not on $Y_1, \dots, Y_n$ or $\delta_1, \dots, \delta_n$. The first set of conditions concerns the weights that are used in the representation (2.5). Throughout this paper, denote by $\|\cdot\|$ the maximum norm on $\mathbb{R}^d$.

(W1) With probability tending to one, the weights in (2.5) can be written in the form

$$W_i(x) = \frac{V_i(x)}{\sum_{j=1}^n V_j(x)},$$

where the real-valued functions V_j $(j = 1, \dots, n)$ have the following properties:

(1) There exist constants $0 < \underline{c} < \overline{c} < \infty$ such that for all $n \in \mathbb{N}$ and all x we have either $V_j(x) = 0$ or $\underline{c}/nh^d \le V_j(x) \le \overline{c}/nh^d$.
(2) If $\|x - X_j\| \le Ch$ for some constant $C < \infty$, then $V_j(x) \ne 0$ and $V_j(x) = 0$ for $\|x - X_j\| \ge c_n$ for some sequence $(c_n)_{n\in\mathbb{N}}$ such that $c_n = \mathrm{O}(h)$. Without loss of generality, we will assume that $C = 1$ throughout this paper.
(3) $\sum_i V_i(x) = C(x)(1 + \mathrm{o}_P(1))$ for some positive function C.
(4) $\sup_t \|\sum_i V_i(x)(x - X_i) I_{\{Y_i \le t\}}\| = \mathrm{o}_P(1/\sqrt{nh^d})$.

Here (and throughout this paper), h denotes a smoothing parameter converging to 0 with increasing sample size.

(W2) We assume that the weak convergence

$$\sqrt{nh^d}\big(H_{0,n}(\cdot|x) - H_0(\cdot|x), H_{2,n}(\cdot|x) - H_2(\cdot|x), H_n(\cdot|x) - H(\cdot|x)\big) \Rightarrow (G_0, G_2, G)$$

holds in $D^3[0,\infty]$, where the limit denotes a centered Gaussian process which has a version with a.s. continuous sample paths and a covariance structure of the form

$$\begin{aligned}\operatorname{Cov}\big(G_i(s|x), G_i(t|x)\big) &= b(x)\big(H_i(s\wedge t|x) - H_i(s|x)H_i(t|x)\big),\\ \operatorname{Cov}\big(G(s|x), G(t|x)\big) &= b(x)\big(H(s\wedge t|x) - H(s|x)H(t|x)\big),\\ \operatorname{Cov}\big(G_i(s|x), G(t|x)\big) &= b(x)\big(H_i(s\wedge t|x) - H_i(s|x)H(t|x)\big)\end{aligned}$$

for some function $b(x)$. Here and throughout this paper, weak convergence is understood as convergence with respect to the sigma algebra generated by the closed balls in the supremum norm (see [38]).

(W3) The estimators $H_{k,n}(\cdot|x)$ $(k = 0, 1, 2)$ and $H_n(\cdot|x)$ are weakly uniformly consistent on the interval $[0,\infty)$.

Remark 3.1. It will be shown in Lemma 3.3 below that, under suitable assumptions on the smoothing parameter h, important examples for weights satisfying conditions (W1)–(W3) are given by the Nadaraya–Watson weights

$$W_i^{\mathrm{NW}}(x) = \frac{(\prod_{k=1}^d K_h((x - X_i)_k))/(nh^d)}{(\sum_j \prod_{k=1}^d K_h((x - X_i)_k))/(nh^d)} =: \frac{V_i^{\mathrm{NW}}(x)}{\sum_j V_j^{\mathrm{NW}}(x)} \tag{3.2}$$

or (in one dimension) by the local linear weights

$$\begin{aligned}W_i^{\mathrm{LL}}(x) &= \frac{K_h(x - X_i)(S_{n,2} - (x - X_i)S_{n,1})/(nh)}{S_{n,2}S_{n,0} - S_{n,1}^2}\\ &= \frac{K_h(x - X_i)(1 - (x - X_i)S_{n,1}/S_{n,2})/(nh)}{(\sum_j K_h(x - X_j)(1 - (x - X_j)S_{n,1}/S_{n,2}))/(nh)} =: \frac{V_i^{\mathrm{LL}}(x)}{\sum_j V_j^{\mathrm{LL}}(x)},\end{aligned} \tag{3.3}$$

where $K_h(\cdot) := K(\cdot/h)$, $S_{n,k} := \frac{1}{nh}\sum_j K_h(x - X_j)(x - X_j)^k$ and the kernel satisfies the following condition.

(K1) The kernel K in (3.2) and (3.3) is a symmetric density of bounded total variation with compact support, say $[-1,1]$, which satisfies $c_1 \le K(x) \le c_2$ for all x with $K(x) \neq 0$ for some constants $0 < c_1 \le c_2 < \infty$.

For the distributions of the random variables (T_i, L_i, R_i, X_i), we assume that for some $\varepsilon > 0$ with $U_\varepsilon(x) := \{y : |y - x| < \varepsilon\}$:

(D1) The conditional distribution function F_R fulfills $F_R(t_{00}|x) < 1$.
(D2) For $i = 0, 1, 2$, we have $\lim_{y\to x}\sup_t |H_i(t|y) - H_i(t|x)| = 0$.

(D3) The conditional distribution functions $F_L(\cdot|x)$, $F_R(\cdot|x)$, $F_T(\cdot|x)$ have densities, say $f_L(\cdot|x)$, $f_R(\cdot|x)$, $f_T(\cdot|x)$, with respect to the Lebesque measure.
(D4) $\int_{t_{00}}^{\infty} \frac{f_L(u|x)}{F_L^2(u|x)F_S(u|x)} \,\mathrm{d}u < \infty$.
(D5) $\sup_{k=1,\dots,d} \int_{t_{00}}^{\infty} \frac{1}{F_L(u|x)F_S(u|x)} |\partial_{x_k} \frac{f_L(u|x)}{F_L(u|x)}| \,\mathrm{d}u < \infty$.
(D6) $\sup_{k,j=1,\dots,d} \sup_{(t,z)\in(t_{00},\infty)\times U_\varepsilon(x)} |\partial_{z_k}\partial_{z_j} \frac{f_L(t|z)}{F_L(t|z)}| < \infty$.
(D7) The functions $H_k(t|x)$ $(k = 0, 1, 2)$ are twice continuously differentiable with respect to the second component in some neighborhood $U_\varepsilon(x)$ of x and for $k = 0, 1, 2$ we have

$$\sup_{k,j=1,\dots,d} \sup_{t} \sup_{|y-x|<\varepsilon} \left|\partial_{y_k}\partial_{y_j} H_k(t|y)\right| < \infty.$$

(D8) The distribution function F_X of the covariates X_i is twice continuously differentiable in $U_\varepsilon(x)$. Moreover, F_X has a uniformly continuous density f_X such with $f_X(x) \neq 0$.
(D9) There exists a constant $C > 0$ such that $H(t|y) \geq CH(t|x)$ for all $(t, y) \in [t_{00}, t_{00} + \varepsilon) \times I$ where I is a set with the property $\int_{I\cap U_\delta(x)} f_X(s)\,\mathrm{d}s \geq c\delta^d$ for some $c > 0$ and all $0 < \delta \leq \varepsilon$.
(D10) $\frac{f_L(t|y)}{F_L(t|y)} = \frac{f_L(t|x)}{F_L(t|x)}(1 + \mathrm{o}(1))$ uniformly in $t \in [t_{00}, t_{00} + \varepsilon)$ as $y \to x$.
(D11) For $\tau_{T,0}(x) := \inf\{t : F_T(t|x) > 0\}$, we have $\inf_{y\in U_\varepsilon(x)} F_L(\tau_{T,0}(y)|y) > 0$.

Remark 3.2. From the definition of t_{00} and H_0, we immediately see that under condition (D1) we have $t_{00} = \tau_{T,0}(x) \vee \tau_{L,0}(x)$ where we use the notation $\tau_{L,0}(x) := \inf\{t : F_L(t|x) > 0\}$. In particular, this implies that under either of the assumptions (D4) or (D11) the equality $t_{00} = \tau_{T,0}(x)$ holds.

Finally, we make some assumptions for the smoothing parameter:

(B1) $nh^{d+4} \log n = \mathrm{o}(1)$ and $nh \longrightarrow \infty$.
(B2) $h \to 0$ and $nh^d/\log n \longrightarrow \infty$.

Some important practical examples for weights satisfying conditions (W1)–(W3) include Nadaraya–Watson and local linear weights. This is the assertion of the next lemma.

Lemma 3.3.

1. *Conditions (W1)(1) and (W1)(2) are fulfilled for the Nadaraya–Watson weights W_i^{NW} with a Kernel K satisfying condition* (K1). *If the density f_X is continuous at the point x, condition (W1)(3) also holds. Finally, if the function $x \mapsto f_X(x)F_Y(t|x)$ is continuously differentiable in a neighborhood of x for every t with uniformly* (*in t*) *bounded first derivative and (B1) is fulfilled, condition (W1)(4) holds.*

   If additionally to these assumptions $d = 1$ and the density f_X of the covariates X is continuously differentiable at x with bounded derivative, condition (W1) also holds for the local linear and rearranged local linear weights W_i^{LL} and W_i^{LLI} defined in (3.3) *and* (2.20), (2.21), *respectively, provided that the corresponding kernel fulfills condition* (K1).
2. *If under assumptions (D7), (D8) and (B1) the density f_X is twice continuously differentiable with uniformly bounded derivative, condition (W2) holds for the Nadaraya–Watson*

(d arbitrary), local linear ($d = 1$) or rearranged local linear ($d = 1$) weights based on a positive, symmetric kernel with compact support.

3. *If under assumptions (B2), (D2), (D3) the density f_X is twice continuously differentiable with uniformly bounded derivative, condition (W3) holds for the Nadaraya–Watson weights W_i based on a positive, symmetric kernel with compact support (d arbitrary). If additionally $d = 1$ and the density f_X of the covariates X is continuously differentiable at x with bounded derivative, condition (W3) also holds for local linear or rearranged local linear weights.*

The proof of this lemma is standard but tedious. Because of space considerations, we do not include it in the present paper, a detailed proof can be found in the technical report [45].

Note that the assumption (B1) does not allow to choose $h \sim n^{-1/(d+4)}$, which would be the MSE-optimal rate for Nadaraya–Watson or local linear weights and functions with two continuous derivatives with respect to the predictor. This assumption has been made for the sake of a transparent presentation and implies that the bias of the estimates is negligible compared to the stochastic part. Such an approach is standard in nonparametric estimation for censored data, see [10] or [31]. In principle, most results of the present paper can be extended to bandwidths $h \sim n^{-1/(d+4)}$ if a corresponding bias term is subtracted.

Another useful property of estimators constructed from weights satisfying condition (W1) is that they are increasing with probability tending to one.

Lemma 3.4. *Under condition (W1)(1), we have*

$$\mathrm{P}\Big(\text{“\textit{The estimates} } H_n(\cdot|x), H_{0n}(\cdot|x), H_{1n}(\cdot|x), H_{2n}(\cdot|x) \text{ \textit{are increasing}”}\Big) \overset{n\to\infty}{\longrightarrow} 1.$$

The lemma follows from the relation

$$\big\{\text{“The estimates } H_n(\cdot|x), H_{0n}(\cdot|x), H_{1n}(\cdot|x), H_{2n}(\cdot|x) \text{ are increasing”}\big\} \supseteq \big\{W_i(x) \geq 0 \ \forall i\big\}$$

and the fact that under assumption (W1) the probability of the event on the right-hand side converges to one. We will use Lemma 3.4 for the analysis of the asymptotic properties of the conditional quantile estimators in Section 3.2. One noteworthy consequence of the lemma is the fact that

$$\mathrm{P}\big(\hat{q}^{\mathrm{IP}}(\cdot|x) \equiv \hat{q}(\cdot|x)\big) \to 1,$$

which follows because the mappings Ψ and the right continuous inversion mapping coincide on the set of nondecreasing functions. In particular, this indicates that, from an asymptotic point of view, it does not matter which of the estimators $\hat{q}, \hat{q}^{\mathrm{IP}}$ is used. The difference between both estimators will only be visible in finite samples – see Section 4. In fact, it can only occur if one of the estimators H_n, $H_{k,n}$ is decreasing at some point.

3.1. Weak convergence of the estimate of the conditional distribution

We are now ready to describe the asymptotic properties of the estimates defined in Section 2. Our first result deals with the weak uniform consistency of the estimate $F_{T,n}(\cdot|x)$ under some rather weak conditions. In particular, it does neither require the existence of densities of the conditional distribution functions [see (D3)] nor integrability conditions like (D4).

Theorem 3.5. *If conditions (D1), (D3), (D11), (W1)(1), (W1)(2) and (W3) are satisfied, then the following statements are correct.*

1. *The estimate $F_{T,n}(\cdot|x)$ defined in (2.12) is weakly uniformly consistent on the interval $[0,\tau]$ for any τ such that $F_S(\tau|x) < 1$.*
2. *If additionally $F_S(\tau_{T,1}(x)|x) = 1$, where*

$$\tau_{T,1}(x) := \sup\{t : F_T(t|x) < 1\},$$

and $F_{T,n}(\cdot|x)$ is increasing and takes values in the interval $[0,1]$, the weak uniform consistency of the estimate $F_{T,n}(\cdot|x)$ holds on the interval $[0,\infty)$.

The next two results deal with the weak convergence of $F_{T,n}$ and require additional assumptions on the censoring distribution. We begin with a result for the estimator $F_{L,n}$, which is computed in the first step of our procedure by formulas (2.6) and (2.7).

Theorem 3.6.

1. *Let the weights used for $H_{2,n}$ and H_n in the definition of the estimate $M^-_{2,n}$ in (2.11) satisfy conditions (W1) and (W2). Moreover, assume that conditions (B1), (D1) and (D3)–(D10) hold. Then we have as $n \to \infty$*

$$\sqrt{nh^d}\big(H_n - H, H_{0,n} - H_0, M^-_{n,2} - M^-_2\big) \Rightarrow (G, G_0, G_M)$$

in $D^3([t_{00},\infty])$, where (G, G_0, G_M) denotes a centered Gaussian process with a.s. continuous sample paths and $G_M(t) = A(t) - B(t)$ is defined by

$$A(t) = \int_t^\infty \frac{\mathrm{d}G_2(u)}{H(u|x)}, \qquad B(t) := \int_t^\infty \frac{G(u)}{H^2(u|x)} H_2(\mathrm{d}u|x). \tag{3.4}$$

Here the process (G_0, G_2, G) is specified in assumption (W2) and the integral with respect to the process $G_2(t)$ is defined via integration-by-parts.

2. *Under the conditions of the first part, we have*

$$\sqrt{nh^d}(H_n - H, H_{0,n} - H_0, F_{L,n} - F_L) \Rightarrow (G, G_0, G_3)$$

in $D^3([t_{00},\infty])$, where the process (G_0, G_2, G) is specified in assumption (W2) and G_3 is a centered Gaussian process with a.s. continuous sample paths which is defined by

$$G_3(t) = F_L(t|x)G_M(t).$$

Remark 3.7. The value of the process G_M at the point t_{00} is defined as its path-wise limit. The existence of this limit follows from assumption (D4) and the representation

$$\mathrm{E}\big[G_M(s)G_M(t)\big] = b(x)\int_{s\vee t}^\infty \frac{1}{H(u|x)} M^-_2(\mathrm{d}u|x)$$

for the covariance structure of G_M, which can be derived by computations similar to those in [35].

Theorem 3.8. *Assume that the conditions of Theorem 3.6 and condition (D11) are satisfied. Moreover, let $t_{00} < \tau$ such that $F_S([0,\tau]|x) < 1$. Then we have the following weak convergence*:

1. $$\sqrt{nh^d}\big(\Lambda^-_{T,n} - \Lambda^-_T\big) \Rightarrow V$$

   in $D([0,\tau])$, where

   $$V(t) := \int_0^t \frac{G_0(\mathrm{d}u)}{(F_L - H)(u-|x)} - \int_0^t \frac{G_3(u-) - G(u-)}{(F_L - H)^2(u-|x)} H_0(\mathrm{d}u|x)$$

   is a centered Gaussian process with a.s. continuous sample paths and the integral with respect to G_0 is defined via integration-by-parts.
2. $$\sqrt{nh^d}(F_{T,n} - F_T) \Rightarrow W$$

   in $D([0,\tau])$, where

   $$W(t) := \big(1 - F_T(t|x)\big)V(t)$$

   is a centered Gaussian process with a.s. continuous sample paths.

Note that the second part of Theorem 3.8 follows from the first part using the representation (2.13) and the delta method.

3.2. Weak convergence of conditional quantile estimators

In this subsection, we discuss the asymptotic properties of the two conditional quantile estimates $\hat{q}$ and $\hat{q}^{\mathrm{IP}}$ defined in (2.17) and (2.25), respectively. As an immediate consequence of Theorem 3.5 and the continuity of the quantile mapping (see [22], Proposition 1), we obtain the weak consistency result.

Theorem 3.9. *If the assumptions of the first part of Theorem 3.5 are satisfied and additionally the conditions $F_S(F_T^{-1}(\tau|x)|x) < 1$ and $\inf_{\varepsilon \le t \le \tau} f_T(t|x) > 0$ hold some some $\varepsilon > 0$, then the estimators $\hat{q}(\cdot|x)$ and $q^{\mathrm{IP}}(\cdot|x)$ defined in (2.17) and (2.25) are weakly uniformly consistent on the interval $[\varepsilon,\tau]$.*

The compact differentiability of the quantile mapping and the delta method yield the following result.

Theorem 3.10. *If the assumptions of Theorem 3.8 are satisfied, then we have for any* $\varepsilon > 0$ *and* $\tau > 0$ *with* $F_S(F_T^{-1}(\tau|x)|x) < 1$ *and* $\inf_{\varepsilon \le t \le \tau} f_T(t|x) > 0$

$$\sqrt{nh^d}\big(\hat{q}(\cdot|x) - F_T^{-1}(\cdot|x)\big) \Rightarrow Z(\cdot) \qquad \text{on } D\big([\varepsilon, \tau]\big),$$
$$\sqrt{nh^d}\big(\hat{q}^{\mathrm{IP}}(\cdot|x) - F_T^{-1}(\cdot|x)\big) \Rightarrow Z(\cdot) \qquad \text{on } D\big([\varepsilon, \tau]\big),$$

where Z *is a centered Gaussian process defined by*

$$Z(\cdot) = -\frac{W \circ F_T^{-1}(\cdot|x)}{f_T(\cdot|x) \circ F_T^{-1}(\cdot|x)}$$

and the centered Gaussian process W *is defined in part* 2 *of Theorem* 3.8.

The proofs Theorems 3.5–3.10 are presented in Appendix A and require several separate steps. A main step in the proof is a result regarding the weak convergence of the Beran estimator on the maximal possible domain in the setting of conditional right censorship. We were not able to find such a result in the literature. Because this question is of independent interest, it is presented separately in the following subsection.

3.3. A new result for the Beran estimator

We consider the common conditional right censorship model (see [10] for details). Assume that our observations consist of the triples (X_i, Z_i, Δ_i) where $Z_i = \min(B_i, D_i)$, $\Delta_i = I_{\{Z_i = D_i\}}$, the random variables B_i, D_i are independent conditionally on X_i and nonnegative almost surely. The aim is to estimate the conditional distribution function F_D of D_i. Following [5], this can be done by estimating F_Z, the conditional distribution function of Z, and $\pi_k(t|x) := \mathrm{P}(Z_i \le t, \Delta_i = k|X = x)$ $(k = 0, 1)$ through

$$F_{Z,n}(t|x) := W_i(x) I_{\{Z_i \le t\}}, \qquad \pi_{k,n}(t|x) := W_i(x) I_{\{Z_i \le t, \Delta_i = k\}} \qquad (k = 0, 1) \tag{3.5}$$

and then defining an estimator for F_D as

$$F_{D,n}(t|x) := 1 - \prod_{[0,t]} \big(1 - \Lambda^-_{D,n}(\mathrm{d}s|x)\big), \tag{3.6}$$

where the quantity $\Lambda^-_{D,n}(\mathrm{d}s|x)$ is given by

$$\Lambda^-_{D,n}(\mathrm{d}s|x) := \frac{\pi_{0,n}(\mathrm{d}s|x)}{1 - F_{Z,n}(s - |x)}, \tag{3.7}$$

and the $W_i(x)$ denote local weights depending on $X_1, \dots, X_n$ (see also the discussion at the beginning of Section 3).

The weak convergence of the process $\sqrt{nh^d}(F_{D,n}(t|x) - F_D(t|x))_t$ in $D([0,\tau])$ with $\pi_0(\tau|x) < 1$ was first established by [10]. An important problem is to establish conditions that ensure that the weak convergence can be extended to $D([0,t_0])$ where $t_0 := \sup\{s : \pi_0(s|x) < 1\}$.

In the unconditional case, such conditions were derived by [21] who used counting process techniques. A generalization of this method to the conditional case was first considered by [33] and later exploited by [13] and [31]. However, none of those authors considered weak convergence on the maximal possible interval $[0,t_0]$. The following theorem provides sufficient conditions for the weak convergence on the maximal possible domain.

Theorem 3.11. *Assume that for some* $\varepsilon > 0$:

(R1) *The conditional distribution functions* $F_D(\cdot|x)$ *and* $F_B(\cdot|x)$ *have densities, say* $f_D(\cdot|x)$ *and* $f_B(\cdot|x)$, *with respect to the Lebesque measure,*

(R2) $\int_0^{t_0} \frac{\lambda_D(t|x)}{1-F_Z(t-|x)} \,\mathrm{d}t < \infty$,

(R3) $\sup_{k=1,\dots,d} \int_0^{t_0} \frac{|\partial_{x_k}\lambda_D(t|x)|}{1-F_Z(t-|x)} \,\mathrm{d}t < \infty$,

(R4) $\sup_{j,k=1,\dots,d} \sup_{(t,y)\in(0,t_0)\times U_\varepsilon(x)} |\partial_{y_k}\partial_{y_j}\lambda_D(t|y)| < \infty$,

(R5) $1 - F_Z(t|y) \geq C(1 - F_Z(t|x))$ *for all* $(t,y) \in (t_0 - \varepsilon, t_0] \times I$ *where* I *is a set with the property* $\int_{I\cap U_\delta(x)} f_X(s)\,\mathrm{d}s \geq c\delta^d$ *for some* $c > 0$ *and all* $0 < \delta \leq \varepsilon$,

(R6) $\lambda_D(t|y) = \lambda_D(t|x)(1 + \mathrm{o}(1))$ *uniformly in* $t \in (t_0 - \varepsilon, t_0]$ *as* $y \to x$.

Moreover, let the weights in (3.5) *satisfy condition* (W1) *and let the weak convergence*

$$\sqrt{nh^d}\big(F_{Z,n}(\cdot|x) - F_Z(\cdot|x), \pi_{0,n}(\cdot|x) - \pi_0(\cdot|x)\big) \Rightarrow (G, G_0) \qquad \text{on } D\big([0,\infty)\big)$$

to a centered Gaussian process (G, G_0) *with covariance structure given by*

$$\begin{aligned}
\operatorname{Cov}\big(G_0(s|x), G_0(t|x)\big) &= b(x)\big(\pi_0(s\wedge t|x) - \pi_0(s|x)\pi_0(t|x)\big),\\
\operatorname{Cov}\big(G(s|x), G(t|x)\big) &= b(x)\big(F_Z(s\wedge t|x) - F_Z(s|x)F_Z(t|x)\big),\\
\operatorname{Cov}\big(G_0(s|x), G(t|x)\big) &= b(x)\big(\pi_0(s\wedge t|x) - \pi_0(s|x)F_Z(t|x)\big)
\end{aligned}$$

for some function $b(x)$ *hold* (*this is the case for Nadaraya–Watson or local linear weights, see Lemma* 3.3). *Then under assumption* (B1)

$$\sqrt{nh^d}\big(F_{D,n}(\cdot|x) - F_D(\cdot|x)\big)_t \Rightarrow G_D(\cdot) \qquad \text{in } D\big([0,t_0]\big), \tag{3.8}$$

where G_D *denotes a centered Gaussian process with covariance structure taking the form*

$$\operatorname{Cov}\big(G_D(t), G_D(s)\big) = b(x)\big(1 - F_D(s|x)\big)\big(1 - F_D(t|x)\big)\int_0^{s\wedge t} \frac{\Lambda_D(\mathrm{d}u|x)}{1 - F_Z(u|x)}.$$

4. Simulation results

We have performed a small simulation study in order to investigate the finite sample properties of the proposed estimates. An important but difficult question in the estimation of the conditional distribution function from censored data is the choice of the smoothing parameter. For conditional right censored data, some proposals regarding the choice of the bandwidth have been made by [13] and [30]. In order to obtain a reasonable bandwidth parameter for our simulations, we used a modification of the cross validation procedure proposed by [1] in the context of nonparametric quantile regression. To address the presence of censoring in the cross validation procedure, we proceeded as follows:

1. Divide the data in blocks of size K with respect to the (ordered) X-components. Let $\{(Y_{jk}, X_{jk}, \delta_{jk})|j = 1, \dots, J_k\}$ denote the points among $\{(Y_i, X_i, \delta_i)|i = 1, \dots, n\}$ which fall in block k ($k = 1, \dots, K$). For our simulations, we used $K = 25$ blocks.
2. In each block, estimate the distribution function F_T as described in Section 2.1. Denote the sizes of the jumps at the jth uncensored observation in the kth block by w_{jk}.
3. Define

$$h := \arg\min_{\alpha} \sum_{k=1}^{K} \sum_{j=1}^{J_k} w_{jk} \rho_\tau \big(Y_{jk} - \tilde{q}_\alpha^{j,k}(\tau|X_{jk})\big),$$

   where ρ_τ denotes the check function and $\tilde{q}_\alpha^{j,k}$ is either the estimator $\hat{q}^{\mathrm{IP}}$ or $\hat{q}$ with bandwidth α based on the sample $\{(Y_i, X_i, \delta_i)|i = 1, \dots, n\}$ without the observation $(Y_{jk}, X_{jk}, \delta_{jk})$.

For a motivation of the proposed procedure, observe that the classical cross validation is based on the fact that each observation is an unbiased 'estimator' for the regression function at the corresponding covariate. In the presence of censoring, such an estimator is not available. Therefore, the cross validation criterion discussed above tries to mimic this property by introducing the weights w_{jk}. A deeper investigation of the theoretical properties of the procedure is beyond the scope of the present paper and postponed to future research. In order to save computing time, the bandwidth that we used for our simulations is an average of 100 cross validation runs in each scenario.

For the calculation of the estimators of the conditional sub-distribution functions, we chose local linear weights (see Remark 3.1) with a truncated version of the Gaussian Kernel, that is,

$$K(x) = \phi(x) I_{\{\phi(x) > 0.001\}},$$

where ϕ denotes the density of the standard normal distribution.

We investigate the finite sample properties of the new estimators in a similar scenario as model 2 in [49] (note that we additionally introduce a censoring mechanism). The model is given by

$$\text{(model 1)} \quad \begin{cases} T_i = 2.5 + \sin(2X_i) + 2\exp\big(-16X_i^2\big) + 0.5\mathcal{N}(0, 1), \\ L_i = 2.6 + \sin(2X_i) + 2\exp\big(-16X_i^2\big) + 0.5\big(\mathcal{N}(0, 1) + q_{0.1}\big), \\ R_i = 3.4 + \sin(2X_i) + 2\exp\big(-16X_i^2\big) + 0.5\big(\mathcal{N}(0, 1) + q_{0.9}\big), \end{cases}$$

where the covariates X_i are uniformly distributed on the interval $[-2, 2]$ and q_p denotes the p-quantile of a standard normal distribution. This means that about 10% of the observations are censored by type $\delta = 1$ and $\delta = 2$, respectively. For the sample size, we use $n = 100, 250, 500$. In

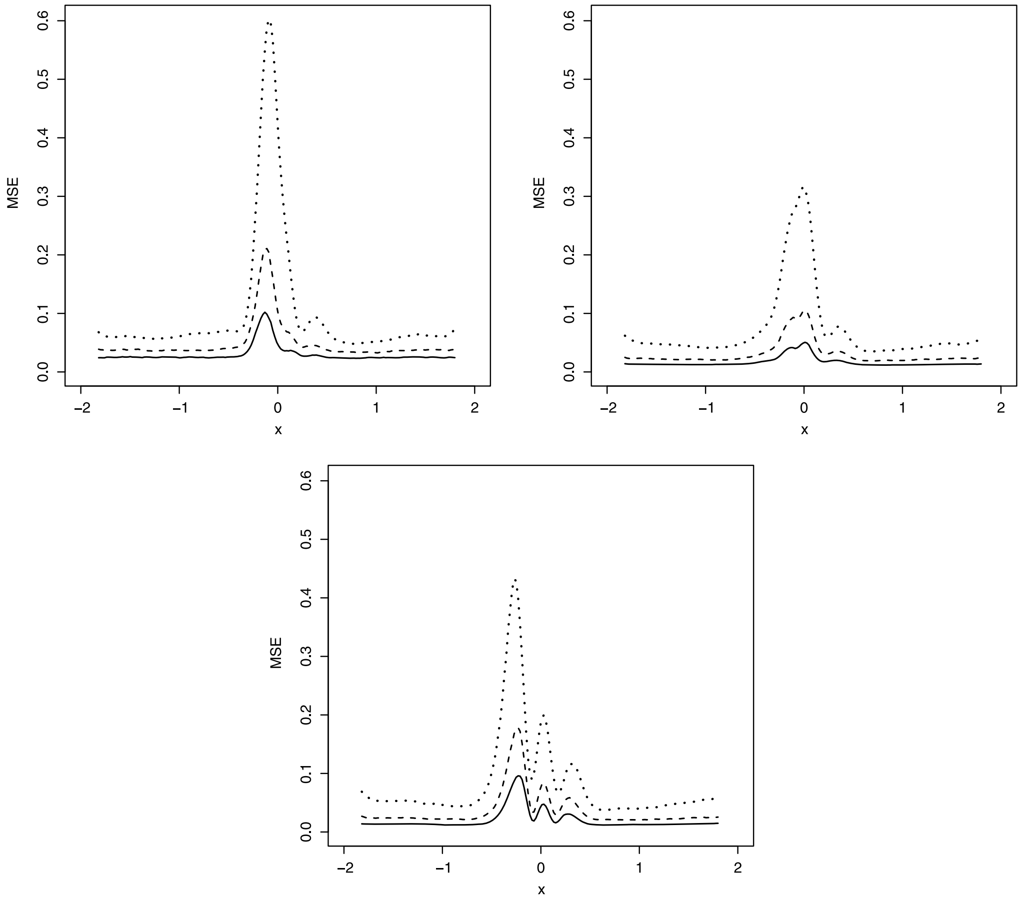

Figure 1. Mean squared error curves of the estimates of the quantile curves in model 1 for different sample sizes: $n = 100$ (dotted line); $n = 250$ (dashed line); $n = 500$ (solid line). Left panel: estimates of the 25%-quantile curves; middle panel: estimates of the 50%-quantile curves; right panel: estimates of the 75%-quantile curves. 10% of the observations are censored by type $\delta = 1$ and $\delta = 2$, respectively.

Figures 1 and 2, we show the mean conditional quantile curves and corresponding mean squared error curves for the 25%, 50% and 75% quantile based on 5000 simulation runs. The cases where the $\hat{q}^{\mathrm{IP}}(\tau|x)$ is not defined are omitted in the estimation of the mean squared error and mean curves (this phenomenon occurred in less than 3% of the simulation runs). Only results for the the estimator $\hat{q}^{\mathrm{IP}}$ are presented because it shows a slightly better performance than the estimator $\hat{q}$. We observe no substantial differences in the performance of the estimates for the 25%, 50% and 75% quantile curves with respect to bias. On the other hand, it can be seen from Figure 1 that the estimates of the quantile curves corresponding to the 25% and 75% quantile have larger variability. In particular, the mse is large at the point 0, where the quantile curves attain their maximum.

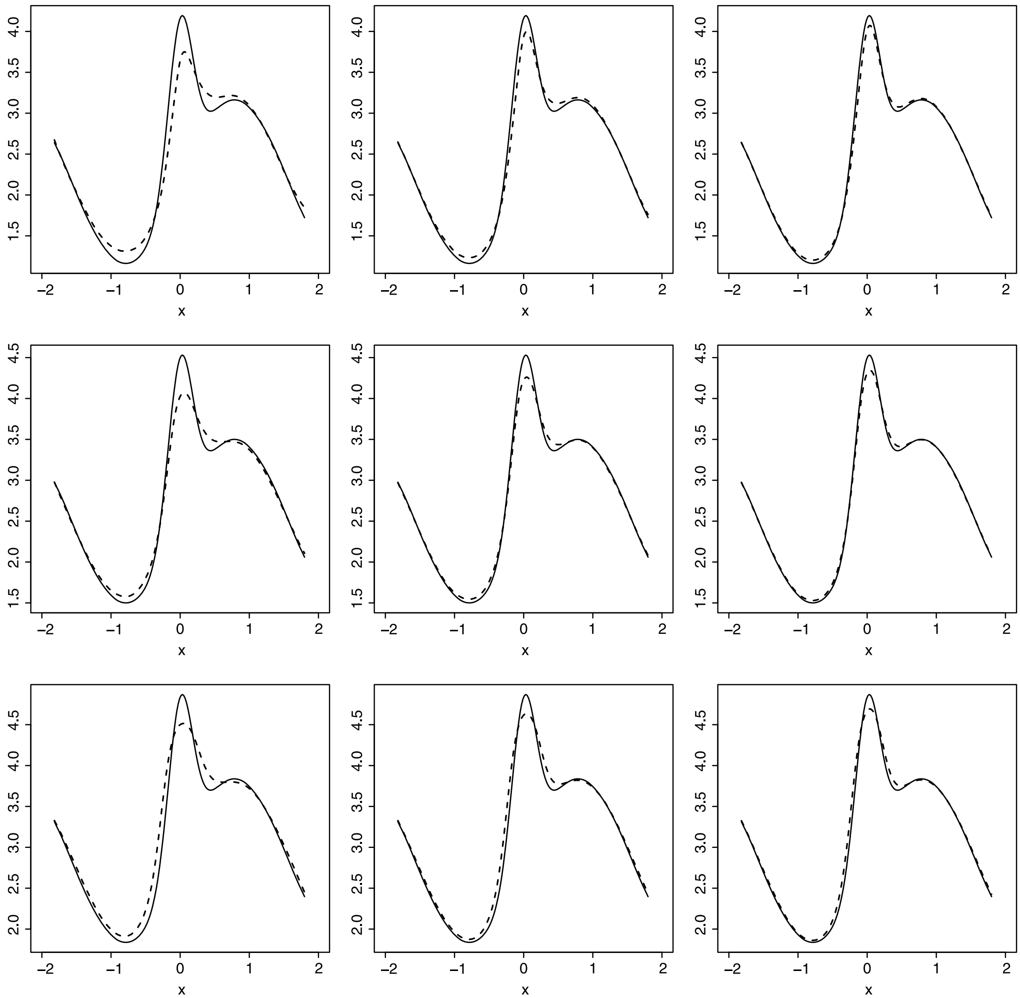

Figure 2. Mean (dashed lines) and true (solid lines) quantile curves for model 1 for different sample sizes: $n = 100$ (left column), $n = 250$ (middle column) and $n = 500$ (right column). Upper row: estimates of the 25% quantile curves; middle row: estimates of the 50% quantile curves; lower row: estimates of the 75% quantile curves. 10% of the observations are censored by type $\delta = 1$ and $\delta = 2$, respectively.

Appendix A: Proofs

Remark A.1. Before we begin with the proof of Theorem 3.5, we observe that condition (W1) implies that we can write the weights $W_i(x)$ in the estimates (2.5) in the form

$$W_i(x) = W_i^{(1)}(x) I_{A_n} + W_i^{(2)}(x) I_{A_n^C},$$

where A_n is some event with $\mathrm{P}(A_n) \to 1$, $W_i^{(1)}(x) = V_i(x)/\sum_j V_j(x)$ and $W_i^{(2)}(x)$ denote some other weights. If we now define modified weights

$$\tilde{W}_i(x) := W_i^{(1)}(x) I_{A_n} + W_i^{\mathrm{NW}}(x) I_{A_n^C},$$

where $W_i^{\mathrm{NW}}(x)$ denote Nadaraya–Watson weights, we obtain: $\mathrm{P}(\exists i \in 1, \dots, n : \tilde{W}_i \neq W_i) \to 0$, that is, any estimator constructed with the weights $\tilde{W}_i(x)$ will have the same asymptotic properties as an estimator based on the original weights $W_i(x)$. Thus, we may confine ourselves to the investigation of the asymptotic distribution of estimators constructed from the statistics in (2.5) that are based on the weights $\tilde{W}_i(x)$. In order to keep the notation simple, the modified estimates are also denoted by H_n, $H_{k,n}$, etc. Finally, observe that we have the representation $\tilde{W}_i(x) = \frac{\tilde{V}_i(x)}{\sum_j \tilde{V}_j(x)}$ with $\tilde{V}_i := V_i I_{A_n} + V_i^{\mathrm{NW}}(x) I_{A_n^C}$. Note that by construction, the random variables $\tilde{V}_i$ satisfy conditions (W1)(1)–(W1)(4) if the kernel in the definition of $W_i^{\mathrm{NW}}(x)$ satisfies assumption (K1).

Proof of Theorem 3.5. The uniform consistency of $F_{L,n}(\cdot|x)$ on $[t_{00} + \varepsilon, \infty)$ with $\varepsilon > 0$ arbitrary can be obtained from the continuity of the maps $(H_2(\cdot|x), H(\cdot|x)) \mapsto M_2^-(\cdot|x)$ and $M_2^-(\cdot|x) \mapsto F_L(\cdot|x)$ on suitable spaces of functions, see [45] for more details.

In the next step, we consider the map

$$\big(H_{0,n}(\cdot|x), H_n(\cdot|x), F_{L,n}(\cdot|x)\big) \mapsto \Lambda_{T,n}(\cdot|x) = \int_0^{\cdot} \frac{H_{0,n}(\mathrm{d}t|x)}{F_{L,n}(t-|x) - H_n(t-|x)}$$

and split the range of integration into the intervals $[0, t_{00} + \varepsilon)$ and $[t_{00} + \varepsilon, t)$. The continuity of the integration and fraction mappings yields the uniform convergence

$$\sup_{t \in [t_{00}+\varepsilon, \tau)} \bigg| \int_{[t_{00}+\varepsilon, t)} \frac{H_{0,n}(\mathrm{d}t|x)}{F_{L,n}(t-|x) - H_n(t-|x)} - \int_{[t_{00}+\varepsilon, t)} \frac{H_0(\mathrm{d}t|x)}{F_L(t-|x) - H(t-|x)} \bigg| \xrightarrow{P} 0 \tag{A.1}$$

for any τ with $F_S(\tau|x) < 1$ [note that $\inf_{t \in [t_{00}+\varepsilon, \tau)} F_L(t-|x) - H(t-|x) > 0$ since $F_L(t-|x) - H(t-|x) = F_L(t-|x)(1 - F_S(t-|x))$ and $F_L(t_{00}-|x) > 0$ by assumption (D11) and continuity of the conditional distribution function $F_L(\cdot|x)$]. We now will show that the integral over the interval $[0, t_{00} + \varepsilon)$ can be made arbitrarily small by an appropriate choice of ε. To this end, denote by $W_1(x,n), \dots, W_k(x,n)$ those values of $Y_1, \dots, Y_n$, whose weights fulfill $W_i(x) \neq 0$ and by $W_{(1)}(x,n), \dots, W_{(k)}(x,n)$ the corresponding increasingly ordered values. By Lemma B.1 in Appendix B, we can find an $\varepsilon > 0$ such that

$$\int_{[W_{(2)}(x,n), t_{00}+\varepsilon)} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \le H_{0,n}(t_{00} + \varepsilon|x) \mathrm{O}_P(1).$$

The integral $\int_{[0, W_{(2)}(x,n))} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)}$ can be bounded by separately considering $\Delta H_{0,n}(W_{(1)}(x,n)|x) = 0$ (in this case, the integral vanishes) and $\Delta H_{0,n}(W_{(1)}(x,n)|x) > 0$ [in this case $F_{L,n}(s|x) = F_{L,n}(W_{(2)}(x,n)|x)$ for all $s \in [0, W_{(2)}(x,n))]$. Summarizing, we have obtained the estimate

$$\int_{[0,t_{00}+\varepsilon)} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \le H_{0,n}(t_{00}+\varepsilon|x)\mathrm{O}_P(1) = H_0(t_{00}+\varepsilon|x)\mathrm{O}_P(1),$$

where the last equality follows from the uniform consistency of $H_{0,n}$ and the remainder $\mathrm{O}_P(1)$ does not depend on ε. Moreover, since the function $\Lambda_{T,n}(\cdot|x)$ is increasing (see Lemma 2.3), the inequality

$$\sup_{t\le t_{00}+\varepsilon} \left|\Lambda_{T,n}(t|x)\right| = \int_{[0,t_{00}+\varepsilon)} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \le H_0(t_{00}+\varepsilon|x)\mathrm{O}_P(1) \qquad \text{(A.2)}$$

follows. Now for any $\delta > 0$, we can choose an $\varepsilon_\delta > 0$ such that $H_0(t_{00}+\varepsilon_\delta|x) < \delta$ [recall the definition of t_{00} in (3.1)] and together with (A.2) we obtain for any τ with $F_S(\tau|x) < 1$

$$\begin{aligned}\mathrm{P}\Big(\sup_{t\in[0,\tau)} \left|\Lambda_{T,n}(t|x) - \Lambda_T(t|x)\right| > 4\alpha\Big) &\le \mathrm{P}\Big(\sup_{t\in[t_{00}+\varepsilon_\delta,\tau)} \left|\Lambda_{T,n}(t|x) - \Lambda_T(t|x)\right| > 2\alpha\Big)\\ &\quad + \mathrm{P}\big(\mathrm{O}_P(1) > \alpha/\delta\big),\end{aligned}$$

and by choosing δ appropriately the right-hand side can be made arbitrarily small for $n \to \infty$.

Thus, we obtain $\lim_{n\to\infty} \mathrm{P}(\sup_{t\in[0,\tau)}|\Lambda_{T,n}(t|x) - \Lambda_T(t|x)| > 4\alpha) = 0$, which implies the weak uniform consistency of $\Lambda_{T,n}(\cdot|x)$ on the interval $[0,\tau)$.

Finally, the continuity of the mapping $\Lambda_T \mapsto F_T$ (see the discussion in [2] following Proposition II.8.7) yields the weak uniform consistency of the estimate $F_{T,n}$ and the first part of the theorem is established.

For a proof of the second part, we use an idea from [46]. Note that, as soon as $F_{T,n}(\cdot|x)$ is increasing and bounded by 1 from above, we have

$$\sup_{t\ge 0}\left|F_{T,n}(t|x) - F_T(t|x)\right| \le 2 \sup_{0\le t\le a}\left|F_{T,n}(t|x) - F_T(t|x)\right| + 2\big(1 - F_T(a|x)\big),$$

and by assumption and part one of the theorem we can make $1 - F_T(a|x)$ arbitrarily small with uniform consistency on the interval $[0,a]$ still holding. Consequently, we obtain the uniform consistency on $[0,\infty)$, which completes the proof of Theorem 3.5. □

Proof of Theorem 3.6. The second part follows from the first one by the Hadamard differentiability of the map $A \mapsto \prod_{(t,\infty]}(1 - A(\mathrm{d}s))$ in definition (2.10) (see [35], Lemma A.1) and the delta method [22]. Note that these results require a.s. continuity of the sample paths which follows from the fact that the process G_M defined in the first part of the Theorem has a.s. continuous sample paths together with the continuity of $F_L(\cdot|x)$.

The proof will now proceed in two steps: first, we will show that weak convergence holds in $D^3([\sigma,\infty])$ for any $\sigma > t_{00}$ and secondly we will extend this convergence to $D^3([t_{00},\infty])$. Note that from condition (D4) we obtain $F_L(t_{00}|x) > 0$, and the continuity of $F_L(\cdot|x)$ yields $t_{00} > 0$.

For the first step, note that $\sigma > t_{00}$ implies $H(\sigma|x) > \varepsilon$ for some $\varepsilon > 0$. The weak convergence in $D^3([\sigma,\infty])$ essentially follows by an application of the delta method (see [22], Theorem 3) to the map $(H, H_0, H_2) \mapsto (H, H_0, M_2^-)$ which is Hadamard differentiable on a suitable domain as noted by [35]. More details can be found in [45].

To obtain the weak convergence in $D^3([t_{00}, \infty])$, we apply a lemma from [38], page 70, Example 11. First, define G_M as the path-wise limit of $G_{M_\sigma}(\sigma)$ for $\sigma \downarrow t_{00}$, the existence of this limit is discussed in Remark 3.7. Note that there exist versions of G_M, G, G_0 with a.s. continuous paths (this holds for G and G_0 by assumption, whereas the paths of G_M are obtained from those of G_2, G by a transformation that preserves continuity [see equation (3.4)]), and hence the condition on the limit process in the lemma is fulfilled.

Hereby, we have obtained a Gaussian process G_M on the interval $[t_{00}, \infty]$ and have taken care of condition (iii) in the lemma in [38]. For arbitrary positive ε and δ, we now have to find a $\sigma = \sigma(\delta, \varepsilon) > t_{00}$ such that

$$P\Big(\sup_{t_{00}<t\leq\sigma} \big|G_M(t)\big| \geq \delta\Big) < \varepsilon, \tag{A.3}$$

$$\limsup_{n\to\infty} P\Big(\sup_{t_{00}<t\leq\sigma} \sqrt{nh^d}\big|\big(M_{2,n}^- - M_2^-\big)(\sigma - |x) - \big(M_{2,n}^- - M_2^-\big)(t - |x)\big| \geq \delta\Big) < \varepsilon. \tag{A.4}$$

Note that once we have found a σ such that (A.4) holds, we can make σ smaller until (A.3) is fulfilled with (A.4) still holding. This is possible because the distribution of $G_M(t)$ corresponds to that of a time-transformed Brownian motion.

In order to prove the existence of a constant σ that ensures (A.4), we reverse time and transform our problem into the setting of conditional right censorship (see Section 3.3). To be more precise, define the function $a(t) := \frac{1}{t}$ which is strictly decreasing and maps the interval $[0, \infty]$ onto itself and consider the random variables $B_i := a(S_i)$, $D_i := a(L_i)$, $Z_i := B_i \wedge D_i$ and $\Delta_i := I_{\{D_i \leq B_i\}} = I_{\{S_i \leq L_i\}}$. It now can be seen that assertion (A.4) follows from (A.6) which is established in the proof of Theorem 3.11 (note that the assumptions (R2)–(R6) can be directly identified with the assumptions of Theorem 3.6). See [45] for more details. □

Proof of Theorem 3.8. First of all, note that the a.s. continuity of the sample paths of the processes $V(\cdot)$ and $W(\cdot)$ follows because these processes are constructed from processes which already have a.s. continuous sample paths in a way that preserves continuity. Thus, it remains to verify the weak convergence. From Theorem 3.6, we obtain

$$\sqrt{nh^d}(H_n - H, H_{0,n} - H_0, F_{L,n} - F_L) \Rightarrow (G, G_0, G_3) \tag{A.5}$$

in $D^3([t_{00}, \infty])$. Now from $F_L(s - |x) - H(s - |x) = F_L(s - |x)(1 - F_S(s - |x))$ and the definition of τ, it follows that

$$F_L(s - |x) - H(s - |x) \geq \varepsilon > 0 \qquad \forall s \in [t_{00}, \tau]$$

[note that the inequality $F_L(t_{00} - |x) > 0$ was derived at the beginning of the proof of Theorem 3.6]. For positive numbers δ, define the event

$$A_n(\delta) := \Big\{\inf_{t\in[t_{00},\tau)} \big(F_{L,n}(t|x) - H_n(t|x)\big) > \delta\Big\}.$$

Because of (A.5) [which implies the uniform consistency of $F_{L,n}(\cdot|x)$ and $H_n(\cdot|x)$], we have that for $\delta < \varepsilon$ $P(I_{A_n(\delta)} \neq 1) \stackrel{n\to\infty}{\longrightarrow} 0$. Define $\tilde{H}_n := H_n I_{A_n(\delta)}$, $\tilde{H}_{0,n} := H_{0,n} I_{A_n(\delta)}$ and $\tilde{F}_{L,n} := F_{L,n} I_{A_n(\delta)} + I_{A_n^C(\delta)}$, then it follows from (A.5)

$$\sqrt{nh^d}\big(\tilde{F}_{L,n} - F_L - (\tilde{H}_n - H), \tilde{H}_{0,n} - H_0\big) \Rightarrow (G_3 - G, G_0) \qquad \text{in } D^3\big([t_{00}, \tau]\big).$$

Moreover, we have $(\tilde{H}_{0,n}, \tilde{F}_{L,n} - \tilde{H}_n) \in \{(A, B) \in BV_1^2([t_{00}, \tau]) : A \geq 0, B \geq \delta > 0\}$. Since the map $(A, B) \mapsto \int_{t_{00}}^{t} \frac{\mathrm{d}A(s)}{B(s)}$ is Hadamard differentiable on this set (see [2], page 113), the delta method (see [22]) yields

$$\sqrt{nh^d}\left(\int_{t_{00}}^{\cdot} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} - \Lambda_T^-(\cdot|x)\right) \Rightarrow V(\cdot)$$

in $D([t_{00}, \tau]]$. Finally, observe that for $t \geq t_{00}$ we have

$$\Lambda_{T,n}^-(t|x) = \int_{t_{00}}^{t} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} + \int_{[0,t_{00})} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)},$$

and thus it remains to prove that the second term in this sum is of order $\mathrm{o}_P(1/\sqrt{nh^d})$. By the same arguments as in the proof of Theorem 3.5, one can obtain the bound

$$\int_{[0,t_{00})} \frac{H_{0,n}(\mathrm{d}s|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \leq H_{0,n}(t_{00}|x)\mathrm{O}_P(1).$$

Standard arguments yield the estimate $H_{0,n}(t_{00}|x) = \mathrm{o}_P(1/\sqrt{nh^d})$ and thus the proof is complete. □

Proof of Theorem 3.9. Note that the estimator $F_{T,n}^{\mathrm{IP}}(\cdot|x)$ is nondecreasing by construction. The assertion for $\hat{q}^{\mathrm{IP}}(\cdot|x)$ now follows from the Hadamard differentiability of the inversion mapping tangentially to the space of continuous functions (see Proposition 1 in [22]), the continuity of $F_T(\cdot|x)$ and the weak uniform consistency of $F_{T,n}^{\mathrm{IP}}(\cdot|x)$ on the interval $[0, \tau]$. The corresponding result for the estimator $\hat{q}(\cdot|x)$ follows from the convergence $\mathrm{P}(\hat{q}^{\mathrm{IP}}(\cdot|x) \equiv \hat{q}(\cdot|x)) \to 1$ (see the discussion after Lemma 3.4). □

Proof of Theorem 3.10. Observe that the estimator $F_{T,n}^{\mathrm{IP}}(\cdot|x)$ is nondecreasing by construction and that Theorem 3.8 yields $\sqrt{nh^d}(F_{T,n}^{\mathrm{IP}}(\cdot|x) - F^T(\cdot|x)) \Rightarrow W(\cdot)$ on $D([0, \tau + \alpha])$ for some $\alpha > 0$ where the process W has a.s. continuous sample paths. Note that the convergence holds on $D([0, \tau + \alpha])$. This follows from the continuity of $F_S(\cdot|x)$ and $F_T^{-1}(\cdot|x)$ at τ which implies $F_S(F_T^{-1}(\tau + \alpha|x)|x) < 1$ for some $\alpha > 0$. By the same arguments $f_T(\cdot|x) \geq \delta > 0$ on the interval $[\varepsilon - \alpha, \tau + \alpha]$ if we choose α sufficiently small. Thus, Proposition 1 from [22] together with the delta method yield the weak convergence of the process for $\hat{q}^{\mathrm{IP}}(\cdot|x)$. The corresponding result for $\hat{q}(\cdot|x)$ follows from the fact that $\mathrm{P}(\hat{q}^{\mathrm{IP}}(\cdot|x) \equiv \hat{q}(\cdot|x)) \to 1$. □

Proof of Theorem 3.11. By the delta method [22], formula (3.6), and the Hadamard differentiability of the product-limit mapping [2] it suffices to verify the weak convergence of $\sqrt{nh^d}(\Lambda^-_{D,n}(t|x) - \Lambda^-_D(t|x))_t$ on $D([0, t_0])$. The corresponding result on $D([0, \tau])$ with $\tau < t_0$ follows from the delta method and the Hadamard differentiability of the mapping $(\pi_{0,n}, F_{Z,n}) \mapsto \Lambda^-_{D,n}$. For the extension of the convergence to $D([0, t_0])$, it suffices to establish the following assertion

$$\limsup_{n\to\infty} P\Big(\sup_{\sigma\le t<t_0} \sqrt{nh^d}\big|\big(\Lambda^-_{D,n} - \Lambda^-_D\big)(t|x) - \big(\Lambda^-_{D,n} - \Lambda^-_D\big)(\sigma - |x)\big| > \delta\Big) < \varepsilon \tag{A.6}$$

(this follows by arguments similar to those in the proof of Theorem 3.6). Define the random variable U as the largest Z_i corresponding non-vanishing weight $\tilde{W}_i(x)$, that is,

$$U = U(x) := \max\big\{Z_i : \tilde{W}_i(x) \neq 0\big\}.$$

Note that for $t \geq U$ we have $F_{Z,n}(t|x) = 1$ for the corresponding estimate of $F_Z(\cdot|x)$. We write

$$\Lambda^-_{D,n}(y - |x) = \sum_{i=1}^n \int_{[0,y)} \frac{d(\tilde{W}_i(x)I_{\{Z_i\le t,\Delta_i=1\}})}{\sum_{j=1}^n \tilde{W}_j(x)I_{\{Z_j\ge t\}}} = \sum_{i=1}^n \int_{[0,y)} C_i(x,t)I_{\{1-F_{Z,n}(t-|x)>0\}}\,\mathrm{d}N_i(t)$$

for the plug-in estimator of $\Lambda^-_D(\cdot|x)$, where

$$C_i(x,t) := \frac{\tilde{W}_i(x)I_{\{Z_i\ge t\}}}{\sum_{j=1}^n \tilde{W}_j(x)I_{\{Z_j\ge t\}}} = \frac{\tilde{V}_i(x)I_{\{Z_i\ge t\}}}{\sum_{j=1}^n \tilde{V}_j(x)I_{\{Z_j\ge t\}}},$$

and the quantity $N_i(t)$ is defined as $N_i(t) := I_{\{Z_i\le t,\Delta_i=1\}}$. In what follows, we will use the notation $G(A) = \int_A G(\mathrm{d}u)$ for a distribution function G and a Borel set A. With the definition

$$\hat{\Lambda}^-_{D,n}(y - |x) := \sum_{i=1}^n \int_{[0,y)} C_i(x,t)I_{\{1-F_{Z,n}(t-|x)>0\}}\Lambda^-_D(\mathrm{d}t|X_i)$$

we obtain the decomposition

$$\begin{aligned}\big|\big(\Lambda^-_{D,n} - \Lambda^-_D\big)\big((\sigma,t]|x\big)\big| \le{}& \big|\big(\Lambda^-_{D,n} - \hat{\Lambda}^-_{D,n}\big)\big((\sigma,U\wedge t]|x\big)\big| + \big|\big(\Lambda^-_{D,n} - \hat{\Lambda}^-_{D,n}\big)\big((U\wedge t,t]|x\big)\big| \\ &+ \big|\big(\hat{\Lambda}_{D,n} - \Lambda^-_D\big)\big((\sigma,t]|x\big)\big|.\end{aligned}$$

Observing that $\Lambda^-_{D,n}((U\wedge t,t]) = \hat{\Lambda}^-_{D,n}((U\wedge t,t]) = 0$ it follows that

$$\begin{aligned}&\big|\big(\Lambda^-_{D,n} - \hat{\Lambda}^-_{D,n}\big)\big((U\wedge t,t]|x\big)\big| = 0,\\ &\big|\big(\hat{\Lambda}^-_{D,n} - \Lambda^-_D\big)\big((\sigma,t]|x\big)\big| \le \big|\big(\hat{\Lambda}^-_{D,n} - \Lambda^-_D\big)\big((\sigma,U\wedge t]|x\big)\big| + \Lambda^-_D\big((U\wedge t,t]|x\big),\\ &\sup_{\sigma\le t<t_0}\big|\big(\hat{\Lambda}^-_{D,n} - \Lambda^-_D\big)\big((\sigma,t\wedge U]|x\big)\big| \le \sup_{\sigma\le t\le U\wedge t_0}\big|\big(\hat{\Lambda}^-_{D,n} - \Lambda^-_D\big)\big((\sigma,t]|x\big)\big|,\end{aligned}$$

where we set the supremum over the empty set to zero. Hence, assertion (A.6) can be obtained from the statements

$$\sqrt{nh^d} \sup_{\sigma \le t < t_0} \Lambda_D^-\big((U \wedge t, t]|x\big) \xrightarrow{P} 0, \tag{A.7}$$

$$\sqrt{nh^d} \sup_{\sigma \le t \le U \wedge t_0} \big|\big(\hat{\Lambda}_{D,n}^- - \Lambda_D^-\big)\big((\sigma, t]|x\big)\big| \xrightarrow{P} 0, \tag{A.8}$$

$$\limsup_{n\to\infty} P\Big(\sqrt{nh^d} \sup_{\sigma \le t < U \wedge t_0} \big|\big(\Lambda_{D,n}^- - \hat{\Lambda}_{D,n}^-\big)\big((\sigma, U \wedge t]|x\big)\big| > \delta\Big) \;<\; \varepsilon/2, \tag{A.9}$$

which will be shown separately.

Proof of (A.7). For a proof of (A.7), note that $\Lambda_D^-((U \wedge t, t]|x) = \Lambda_D^-((U, t]|x) I\{U < t\}$ and $\Lambda_D^-((U, t]|x) \le \Lambda_D^-((U \wedge t_0, t_0]|x)$ whenever $U < t \le t_0$. Hence, the supremum in (A.7) can be bounded by [note that by (R2) we have $F_D([t_0, \infty]|x) > 0$]

$$\begin{aligned}\sup_{\sigma \le t < t_0} \Lambda_D^-\big((U \wedge t, t]|x\big) \le \Lambda_D^-\big((U \wedge t_0, t_0]|x\big) &\le \int_{(U \wedge t_0, t_0]} \frac{F_D(\mathrm{d}t|x)}{F_D([t_0, \infty]|x)} \\ &= \frac{F_D((U \wedge t_0, t_0]|x)}{F_D([t_0, \infty]|x)}.\end{aligned}$$

Thus, it suffices to verify the convergence $\sqrt{nh^d} F_D((U \wedge t_0, t_0]|x) \xrightarrow{P} 0$. For this purpose, we introduce the notation

$$u_n^\alpha = u_n^\alpha(x) := \inf\big\{s : \sqrt{nh^d} F_D\big((s, t_0]|x\big) \le \alpha\big\}$$

(note that $u_n^\alpha \le t_0$). Assume that the interval I in condition (R5) contains the set $[x, x + \beta)$ for some $\beta > 0$ [the other case $(x - \beta, x] \subseteq I$ can be treated analogously]. Then we obtain for any fixed $\alpha > 0$ and sufficiently large n (see [45] for more details)

$$\begin{aligned}&P\big(\sqrt{nh^d} F_D\big((U \wedge t_0, t_0]|x\big) > \alpha\big) \\ &\quad\le \mathrm{E}[I_{\{U \wedge t_0 < u_n^\alpha\}}] = \mathrm{E}\big[\mathrm{E}[I_{\{U \wedge t_0 < u_n^\alpha\}}|X_1, \ldots, X_n]\big] \\ &\quad\le \mathrm{E}\left[\mathrm{E}\left[\prod_{j=1}^n \{1 - I_{\{Z_j \ge u_n^\alpha\}} I_{\{\tilde{W}_i(x) \ne 0\}}\}|X_1, \ldots, X_n\right]\right] \\ &\quad\le \mathrm{E}\left[\prod_{j=1}^n \big\{1 - \mathrm{E}[I_{\{Z_j \ge u_n^\alpha\}}|X_j] I_{\{\|X_j - x\| \le c_n\}}\big\}\right] \\ &\quad\le \bigg(1 - C\frac{\alpha^2}{n} \frac{F_B([u_n^\alpha, \infty]|x)}{F_D([u_n^\alpha, t_0)|x)} f_X(x)\big(1 + \mathrm{o}(1)\big)\bigg).\end{aligned}$$

Now we have

$$\frac{F_D([u_n^\alpha, t_0)|x)}{F_B([u_n^\alpha, \infty]|x)} \leq \int_{[u_n^\alpha, t_0)} \frac{F_D(\mathrm{d}s|x)}{F_B((s,\infty]|x)}$$
$$\leq \int_{[u_n^\alpha, t_0)} \frac{\Lambda_D^-(\mathrm{d}s|x)}{F_Z((s,\infty]|x)} \longrightarrow 0$$

by (R2) (note that $u_n^\alpha \to t_0$ if $n \to \infty$) and hence the proof of (A.7) is complete.

Proof of (A.8). For fixed $\sigma \leq s \leq U \wedge t_0$ and sufficiently small h combining Taylor expansions and (R4) yields

$$\begin{aligned}
&\big|\big(\hat{\Lambda}_{D,n}^- - \Lambda_D^-\big)\big((\sigma,s]|x\big)\big| \\
&\quad \leq \left|\int_\sigma^s \sum_{i=1}^n C_i(x,t)(x - X_i)'\partial_x \lambda_D(t|x)\,\mathrm{d}t\right| \\
&\qquad + \int_\sigma^s \sum_{i=1}^n C_i(x,t)(x - X_i)^2 \frac{C}{2}\,\mathrm{d}t
\end{aligned}$$

with some positive constant C (here, $\partial_x \lambda_D$ is interpreted as the vector of partial derivatives with respect to the components of x). Noting that $C_i(x,t) \geq 0$ and $\sum_i C_i(x,t) = 1$ we see that the second term in the above expression is of order $\mathrm{O}(h^2) = \mathrm{o}((nh^d)^{-1/2})$ uniformly in $s \in [\sigma, t_0]$. For the first term can be represented as

$$R_n = \left|\frac{1}{\sum_{k=1}^n \tilde{V}_k(x)} \int_\sigma^s \sum_{i=1}^n \tilde{V}_i(x) I_{\{Z_i \geq t\}} \left(\frac{1 - F_Z(t-|x)}{1 - F_{Z,n}(t-|x)}\right)(x - X_i)' \frac{\partial_x \lambda_D(t|x)}{1 - F_Z(t-|x)}\,\mathrm{d}t\right|.$$

Using conditions (W1)(3), (W1)(4), (R3) and Lemma B.2 in Appendix B, we obtain $R_n = \mathrm{o}_P(1/\sqrt{nh^d})$ uniformly in $s \in [\sigma, t_0]$, and hence assertion (A.8) is established.

Proof of (A.9). Observe that $|(\Lambda_{D,n}^- - \hat{\Lambda}_{D,n}^-)((\sigma, U \wedge t_0]|x)| \leq |D_1(U \wedge t_0) - D_1(\sigma)|$, where we defined $D_1(t) := \sum_{i=1}^n \int_{[0,t]} C_i(x,t) I_{\{1 - F_{Z,n}(t-|x) > 0\}}\,\mathrm{d}M_i(t)$ and $M_i(t) := N_i(t) - \int_0^t I_{\{Z_i \geq s\}} \Lambda_D^-(\mathrm{d}s|X_i)$. Setting $\mathcal{F}_t := \sigma(X_i, I_{\{Z_i \leq t, \Delta_i = 1\}}, I_{\{Z_i \leq t, \Delta_i = 0\}} : i = 1, \ldots, n)$, classical arguments from counting process theory (see [45] for more details) show that for $t \in [\sigma, t_0]$ $D_1(t) - D_1(\sigma)$ is a locally bounded martingale with respect to $\mathcal{F}_t$. Its predictable variation is given by $P_t = \int_{[\sigma, U \wedge t]} \sum_{i=1}^n C_i^2(x,s) \Lambda_D^-(\mathrm{d}s|X_i)$. Hence from a version Lenglart's inequality (see [42], page 893, Example 1), we obtain

$$P\Big(\sup_{\sigma \leq t \leq U \wedge t_0} nh^d \big(D_1(t) - D_1(\sigma)\big)^2 \geq \varepsilon\Big) \leq \frac{\eta}{\varepsilon} + P\big(nh^d P_{t_0} \geq \eta\big). \tag{A.10}$$

If σ is sufficiently close to t_0 it follows

$$
\begin{aligned}
nh^d P_{t_0} &= nh^d \int_{[\sigma, U\wedge t_0]} \sum_{i=1}^{n} C_i^2(x,t)\Lambda_D^-(\mathrm{d}t|X_i) \\
&\leq \frac{nh^d \sup_j \tilde{V}_j(x)}{\sum_{k=1}^n \tilde{V}_k(x)} \int_{[\sigma, U\wedge t_0]} \sum_{i=1}^{n} \frac{C_i(x,t)\Lambda_D^-(\mathrm{d}t|X_i)}{(1-F_{Z,n}(t-|x))} \\
&\overset{(*)}{=} \mathrm{O}_P(1) \int_{[\sigma, U\wedge t_0]} \sum_{i=1}^{n} \frac{C_i(x,t)\lambda_D(t|x)}{(1-F_{Z,n}(t-|x))} \,\mathrm{d}t\big(1+\mathrm{o}_P(1)\big) \\
&= \mathrm{O}_P(1) \int_{[\sigma, U\wedge t_0]} \frac{\lambda_D(t|x)}{1-F_Z(t-|x)} \,\mathrm{d}t,
\end{aligned}
$$

where we have used (R6), (W1)(1) and (W1)(3) in equality $(*)$ [note that the $(1+\mathrm{o}_P(1))$ holds uniformly in i and t] and Lemma B.2 in the last equality. Now we obtain from (R2) the a.s. convergence $\int_{[\sigma,U\wedge t_0]} \frac{\lambda_D(t|x)}{1-F_Z(t-|x)}\,\mathrm{d}t \xrightarrow{\sigma\to t_0} 0$ and hence assertion (A.9) is established [first choose η in (A.10) small enough to make η/ε small and then choose σ close enough to t_0].

Thus, we have established (A.7)–(A.9) and the proof of the theorem is complete. □

Appendix B: Auxiliary results: Technical details

Lemma B.1. *Assume that conditions (D3) and (D11) hold. Denote by $W_1(x,n),\dots,W_k(x,n)$ those values of $Y_1,\dots,Y_n$, whose weights fulfill $W_i(x)\neq 0$ and by $W_{(1)}(x,n),\dots,W_{(k)}(x,n)$ the corresponding increasingly ordered values. Assume that the estimators $F_{L,n}$ and H_n are based on weights $W_i(x)=V_i(x)/\sum_j V_j(x)$ with $V_i(x)$ satisfying the conditions (W1)(1) and (W1)(2), that $F_{S,n}(r|x) := H_n(r|x)/F_{L,n}(r|x)$ is consistent for some $r > t_{00}$ with $F_S(r|x) < 1$ and that all the observations Y_i are distinct. Then we have for any $b < r$:*

$$
\sup_{b\geq s\geq W_{(2)}(x,n)} \frac{1}{F_{L,n}(s-|x)-H_n(s-|x)} = \mathrm{O}_P(1).
$$

Proof. Because of space considerations, we only sketch the main arguments, a much more detailed proof can be found in [45]. As in the proof of Theorem 3.6, we reverse the time and use the same notation. Write $V_x := a(W_{(2)}(x,n))$, $v=a(r)$, $w=a(b)$, then the statement of the lemma can be reformulated as

$$
\sup_{w\leq s\leq V_x} \frac{1}{1-F_{D,n}(s|x)-(1-F_{Z,n}(s|x))} = \mathrm{O}_P(1).
$$

With the notation $F_{B,n}(s|x) := 1-(1-F_{Z,n}(s|x))/(1-F_{D,n}(s|x))$, the denominator in this expression can be rewritten as $1-F_{D,n}(s|x)-(1-F_{Z,n}(s|x)) = (1-F_{D,n}(s|x))F_{B,n}(s|x)$ [note that $F_{B,n}(v|x) = 1-F_{S,n}(r-|x)$]. Since $F_{B,n}(s|x)$ is increasing in s and consistent at some point $v \le w$ with $F_{B,n}(v|x) > 0$, we only need to worry about finding a bound in probability for the term $1/(1 - F_{D,n}(s|x))$. Noting that $1 - F_{D,n}(t|x) = \prod_{[0,t]}(1 - \Lambda^-_{D,n}(\mathrm{d}s|x))$ and applying exactly the same arguments as given in the proof of Lemma 6 in [23] gives

$$1 - F_{D,n}(t|x) \ge \exp\bigl(-C_1 \Lambda^-_{D,n}(t|x)\bigr) \qquad \text{a.s.}$$

for some finite constant C_1. The bound $\sup_{t \le V_x} \Lambda^-_{D,n}(t|x) = \mathrm{O}_P(1)$ can be used by exploiting the martingale structure of $\Lambda^-_{D,n}(t \wedge V_x|x) - \hat{\Lambda}^-_{D,n}(t \wedge V_x|x)$ and applying Doob's submartingale inequality. □

Lemma B.2. *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ denote i.i.d. random variables with $F(y|x) := P(Y_1 \le y|X_1 = x)$. Define $\hat{F}(y|x) := \sum_i \frac{V_i(x) I_{\{Y_i \le y\}}}{\sum_j V_j(x)}$, which is an estimator of the conditional distribution function $F(y|x)$ and assume that the weights weights $V_i(x)$ satisfy conditions (W1)(1)–(W1)(3), the bandwidth h fulfills $nh^d \to \infty$, $h \to 0$ and that additionally the following conditions hold:*

1. *$F(t|x)$ is continuous at (t_0, x_0),*
2. *there exist constants $C > 0, \delta > 0$ such that $1 - F(t|y) \ge C(1 - F(t|x))$ for all $(t, y) \in (t_0 - \delta, t_0] \times I$ where I is a set with the property $\int_{I \cap U_\delta(x)} f_X(s)\,\mathrm{d}s \ge c\delta^d$ for some $c > 0$ and all $0 < \delta \le \varepsilon$,*
3. *$F(t_0 - \delta|z)$ is continuous in the second component at the point $z = x$,*
4. *the distribution function G of the random variables X_i has a continuous density g with $g(x) > 0$.*

Then, with the notation $U := \max\{Y_i : V_j(x) \ne 0\}$, we have for $n \to \infty$

$$\sup_{0 \le y \le t_0 \wedge U} \frac{1 - F(y-|x)}{1 - \hat{F}_n(y-|x)} = \mathrm{O}_P(1).$$

Proof. Define

$$\bar{F}_n(y|x) := \frac{\sum_{i=1}^n F(y|X_i) I_{\{\|x - X_i\| \le h\}}}{\sum_{i=1}^n I_{\{\|x - X_i\| \le h\}}}$$

and observe the representation (see [45] for details)

$$\frac{1 - F(y-|x)}{1 - \hat{F}_n(y-|x)} = \frac{1 - \bar{F}_n(y-|x)}{1 - \hat{F}_n(y-|x)} \frac{1 - F(y-|x)}{1 - \bar{F}_n(y-|x)} = \frac{1 - \bar{F}_n(y-|x)}{1 - \hat{F}_n(y-|x)} \mathrm{O}_P(1),$$

where the $\mathrm{O}_P(1)$ is uniform over $0 \le y \le t_0$. Next, observe that

$$1 - \hat{F}(y-|x) = \sum_i \frac{V_i(x)(1 - I_{\{Y_i < y\}})}{\sum_j V_j(x)} \le \underline{c} f_X(x) \frac{1 + \mathrm{o}_P(1)}{C(x)} \frac{\sum_i I_{\{\|x - X_i\| \le h\}}(1 - I_{\{Y_i \le y\}})}{\sum_j I_{\{\|x - X_j\| < h\}}},$$

uniformly in y. The fraction $(1 - \bar{F}_n(y-|x)) \sum_j I_{\{\|x - X_j\| < h\}} / \sum_i I_{\{\|x - X_i\| \le h\}}(1 - I_{\{Y_i \le y\}})$ can be bounded by applying the results from [44] conditionally on X_i, see [45] for more details. This completes the proof. □

Acknowledgements

The authors are grateful to Martina Stein who typed parts of this paper with considerable technical expertise. We are also grateful to an unknown referee and the associate editor for their constructive comments on an earlier version of this manuscript. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C1) of the German Research Foundation (DFG).

References

[1] Abberger, K. (2001). Cross validation in nonparametric quantile regression. *Allg. Statist. Archiv.* **82** 149–161.

[2] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes. Springer Series in Statistics*. New York: Springer. MR1198884

[3] Anevski, D. and Fougères, A.L. (2007). Limit properties of the monotone rearrangement for density and regression function estimation. Available at arXiv:0710.4617v1.

[4] Bennett, C. and Sharpley, R. (1988). *Interpolation of Operators. Pure and Applied Mathematics* **129**. Boston, MA: Academic Press. MR0928802

[5] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

[6] Chang, M.N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** 391–404. MR1041399

[7] Chang, M.N. and Yang, G.L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547. MR0913572

[8] Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. MR1105843

[9] Chernozhukov, V., Fernandéz-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125.

[10] Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Stat.* **14** 181–197. MR0932943

[11] Dabrowska, D.M. (1989). Uniform consistency of the kernel conditional Kaplan–Meier estimate. *Ann. Statist.* **17** 1157–1167. MR1015143

[12] Dabrowska, D.M. (1992). Nonparametric quantile regression with censored data. *Sankhyā Ser. A* **54** 252–259. MR1192099

[13] Dabrowska, D.M. (1992). Variable bandwidth conditional Kaplan–Meier estimate. *Scand. J. Stat.* **19** 351–361. MR1211789

[14] Dette, H., Neumeyer, N. and Pilz, K.F. (2005). A note on nonparametric estimation of the effective dose in quantal bioassay. *J. Amer. Statist. Assoc.* **100** 503–510. MR2160554

[15] Dette, H., Neumeyer, N. and Pilz, K.F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12** 469–490. MR2232727

[16] Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 609–627. MR2420417

[17] Efron, B. (1967). The two-sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics IV* (L. Le Cam and J. Neyman, eds.) 831–853. New York: Prentice Hall.

[18] El Ghouch, A. and Van Keilegom, I. (2008). Non-parametric regression with dependent censored data. *Scand. J. Stat.* **35** 228–247. MR2418738

[19] Fitzenberger, B. (1997). A guide to censored quantile regressions. In *Robust Inference*. *Handbook of Statist*. **15** 405–437. Amsterdam: North-Holland. MR1492720

[20] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR1100924

[21] Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann*. *Statist*. **11** 49–58. MR0684862

[22] Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand*. *J*. *Stat*. **16** 97–128. MR1028971

[23] Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann*. *Statist*. **18** 1501–1555. MR1074422

[24] Gu, M.G. and Zhang, C.H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann*. *Statist*. **21** 611–624. MR1232508

[25] Hardy, G.H., Littlewood, J.E. and Pólya, G. (1988). *Inequalities*. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. Reprint of the 1952 edition. MR0944909

[26] Koenker, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge: Cambridge Univ. Press. MR2268657

[27] Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software* **27** 1–25.

[28] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[29] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J*. *Amer. Statist*. *Assoc*. **96** 458–468. MR1939348

[30] Li, G. and Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Ann*. *Inst*. *Statist*. *Math*. **53** 708–729. MR1879606

[31] Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann*. *Statist*. **23** 787–823. MR1345201

[32] Lorentz, G.G. (1953). An inequality for rearrangements. *Amer*. *Math*. *Monthly* **60** 176–179. MR0052476

[33] McKeague, I.W. and Utikal, K.J. (1990). Inference for a nonlinear counting process regression model. *Ann*. *Statist*. **18** 1172–1187. MR1062704

[34] Meeker, W.Q. and Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. *Wiley Series in Probability and Mathematical Statistics*, *Applied Section* **xxii** 680 p. New York: Wiley.

[35] Patilea, V. and Rolin, J.M. (2001). Product-limit estimators of the survival function for doubly censored data. DP 0131, Institut de Statistique, Louvain-la-Neuve.

[36] Patilea, V. and Rolin, J.M. (2006). Product-limit estimators of the survival function with twice censored data. *Ann*. *Statist*. **34** 925–938. MR2283398

[37] Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J*. *Amer. Statist*. *Assoc*. **103** 637–649. MR2435468

[38] Pollard, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. New York: Springer. MR0762984

[39] Portnoy, S. (2003). Censored regression quantiles. *J*. *Amer. Statist*. *Assoc*. **98** 1001–1012. MR2041488

[40] Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model. *J*. *Econometrics* **25** 303–325. MR0752444

[41] Powell, J.L. (1986). Censored regression quantiles. *J*. *Econometrics* **32** 143–155. MR0853049

[42] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838963

[43] Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J*. *Amer. Statist*. *Assoc*. **69** 169–173. MR0381120

[44] van Zuijlen, M.C.A. (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann*. *Probab*. **6** 250–266. MR0474624

[45] Volgushev, S. and Dette, H. (2012). Nonparametric quantile regression for twice censored data. Technical report. Available at arXiv:1007.3376v2 [stat.ME].

[46] Wang, J.G. (1987). A note on the uniform consistency of the Kaplan–Meier estimator. *Ann*. *Statist*. **15** 1313–1316. MR0902260

[47] Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J*. *Amer*. *Statist*. *Assoc*. **94** 137–145. MR1689219

[48] Ying, Z., Jung, S.H. and Wei, L.J. (1995). Survival analysis with median regression models. *J*. *Amer*. *Statist*. *Assoc*. **90** 178–184. MR1325125

[49] Yu, K. and Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimators. *Comput*. *Statist*. *Data Anal*. **25** 159–166.

[50] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J*. *Amer*. *Statist*. *Assoc*. **93** 228–237. MR1614628

Received March 2010 and revised June 2012