Nonparametric quantile regression for twice censored data

Stanislav Volgushev, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

FAX: +49 234 3214 559

September 16, 2009

Abstract

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.

AMS Subject Classification: 62G08, 62N02, 62E20

Keywords and Phrases: quantile regression, crossing quantile curves, censored data, monotone rearrangements, survival analysis, Beran estimator

1 Introduction

Quantile regression offers great flexibility in assessing covariate effects on event times. The method was introduced by Koenker and Bassett (1978) 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 Koenker (2005)]. Recently Koenker and Geling (2001) 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 Portnoy (2003)]. 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 Powell (1984, 1986), 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 for example Fitzenberger (1997)]. Even worse, recent research indicates that using the information contained in the observed censored data actually reduces the estimation accuracy [see Koenker (2008)].

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, Ying et al. (1995) assumed that the responses and censoring times are independent, which is stronger than the usual assumption of conditional independence. Yang (1999) proposed a method for median regression under the assumption of i.i.d. errors, which is computationally difficult to evaluate and cannot be directly generalized to the heteroscedastic case. Recently, Portnoy (2003) 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 Efron (1967)] and can be considered as a direct generalization of this classical estimate in survival analysis. Peng and Huang (2008) 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 Chaudhuri (1991) and Yu and Jones (1997, 1998). Chernozhukov et al. (2006) and Dette and Volgushev

(2008) 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 Dabrowska (1987, 1989) or Li and Doss (1995). The estimation of conditional quantile functions in the same setting is briefly stressed in Dabrowska (1987) and further elaborated in Dabrowska (1992a), while El Ghouch and Van Keilegom (2008) 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 Turnbull (1974). More recent references are Chang and Yang (1987); Chang (1990); Gu and Zhang (1993) and Patilea and Rolin (2006). On the other hand- to their best knowledge- the authors are not aware of literature on nonparametric conditional quantile regression 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 Patilea and Rolin (2006) 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 Dette et al. (2005)] and monotone rearrangements [see Dette et al. (2006), Chernozhukov et al. (2006) or Anevski and Fougères (2007) 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 an 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 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 $F_R(.|x)$ and $F_T(.|x)$ 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 $F_T^{-1}(.|x)$. 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

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

An unconditional version of this censoring mechanism was introduced by Patilea and Rolin (2006), and some applications of this model can also be found in the corresponding paper. 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 $F_L(.|x)$ 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 Dette et al. (2005)] and the monotone rearrangements, which was recently introduced by Dette et al. (2006) 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 $H_k(A|x) = \mathrm{P}\Big(A \cap \{\delta = k\}|X = x\Big)$ and obtain the decomposition $H = H_0 + H_1 + H_2$ for the conditional distribution of Y_i. The subdistribution functions H_k $(k = 0, 1, 2)$ can be represented as follows

$$H_0(dt|x) \;=\; F_L(t-|x)(1 - F_R(t-|x))F_T(dt|x) \tag{2.2}$$

$$(2.3) \qquad H_1(dt|x) = F_L(t-|x)(1-F_T(t|x))F_R(dt|x)$$

$$(2.4) \qquad H_2(dt|x) = \{1-(1-F_T(t|x))(1-F_R(t|x))\}\, F_L(dt|x) = F_S(t|x)F_L(dt|x).$$

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

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

where the quantities $W_i(x)$ denote local weights depending on the covariates $X_1, ..., X_n$, 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

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

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

$$(2.7) \qquad F_L(t|x) = \prod_{(t,\infty]} (1 - M_2^-(ds|x))$$

[see e.g. Patilea and Rolin (2006)]. Now having a representation for the conditional distribution function F_L we can define in a second step

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

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

$$(2.9) \qquad 1 - F_T(t|x) = \prod_{[0,t]} (1 - \Lambda_T^-(ds|x))$$

[see e.g. Gill and Johansen (1990)]. Note that formula (2.9) yields an explicit representation of the conditional distribution function $F_T(.|x)$ 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 $F_L(.|x)$, that is

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

where

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

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

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

where

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

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

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

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 $F(.|x)$ are not necessarily increasing for finite sample sizes [see e.g. Yu, Jones (1998)]. 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 Koenker (2005) or Yu and Jones (1998)] 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 Dette and Volgushev (2008) 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

$$\Psi : \begin{cases} L^\infty(J) \to L^\infty(\mathbb{R}) \\ f \mapsto \big(y \mapsto \int_J I_{\{f(u) \le y\}} du\big) \end{cases} \tag{2.14}$$

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), i.e.

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

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 Dette and Volgushev (2008). In the present paper we use different approach which is a slightly modified version of the ideas from Anevski and Fougères (2007). To be precise note that estimators of the conditional distribution function $F(.|x)$ [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 $F_{T,n}(.|x)$ implies that $F_{T,n}(.|x)$ will also be constant outside of J. We thus propose to consider the modified operator $\tilde{\Psi}_J$ defined as

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

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

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

2. Use the concept of increasing rearrangements [see Dette et al. (2006) and Chernozhukov et al. (2006) 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

$$\Phi : \begin{cases} L^\infty(J) \to L^\infty(I\!R) \\ f \mapsto (y \mapsto (\Psi f(.))^{-1}(y)) \end{cases} \tag{2.18}$$

where Ψ is introduced in (2.14). Note that for a strictly increasing right continuous function f this operator reproduces f, i.e. $\Phi(f) = f$. On the other hand, if f is not isotone, $\Phi(f)$ is an isotone function and the operator preserves the L^p-norm, i.e.

$$\int_J |\Phi(f(u))|^p \, du = \int_J |f(u)|^p \, du.$$

Moreover, the operator also defines a contraction, i.e.

$$\int_J |\Phi(f_1)(u) - \Phi(f_2)(u)|^p \, du \le \int_J |f_1 - f_2|^2 \, du \quad \forall p \ge 1$$

[see Hardy et al. (1988) or Lorentz (1953)]. 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 Bennett and Sharpley (1988), while some statistical applications can be found in Dette et al. (2006) and Chernozhukov et al. (2006).

The idea is now to use rearranged estimators of $H_i(.|x)$ and $H(.|x)$ 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

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

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

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

  [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 $H_n^{IP}(t|x) := \tilde{\Phi}_{J_Y}(\tilde{H}_n(\cdot|x))(t)$ as estimator for $H(t|x)$.

- Observe that the estimator $H_n^{IP}(t|x)$ is by construction an increasing step function which can only jump in the points $t = Y_i$, i.e. it admits the representation

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

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

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

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

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:

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

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

Lemma 2.2 *Assume that $Y_i \neq Y_j$ for $i \neq j$. Then the function*

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

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

This in turn yields the estimate

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

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

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

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

Remark 2.3 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. Fleming and Harrington (1991), 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 Fleming and Harrington (1991), page 105 and Anderson et al. (1993), page 260]. When censoring is light, the first of the above mentioned approaches seems to yield the best results [see Anderson et al. (1993), 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 $\hat{q}^{IP}(\tau|x)$ are not defined as soon as $\sup_t F_{T,n}(t|x) < \tau$ or $\sup_t F_{T,n}^{IP}(t|x) < \tau$, respectively. A simple ad-hoc solution to this problem is to define the estimator $F_{T,n}$ or $F_{T,n}^{IP}$ 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 2.2.

Proof of Lemma 2.2 In order to see that $\tilde{\Lambda}_{T,n}^{-}(dt|x)$ is increasing, we note that

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

Thus $\tilde{F}_{L,n}(t-|x) - H_n^{IP}(t-|x) \geq 0$ and the nonnegativity of $\tilde{\Lambda}_{T,n}^{-}(dt|x)$ is established. In order to prove the inequality $\tilde{\Lambda}_{T,n}^{-}(dt|x) \leq 1$ 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 $k \geq 2$

$$\begin{aligned} & \tilde{F}_{L,n}(Y_k-|x) - H_n^{IP}(Y_k-|x) \\ = & \Big[1-\prod_{[Y_k,\infty)}\Big(1-\frac{H_{0,n}^{IP}(ds|x)+H_{1,n}^{IP}(ds|x)}{H_n^{IP}(s|x)}\Big)\Big]\prod_{[Y_k,\infty)}\Big(1-\frac{H_{2,n}^{IP}(ds|x)}{H_n^{IP}(s|x)}\Big) \\ \overset{(*)}{=} & \Big[1-\prod_{j\geq k,\delta_j\neq 2}\Big(1-\frac{\Delta H_{0,n}^{IP}(Y_j|x)+\Delta H_{1,n}^{IP}(Y_j|x)}{H_n^{IP}(Y_j|x)}\Big)\Big]\prod_{j\geq k+1,\delta_j=2}\Big(1-\frac{\Delta H_{2,n}^{IP}(Y_j|x)}{H_n^{IP}(Y_j|x)}\Big) \\ = & \Big[1-\prod_{j\geq k,\delta_j\neq 2}\Big(\frac{H_n^{IP}(Y_{j-1}|x)}{H_n^{IP}(Y_j|x)}\Big)\Big]\prod_{j\geq k+1,\delta_j=2}\Big(\frac{H_n^{IP}(Y_{j-1}|x)}{H_n^{IP}(Y_j|x)}\Big) \end{aligned}$$

$$
\begin{aligned}
&\overset{(**)}{=} \Big[1 - \frac{H_n^{IP}(Y_{k-1}|x)}{H_n^{IP}(Y_k|x)} \prod_{j\geq k+1,\delta_j\neq 2} \Big(\frac{H_n^{IP}(Y_{j-1}|x)}{H_n^{IP}(Y_j|x)}\Big)\Big] \prod_{j\geq k+1,\delta_j=2} \Big(\frac{H_n^{IP}(Y_{j-1}|x)}{H_n^{IP}(Y_j|x)}\Big) \\
&\geq \Big[1 - \frac{H_n^{IP}(Y_{k-1}|x)}{H_n^{IP}(Y_k|x)}\Big] \prod_{j\geq k+1} \Big(\frac{H_n^{IP}(Y_{j-1}|x)}{H_n^{IP}(Y_j|x)}\Big) \\
&= \Big[\frac{H_n^{IP}(Y_k|x) - H_n^{IP}(Y_{k-1}|x)}{H_n^{IP}(Y_k|x)}\Big] \frac{H_n^{IP}(Y_k|x)}{H_n^{IP}(Y_n|x)} \\
&= \Delta H_n^{IP}(Y_k|x),
\end{aligned}
$$

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 $\Delta\tilde{\Lambda}_{T,n}^-(Y_k|x) \leq 1$, 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 Gill (1989)] 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 Patilea and Rolin (2001)], 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

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

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 $X_1, ..., X_n$ but not on $Y_1, ..., Y_n$ or $\delta_1, ..., \delta_n$. The first set of conditions concerns the weights that are used in the representation (2.5).

(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 functions V_j $(j = 1, \ldots, n)$ 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 $\underline{c}/nh \leq V_j(x) \leq \overline{c}/nh$

(2) If $|x - X_j| \leq Ch$ for some constant $C < \infty$, then $V_j(x) \neq 0$ and $V_j(x) = 0$ for $|x - X_j| \geq c_n$ for some sequence $(c_n)_{n\in\mathbb{N}}$ such that $c_n = O(h)$. Without loss of generality, we will assume that $C = 1$ throughout this paper.

(3) $\sum_i V_i(x) = C(x)(1 + o_P(1))$ for some positive function C

(4) $\sup_t \left|\sum_i V_i(x)(x - X_i)I_{\{Y_i \leq t\}}\right| = o_P(1/\sqrt{nh})$

Here [and throughout this paper] h denotes a smoothing parameter converging to 0 with increasing sampling size.

(W2) We assume that the weak convergence

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

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

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

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 Pollard (1984)].

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

Remark 3.1 It will be shown in Lemma 3.3 below that important examples for weights satisfying conditions (W1)-(W3) are given by the Nadaraya-Watson weights

$$W_i^{NW}(x) = \frac{\frac{1}{nh}K_h(x - X_i)}{\frac{1}{nh}\sum_j K_h(x - X_j)} =: \frac{V_i^{NW}(x)}{\sum_j V_j^{NW}(x)}, \tag{3.2}$$

or by the local linear weights

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

where $K_h(.) := K(./h)$, $S_{n,k} := \frac{1}{nh}\sum_j K_h(x - X_j)(x - X_j)^k$ 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 $c_1 \leq K(x) \leq c_2$ for all x with $K(x) \neq 0$ for some constants $0 < c_1 \leq c_2 < \infty$.

For the distributions of the random variables (T_i, L_i, R_i, X_i) we assume that for some $\varepsilon > 0$:

(D1) The conditional distribution function F_R fulfills $F_R(t_{00}|x) < 1$

(D2) The conditional distribution functions $F_L(.|x), F_R(.|x), F_T(.|x)$ are continuous

(D3) For $i = 0, 1, 2$ we have $\lim_{y \to x} \sup_t |H_i(t|y) - H_i(t|x)| = 0$

(D4) The conditional distribution functions $F_L(.|x), F_R(.|x), F_T(.|x)$ have densities, say $f_L(.|x), f_R(.|x), f_T(.|x)$, with respect to the Lebesque measure

(D5) $\int_{t_{00}}^{\infty} \frac{f_L(u|x)}{F_L^2(u|x)F_S(u|x)} du < \infty$

(D6) $\int_{t_{00}}^{\infty} \frac{1}{F_L(u|x)F_S(u|x)} \left| \partial_x \frac{f_L(u|x)}{F_L(u|x)} \right| du < \infty$

(D7) $\sup_{(t,z) \in (t_{00}, \infty) \times U_\varepsilon(x)} \left| \partial_z^2 \frac{f_L(t|z)}{F_L(t|z)} \right| < \infty$

(D8) 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

$$\sup_t \sup_{|y-x|<\varepsilon} |\partial_y^2 H_k(t|y)| < \infty$$

(D9) The distribution function F_X of the covariates X_i is twice continuously differentiable with density f_X such that $f_X(x) \neq 0$

(D10) 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 an interval of positive length with $x \in I$.

(D11) $\frac{f_L(t|y)}{F_L(t|y)} = \frac{f_L(t|x)}{F_L(t|x)}(1 + o(1))$ uniformly in $t \in [t_{00}, t_{00} + \varepsilon)$ as $y \to x$

(D12) 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 (D5) or (D12) the equality $t_{00} = \tau_{T,0}(x)$ holds.

Finally, we make some assumptions for the smoothing parameter

(B1) $n \log(n) h^5 = o(1)$ and $nh \longrightarrow \infty$.

(B2) $h \to 0$ and $nh/\log(n) \longrightarrow \infty$.

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 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. *Under assumptions (D8), (D9) and (B1) condition (W2) holds for the Nadaraya-Watson, local linear or rearranged local linear weights based on a positive, symmetric kernel with compact support.*

3. *Under assumptions (B2), (D2), (D3) condition (W3) holds for the Nadaraya-Watson weights W_i based on a positive, symmetric kernel with compact support. If additionally 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.*

Note that the assumption (B1) does not allow to choose $h \sim n^{-1/5}$, 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 Dabrowska (1987) or Li and Doss (1995). In principle, most results of the present paper can be extended to bandwidths $h \sim n^{-1/5}$ if a corresponding bias term is substracted.

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*

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

The Lemma follows from the relation

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

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

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

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 $\hat{q}, \hat{q}^{IP}$ 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 $F_{T,n}(.|x)$ under some rather weak conditions. In particular, it does neither require the existence of densities of the conditional distribution functions [see (D4)] nor integrability conditions like (D5).

Theorem 3.5 *If conditions (D1), (D2), (D12), (W1)(1)-(W1)(2) and (W3) are satisfied, then the following statements are correct.*

1. *The estimate $F_{T,n}(.|x)$ 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 $F_{T,n}(.|x)$ is increasing and takes values in the interval $[0, 1]$, the weak uniform consistency of the estimate $F_{T,n}(.|x)$ 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 (D4)-(D11) hold. Then we have as $n \to \infty$*

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

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

$$(3.4) \qquad A(t) = \int_t^\infty \frac{dG_2(u)}{H(u|x)}, \qquad B(t) := \int_t^\infty \frac{G(u)}{H^2(u|x)} H_2(du|x).$$

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*

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

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 (D5) and the representation

$$\mathrm{E}[G_M(s)G_M(t)] = b(x)\int_{s\vee t}^{\infty} \frac{1}{H(u|x)} M_2^-(du|x)$$

for the covariance structure of G_M, which can be derived by computations similar to those in Patilea and Rolin (2001).

Theorem 3.8 *Assume that the conditions of Theorem 3.6 and condition (D12) are satisfied. Moreover, let* $t_{00} < \tau$ *such that* $F_S([0,\tau]|x) < 1$*. Then we have the following weak convergence*

1.

$$\sqrt{nh}(\Lambda_{T,n}^- - \Lambda_T^-) \Rightarrow V$$

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

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

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.

$$\sqrt{nh}(F_{T,n} - F_T) \Rightarrow W$$

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

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

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 $\hat{q}^{IP}$ defined in (2.17) and (2.25), respectively. As an immediate consequence of Theorem 3.5 and the continuity of the quantile mapping [see Gill (1989), 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 $\hat{q}(.|x)$ and $q^{IP}(.|x)$ 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 $\inf_{\varepsilon \le t \le \tau} f_T(t|x) > 0$*

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

where Z is a centered Gaussian process defined by

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

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

The proof Theorem 3.5 - 3.10 is presented in the Appendix A and requires 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 Dabrowska (1987) 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 Beran (1981) this can be done by estimating F_Z, the conditional distribution function of Z, and $\pi_k(t|x) := \mathrm{P}\Big(Z_i \le t, \Delta_i = k|X = x\Big)$ $(k = 0, 1)$ through

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

and then defining an estimator for F_D as

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

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

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

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

The weak convergence of the process $\sqrt{nh}(F_{D,n}(t|x) - F_D(t|x))_t$ in $D([0, \tau])$ with $\pi_0(\tau|x) < 1$ was first established by Dabrowska (1987). 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 Gill (1983) who used counting process techniques. A generalization of this method to the conditional case was first considered by McKeague and Utikal (1990) and later exploited by Dabrowska (1992b) and Li and Doss (1995). 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 $F_D(.|x)$ *and* $F_B(.|x)$ *have densities, say* $f_D(.|x)$ *and* $f_B(.|x)$*, with respect to the Lebesque measure*

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

(R3) $\int_0^{t_0} \frac{|\partial_x \lambda_D(t|x)|}{1 - F_Z(t-|x)} dt < \infty$,

(R4) $\sup_{(t,y) \in (0,t_0) \times U_\varepsilon(x)} \left|\partial_y^2 \lambda_D(t|y)\right| < \infty$,

(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 an interval of positive length with* $x \in I$*,*

(R6) $\lambda_D(t|y) = \lambda_D(t|x)(1 + o(1))$ *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

$$\sqrt{nh}(F_{Z,n}(.|x) - F_Z(.|x), \pi_{0,n}(.|x) - \pi_0(.|x)) \Rightarrow (G, G_0) \quad \textit{on} \quad D([0, \infty))$$

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

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

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)*

$$\sqrt{nh}(F_{D,n}(.|x) - F_D(.|x))_t \Rightarrow G_D(.) \quad \textit{in} \quad D([0, t_0]), \tag{3.8}$$

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

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

4 Finite sample properties

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 Dabrowska (1992b) and Li and Datta (2001). In order to obtain a reasonable bandwidth parameter for our simulations, we used a modification of the cross validation procedure proposed by Abberger (2001) 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 $\{(Y_{jk}, X_{jk}, \delta_{jk})|\ j = 1, \ldots, J_k\}$ denote the points among $\{(Y_i, X_i, \delta_i)|\ i = 1, \ldots, n\}$ which fall in block k $(k = 1, \ldots, K)$. 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
$$h := \text{argmin}_\alpha \sum_{k=1}^{K} \sum_{j=1}^{J_k} w_{jk} \rho_\tau (Y_{jk} - \tilde{q}_\alpha^{j,k}(\tau | X_{jk}))$$
where ρ_τ denotes the check function and $\tilde{q}_\alpha^{j,k}$ is either the estimator $\hat{q}^{IP}$ or $\hat{q}$ with bandwidth α based on the sample $\{(Y_i, X_i, \delta_i)|\ i = 1, \ldots, n\}$ 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, i.e.

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

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 models 2 and 3 in Yu and Jones (1997) [note that we additionally introduce a censoring mechanism]. The first model is given by

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

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 Figures 2 and 1 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 $\hat{q}^{IP}(\tau|x)$ 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 $\hat{q}^{IP}$ 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

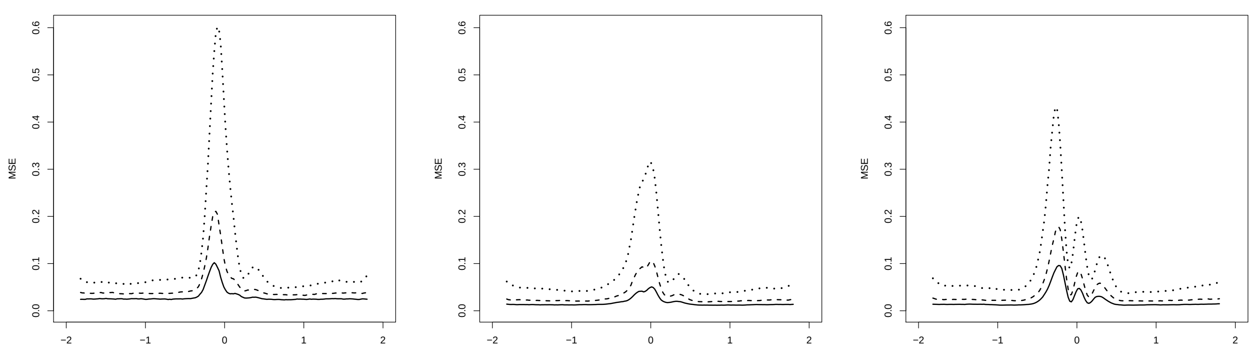

Figure 1: *Mean squared error curves of the estimates of the quantile curves in model 1 for different sample sizes:* $n = 100$ *(dotted line);* $n = 250$ *(dashed line);* $n = 500$ *(solid line). Left panel: estimates of the* 25%*-quantile curves; middle panel: estimates of the* 50%*-quantile curves;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

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.

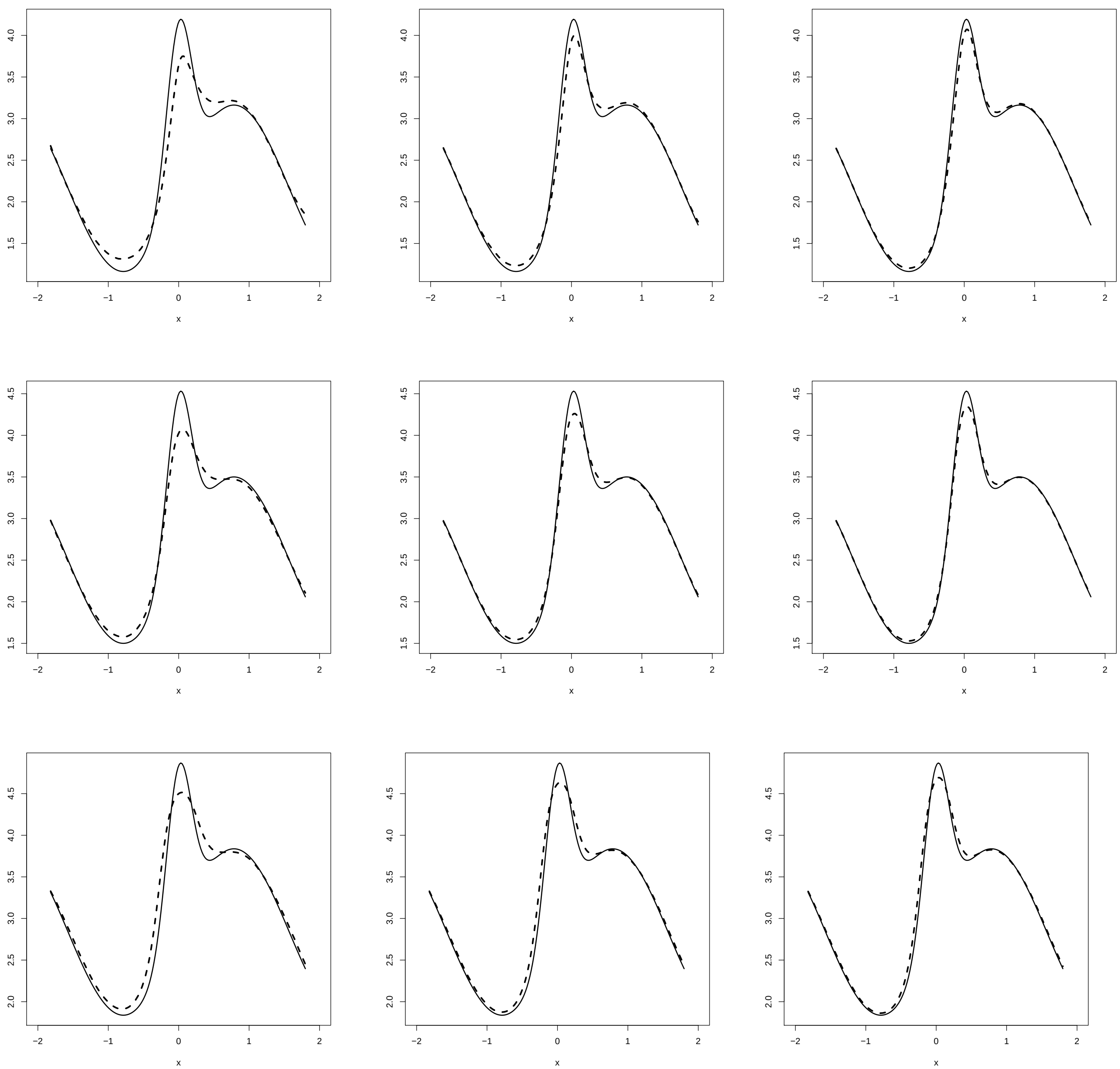

Figure 2: *Mean (dashed lines) and true (solid lines) quantile curves for model 1 for different sample sizes:* $n = 100$ *(left column),* $n = 250$ *(middle column) and* $n = 500$ *(right column). Upper row: estimates of the* 25% *quantile curves; middle row: estimates of the* 50% *quantile curves; lower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

As a second example we investigate the effect of different censoring types. To this end, we consider a similar example as in model 3 of Yu and Jones (1997), that is

$$\text{(model 2)} \qquad \begin{cases} T_i = 2 + 2\cos(X_i) + \exp(-4X_i^2) + \mathcal{E}(1) \\ L_i = 2 + 2\cos(X_i) + \exp(-4X_i^2) + (c_L + \mathcal{U}[0,1]) \\ R_i = 2 + 2\cos(X_i) + \exp(-4X_i^2) + (c_R + \mathcal{E}(1)) \end{cases}$$

where the covariates X_i are uniformly distributed on the interval $[-2, 2]$, $\mathcal{E}(1)$ denotes an exponentially distributed random variable with parameter 1, $\mathcal{U}[0, 1]$ is a uniformly distributed random variable on $[0, 1]$ and the parameters (c_L, c_R) are used to control the amount of censoring. For this purpose we investigate three different cases for the parameters (c_L, c_R), namely $(-0.5, 1.5)$, $(-0.5, 0.5)$ and $(-0.2, 1.5)$, which corresponds to approximately $(10\%, 11\%)$, $(30\%, 11\%)$ and $(11\%, 25\%)$ of type $\delta = 1$ and $\delta = 2$ censoring, respectively. The corresponding results for the estimators of the 25%, 50% and 75% quantile on the basis of a sample of $n = 250$ observations are presented in Figures 3 and 4.

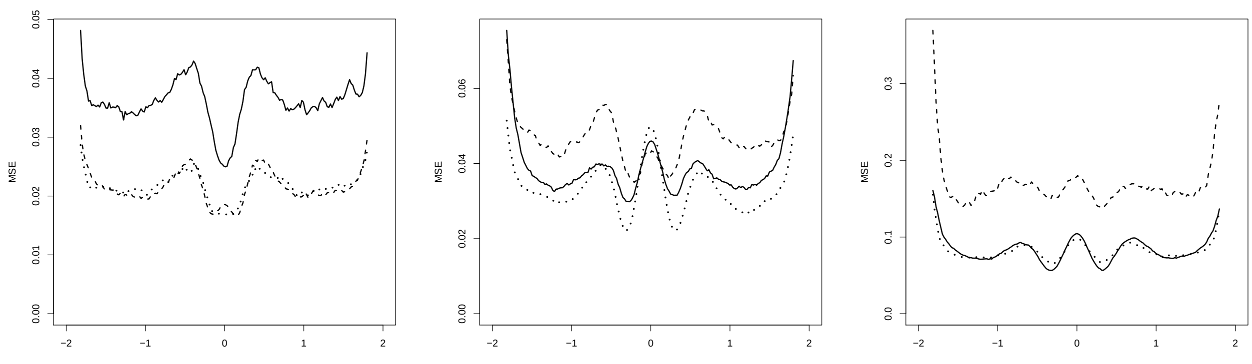

Figure 3: *Mean squared error curves of the estimates of the quantile curves in model 2 for different censoring:* $(10\%, 11\%)$ *censoring (dotted line);* $(30\%, 11\%)$ *censoring (dashed line);* $(11\%, 25\%)$ *censoring (solid line). Left panel: estimates of the* 25%*-quantile curves; middle panel: estimates of the* 50%*-quantile curves; right panel: estimates of the* 75%*-quantile curves. The sample size is* $n = 250$.

We observe a slight increase in bias when estimating upper quantile curves. An additional amount of censoring results in a slightly worse average behavior of the estimates. More censoring of type $\delta = 2$ has an impact on the accuracy of the estimates of the lower quantiles, while more censoring of type $\delta = 1$ has a stronger effect for the upper quantile curves. Upper quantile curves are always estimated with more variability which is in accordance with the factor $1/f_T(F_T^{-1}(p|x)|x)$ in their limiting process.

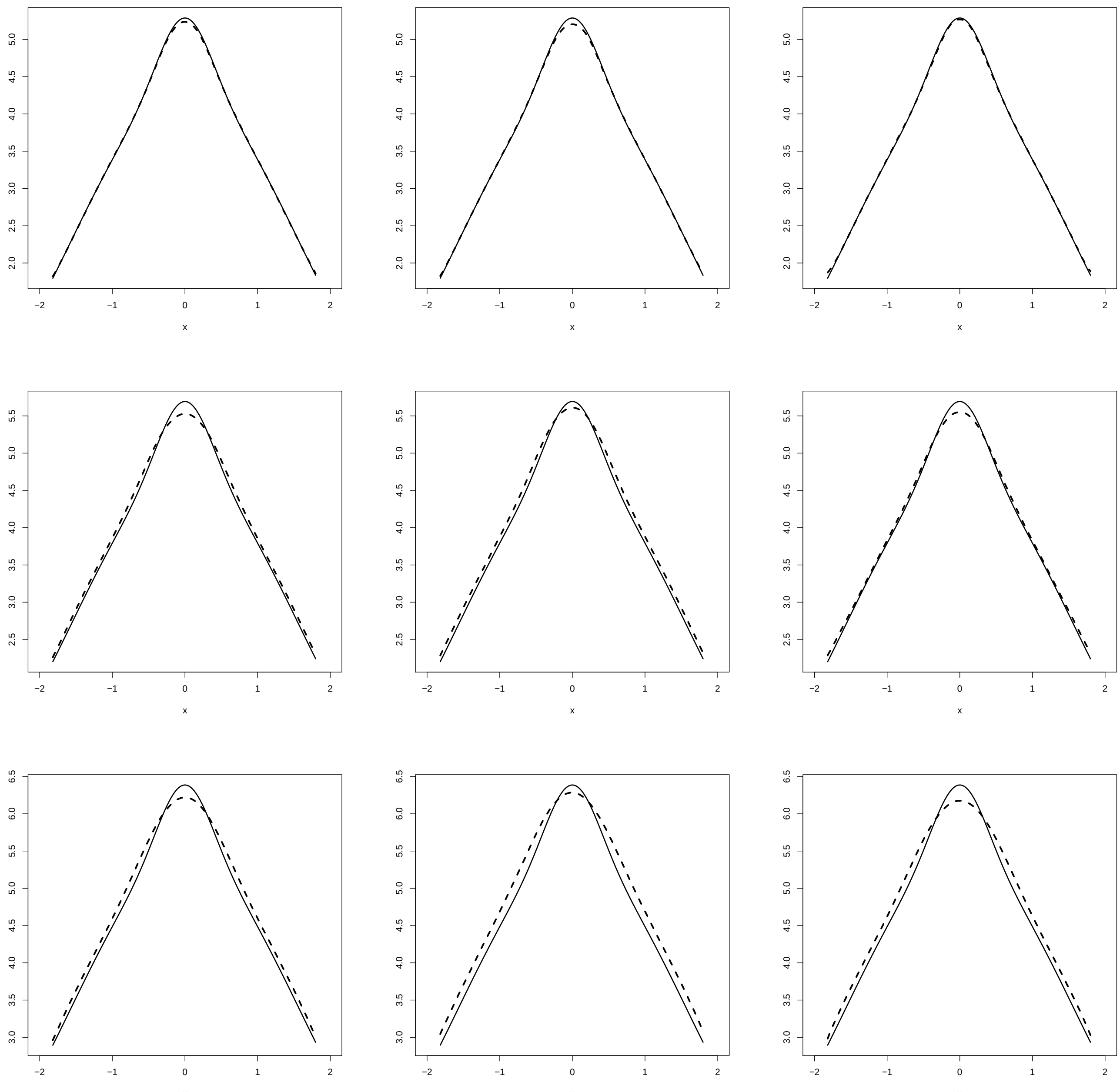

Figure 4: *Mean (dashed lines) and true (solid lines) quantile curves for model 2 and different censoring: left column:* $(10\%, 11\%)$ *censoring; middle column:* $(30\%, 11\%)$ *censoring; right column:* $(11\%, 25\%)$ *censoring. Upper row:* 25% *quantile curves; middle row:* 50% *quantile curves; lower row:* 75% *quantile curves. The sample sizes is* 250.

Acknowledgements. The authors are grateful to Martina Stein who typed parts of this paper with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) and in part by an NIH grant award IR01GM072876:01A1.

A Appendix: Proofs

Proof of Lemma 3.3 We begin with the proof of the first part. Recalling the definition of the Nadaraya-Watson weights in (3.2), we see that (W1)(1) follows easily from the inequality $c_1 \leq K(x) \leq c_2$ for all x in the support of K. Conditions (W1)(2) and (W1)(3) hold with $C(x) = f_X(x)$, which is a standard result from density estimation [see e.g. Parzen (1962)]. Finally, for assumption (W1)(4) we note that, as soon as the function $f_X(.)F_Y(t|.)$ is continuously differentiable in a neighborhood of x with uniformly (in t) bounded derivative, we have

$$\sup_t \Big| \frac{1}{nh} \mathrm{E}\Big[\sum_i K_h(x - X_i)(x - X_i) I_{\{Y_i \leq t\}}\Big]\Big| = O(h^2).$$

From standard empirical process arguments [see for example Pollard (1984)] we therefore obtain

$$\sup_t \frac{1}{nh} \Big| \sum_i K_h(x - X_i)(x - X_i) I_{\{Y_i \leq t\}} - \mathrm{E}\Big[\sum_i K_h(x - X_i)(x - X_i) I_{\{Y_i \leq t\}}\Big]\Big| = O\Big(\sqrt{\frac{h \log n}{n}}\Big)$$

a.s. and the assertion now follows from condition (B1).

To see that we can also use the local linear weights defined in (3.3), we note that

$$S_{n,0} = f_X(x)(1 + o_P(1)) \tag{A.1}$$

$$S_{n,1} = h^2 \mu_2(K) f_X'(x) + o_P(h^2), \tag{A.2}$$

$$S_{n,2} = h^2 \mu_2(K) f_X(x) + o_P(h^2) \tag{A.3}$$

and from the compactness of the support of K, which implies: $|x - X_j| = O(h)$ uniformly in j, we obtain the representation $V_i^{LL} = V_i^{NW}(1 + o_P(1))$ uniformly in i. Conditions (W1)(1) and (W1)(4) for the local linear follow from the corresponding properties of the Nadaraya-Watson weights (possibly with slightly smaller and larger constants $\underline{c}$ and $\overline{c}$, respectively).

Finally, from the fact that, with probability tending to one, the local linear weights are positive, it follows that the corresponding estimators H_n, H_{ni} are increasing and hence unchanged by the rearrangement. This implies $\mathrm{P}\Big(\exists i \in 1, ..., n : W_i^{LL} \neq W_i^{LLI}\Big) \stackrel{n\to\infty}{\longrightarrow} 0$, where W_i^{LLI} denote the weights of the rearranged local linear estimator. Thus condition (W1) also holds for the weights W_i^{LLI} and the proof of the first part is complete.

For a proof of the second part of the Lemma we note that the same arguments as given in Dabrowska (1987), Section 3.2, yield condition (W2) for the Nadaraya-Watson weights [here we used assumptions (D8), (D9) and (B1)].

The corresponding result for the local linear weights can be derived by a closer examination of the weights W_i^{LL}. For the sake of brevity, we will only consider the estimate H_n defined in (2.5), the results for $H_{k,n}$ $(k = 0, 1, 2)$ follow analogously. From the definition of the weights W_i^{LL} we obtain the representation

$$\begin{aligned} H_n^{LL}(t|x) &= \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h}\right)(S_{n,2}-(x-X_i)S_{n,1})}{S_{n,2}S_{n,0}-S_{n,1}^2} I_{\{Y_i \le t\}} \\ &= \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h}\right)}{S_{n,0}} \frac{1}{1-S_{n,1}^2/(S_{n,0}S_{n,2})} I_{\{Y_i \le t\}} - \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h}\right)(x-X_i)S_{n,1}}{S_{n,2}S_{n,0}-S_{n,1}^2} I_{\{Y_i \le t\}} \\ &= H_n^{NW}(t|x) + O_P(h^2) \end{aligned}$$

uniformly in t where the last equality follows from the estimates $H_n^{NW}(t|x) = O_P(1)$ and (A.1) - (A.3). Now condition (B1) ensures $h^2 = o(1/\sqrt{nh})$ and thus the difference $H_n^{NW} - H_n^{LL}$ is asymptotically negligible. From Lemma 3.4 we immediately obtain that, with probability tending to one, the rearranged estimators H_n^{LLI} and $H_{i,n}^{LLI}$ defined in (2.20) and (2.21) coincide with the estimates H_n^{LL} and $H_{i,n}^{LL}$ respectively. Thus condition $(W2)$ also holds for $(H_n^{LLI}, H_{0,n}^{LLI}, H_{2,n}^{LLI})$ and the second part of Lemma 3.3 has been established.

We now turn to the proof of the last part. Again we only consider the process $H_n(.|x)$, and note that the uniform consistency of $H_{k,n}(.|x)$ follows analogously. First, observe the estimate

$$\mathrm{E}\Big[\frac{1}{nh}\sum_i K_h(x-X_i)I_{\{Y_i\le t\}}\Big] = \frac{1}{h}\int K_h(x-u)F_Y(t|u)f_X(u)du = f_X(x)F_Y(t|u)(1+o(1))$$

uniformly in t, which is a consequence of condition (D3). From standard empirical process arguments [see Pollard (1984)] it follows that almost surely

$$\sup_t\Big|\frac{1}{nh}\sum_i K_h(x-X_i)I_{\{Y_i\le t\}} - \mathrm{E}\Big[\frac{1}{nh}\sum_i K_h(x-X_i)I_{\{Y_i\le t\}}\Big]\Big| = O\Big(\sqrt{\frac{\log n}{nh}}\Big),$$

and with condition (B2) the assertion for the Nadaraya-Watson weights follows. The extension of the result to local linear and rearranged local linear weights can be established by the same arguments as presented in the second part of the proof. □

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

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

where $W_i^{NW}(x)$ denote Nadaraya-Watson weights, we obtain: $\mathrm{P}(\exists i \in 1,...,n : \tilde{W}_i \neq W_i) \to 0$, i.e. 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 $\tilde{V}_i := V_i I_{A_n} + V_i^{NW}(x) I_{A_n^C}$. Note that by construction, the random variables $\tilde{V}_i$ satisfy conditions (W1)(1)-(W1)(4) if the kernel in the definition of $W_i^{NW}(x)$ satisfies assumption (K1).

<u>Proof of Theorem 3.5</u>: Let S denote the set of pairs of functions $(H_2(.|x), H(.|x))$ of bounded variation such that $H(.|x) \geq \beta > 0$. Since the map $(H_2(.|x), H(.|x)) \mapsto M_2^-(.|x)$ is continuous on S with respect to the supremum norm [see the discussion in Anderson et al. (1993) following Proposition II.8.6], and H_n is uniformly consistent [which implies $\mathrm{P}((H_{2,n}, H_n) \in S] \to 1]$, the weak uniform consistency of M_{2n}^- on $[t_{00} + \varepsilon, \infty)$ [$\varepsilon > 0$ is arbitrary] follows from the uniform consistency of $H_{2,n}$ and H_n. This can be seen by similar arguments as given in Dabrowska (1987), p. 184.

Moreover, the map $M_2^-(.|x) \mapsto F_L(.|x)$ is continuous on the set of functions of bounded variation [reverse time and use the discussion in Andersen et.al. (1993) following Proposition II.8.7], and thus the uniform consistency of $F_{L,n}(.|x)$ on $[t_{00} + \varepsilon, \infty)$ follows for any positive $\varepsilon > 0$.

In the next step, we consider the map

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

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

$$\text{(A.4)} \quad \sup_{t \in [t_{00}+\varepsilon, \tau)} \left| \int_{[t_{00}+\varepsilon, t)} \frac{H_{0,n}(dt|x)}{F_{L,n}(t-|x) - H_n(t-|x)} - \int_{[t_{00}+\varepsilon, t)} \frac{H_0(dt|x)}{F_L(t-|x) - H(t-|x)} \right| \xrightarrow{P} 0$$

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 (D12) and continuity of the conditional distribution function $F_L(.|x)$]. 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 $W_1(x,n), ..., W_k(x,n)$ those values of $Y_1, ..., Y_n$, whose weights fulfill $W_i(x) \neq 0$ and by $W_{(1)}(x,n), ..., W_{(k)}(x,n)$ the corresponding increasingly ordered values. By Lemma B.2 in Appendix B we can find an $\varepsilon > 0$ such that:

$$\sup_{t_{00}+\varepsilon\geq t\geq W_{(2)}(x,n)} \frac{1}{F_{L,n}(s-|x) - H_n(s-|x)} = O_P(1),$$

and it follows

$$\int_{[W_{(2)}(x,n),t_{00}+\varepsilon)} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \leq H_{0,n}(t_{00}+\varepsilon|x)O_P(1).$$

Therefore it remains to find a bound for the integral $\int_{[0,W_{(2)}(x,n))} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x)-H_n(s-|x)}$. For this purpose we consider two cases. The first one appears if the δ_i corresponding to $W_{(1)}(x,n)$ equals 0. In this case there is positive mass at the point $W_{(1)}(x,n)$ but at the same time $F_{L,n}(s|x) = F_{L,n}(W_{(2)}(x,n)|x)$ for all $s \in [0, W_{(2)}(x,n))$ and hence $\int_{[0,t_{00}+\varepsilon)} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x)-H_n(s-|x)} \leq H_{0,n}(t_{00} + \varepsilon|x)O_P(1)$. For all other values of the corresponding δ_i the mass of $H_{0,n}(ds|x)$ at the point $W_{(1)}(x,n)$ equals zero and thus the integral vanishes. Summarizing, we have obtained the estimate

$$\int_{[0,t_{00}+\varepsilon)} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \leq H_{0,n}(t_{00}+\varepsilon|x)O_P(1) = H_0(t_{00}+\varepsilon|x)O_P(1),$$

where the last equality follows from the uniform consistency of $H_{0,n}$ and the remainder $O_P(1)$ does not depend on ε. Moreover, since the function $\Lambda_{T,n}(.|x)$ is increasing [see Lemma 2.2], the inequality

$$\text{(A.5)} \qquad \sup_{t\leq t_{00}+\varepsilon} |\Lambda_{T,n}(t|x)| = \int_{[0,t_{00}+\varepsilon)} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \leq H_0(t_{00}+\varepsilon|x)O_P(1)$$

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 we have

$$\mathrm{P}\Big(\sup_{t\in[0,t_{00}+\varepsilon_\delta)} |\Lambda_{T,n}(t|x) - \Lambda_T(t|x)| > 2\alpha\Big) \leq \mathrm{P}\Big(\sup_{t\in[0,t_{00}+\varepsilon_\delta)} |\Lambda_{T,n}(t|x)| > \alpha\Big) \leq \mathrm{P}\Big(O_P(1) > \alpha/\delta\Big),$$

whenever $\Lambda_T(t_{00}+\varepsilon|x) < \alpha$, where the last inequality follows from (A.5) and the remainder $O_P(1)$ does not depend on α and δ. From this estimate we obtain for any τ with $F_S(\tau|x) < 1$

$$\mathrm{P}\Big(\sup_{t\in[0,\tau)} |\Lambda_{T,n}(t|x) - \Lambda_T(t|x)| > 4\alpha\Big) \leq \mathrm{P}\Big(\sup_{t\in[t_{00}+\varepsilon_\delta,\tau)} |\Lambda_{T,n}(t|x) - \Lambda_T(t|x)| > 2\alpha\Big) + \mathrm{P}\Big(O_P(1) > \alpha/\delta\Big).$$

By (A.4) The first probability on the right hand side of the inequality converges to zero as n tends to infinity for any $\alpha, \varepsilon_\delta > 0$, and the limit of the second one can be made arbitrarily small by choosing δ appropriately. Thus we obtain $\lim_{n\to\infty} \mathrm{P}\Big(\sup_{t\in[0,\tau)} |\Lambda_{T,n}(t|x) - \Lambda_T(t|x)| > 4\alpha\Big) = 0$, which implies the weak uniform consistency of $\Lambda_{T,n}(.|x)$ on the interval $[0,\tau)$.

Finally, the continuity of the mapping $\Lambda_T \mapsto F_T$ [see the discussion in Anderson et al. (1993) 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 Wang (1987). Note that, as soon as $F_{T,n}(.|x)$ is increasing and bounded by 1 from above, we have the inequality $\sup_{t\geq a} |F_{T,n}(t|x) - F_T(t|x)| \leq |F_{T,n}(a|x) - F_T(a|x)| + (1 - F_T(a|x))$. Thus

$$\sup_{t\geq 0} |F_{T,n}(t|x) - F_T(t|x)| \leq 2 \sup_{0\leq t\leq a} |F_{T,n}(t|x) - F_T(t|x)| + 2(1 - F_T(a|x)),$$

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. □

<u>Proof of Theorem 3.6:</u> The second part follows from the first one by the Hadamard differentiability of the map $A \mapsto \prod_{(t,\infty]}(1 - A(ds))$ in definition (2.10) [see Patilea and Rolin (2001), Lemma A.1] and the delta method [Gill (1989)]. 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 $F_L(.|x)$.

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 (D5) we obtain $F_L(t_{00}|x) > 0$, and the continuity of $F_L(.|x)$ yields $t_{00} > 0$.

Set $\varepsilon > 0$ and choose $\sigma > t_{00}$ such that $H(\sigma|x) > \varepsilon$. Recall that the map

$$(H, H_0, H_2) \mapsto (H, H_0, M_2^-)$$

is Hadamard differentiable on the domain $\tilde{D} := \{(A_1, A_2, A_3) \in BV_1^3([\sigma,\infty]) : A_1 \geq 0, A_3 \geq \varepsilon/2\}$ [see Patilea and Rolin (2001)] and takes values in $BV_C^3([\sigma,\infty])$. Here BV_C denotes the space of functions of bounded variation with elements uniformly bounded by the constant C. Moreover, assumption (W2) implies weak convergence and weak uniform consistency of the estimator H_n on $D([\sigma,\infty])$. Therefore $(H_{0,n}, H_{2,n}, H_n)$ will belong to the domain $\tilde{D}$ with probability tending to one if $n \to \infty$. Hence, we can define the random variable $\bar{H}_n := I_{A_n} H_n + I_{A_n^C}$ where $A_n :=$

$\{\inf_{t\in[\sigma,\infty]} H_n(t) \geq \varepsilon/2\}$, which certainly has the property $\bar{H}_n \geq \varepsilon/2$ on $[\sigma,\infty]$ almost surely. Now, since $\mathrm{P}(\bar{H}_n \neq H_n] = 1 - \mathrm{P}(A_n) \to 0$, the weak convergence result in (W2) continues to hold on $D^3([\sigma,\infty])$ with H_n replaced by $\bar{H}_n$. By the same argument, we may replace the H_n in the definition of $M_{2,n}^-$ by $\bar{H}_n$ without changing the asymptotics. Thus we can apply the delta method [see Gill (1989), Theorem 3] to $(H_{0,n}, H_{2,n}, \bar{H}_n)$ and deduce the weak convergence

$$\sqrt{nh}(H_n - H, H_{0,n} - H_0, M_{2,n}^- - M_2^-) \Rightarrow (G, G_0, G_{M_\sigma}) \quad \text{in} \quad \mathrm{D}^3([\sigma,\infty]).$$

To obtain the weak convergence in $D^3([t_{00},\infty])$, we apply a Lemma from Pollard (1984, page 70, Example 11). First define G_M as the pathwise 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 Pollard (1984). For arbitrary positive ε and δ we now have to find a $\sigma = \sigma(\delta,\varepsilon) > t_{00}$ such that

$$\text{(A.6)} \qquad P\left(\sup_{t_{00}<t\leq\sigma} |G_M(t)| \geq \delta\right) < \varepsilon$$

$$\text{(A.7)} \qquad \limsup_{n\to\infty} P\left(\sup_{t_{00}<t\leq\sigma} \sqrt{nh}\left|(M_{2,n}^- - M_2^-)(\sigma - |x) - (M_{2,n}^- - M_2^-)(t - |x)\right| \geq \delta\right) < \varepsilon.$$

Note that once we have found a σ such that $(A.7)$ holds, we can make σ smaller until $(A.6)$ is fulfilled with $(A.7)$ still holding. This is possible because for every $\delta > 0$, $\lim_{\sigma\downarrow t_{00}} P\left(\sup_{t_{00}<t\leq\sigma} |G_M(t)| \geq \delta\right) = 0$, which can be established as follows. Define the function $\kappa(t) := \int_t^\infty \frac{M_2^-(ds|x)}{H(s|x)}$ and denote by W_t a Brownian motion on $[0,\infty]$. Then we have

$$\mathrm{Cov}(\sqrt{b(x)}W_{\kappa(s)}, \sqrt{b(x)}W_{\kappa(t)}) = b(x)(\kappa(s) \wedge \kappa(t)) = b(x)\int_{s\vee t}^\infty \frac{M_2^-(ds|x)}{H(s|x)} = \mathrm{Cov}(G_M(s), G_M(t)),$$

where the last equality follows from Remark 3.7. Thus we have represented the process G_M in terms of a Brownian motion and the assertion follows from the finiteness of $\kappa(t_{00})$ [by assumption (D5)] and the properties of the Brownian motion.

In order to prove the existence of a constant σ that ensures (A.7), 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. Consider the random variables $B_i := a(S_i)$, $D_i := a(L_i)$, $Z_i := B_i \wedge D_i$ and $\Delta_i := I_{\{D_i \leq B_i\}} =$

$I_{\{S_i \le L_i\}}$. This is a conditional right censorship model with the useful property that $\Lambda_D^-(.|X_i)$, the predictable hazard function of D_i, is closely connected to the reverse hazard function $M_2^-(.|X_i)$ by the identity

$$\Lambda_D^-(a(t)|x) = M_2^-(\infty|x) - M_2^-(t-|x)$$

It is easy to verify that the conditional Nelson-Aalen estimator $\Lambda_{D,n}^-(dt|x)$ in the new model is related to the estimator $M_{2,n}^-$ in a similar way, i.e. $\Lambda_{D,n}^-(a(t)|x) = M_{2,n}^-(\infty|x) - M_{2,n}^-(t|x)$. Thus to prove $(A.7)$ it suffices to find a σ such that in the new model the following inequality is fulfilled

$$\text{(A.8)} \qquad \limsup_{n\to\infty} P\left(\sup_{\sigma \le t < t_0} \sqrt{nh}\left|(\Lambda_{D,n}^- - \Lambda_D^-)(t|x) - (\Lambda_{D,n}^- - \Lambda_D^-)(\sigma-|x)\right| > \delta\right) < \varepsilon,$$

where we define $t_0 = a(t_{00}) < \infty$. This assertion 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]. □

Proof of Theorem 3.8: First of all note that the a.s. continuity of the sample paths of the processes $V(.)$ and $W(.)$ 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

$$\text{(A.9)} \qquad \sqrt{nh}(H_n - H, H_{0,n} - H_0, F_{L,n} - F_L) \Rightarrow (G, G_0, G_3)$$

in $D^3([t_{00}, \infty])$. 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

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

[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

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

Because of (A.9) [which implies the uniform consistency of $F_{L,n}(.|x)$ and $H_n(.|x)$], we have that for $\delta < \varepsilon$ $P(I_{A_n(\delta)} \ne 1) \overset{n\to\infty}{\longrightarrow} 0$. 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 $(A.9)$

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

Moreover, the pair $(\tilde{H}_{0,n}, \tilde{F}_{L,n} - \tilde{H}_n)$ is an element of $\{(A,B) \in BV_1^2([t_{00},\tau]) : A \geq 0, B \geq \delta > 0\}$. Since the map $(A,B) \mapsto \int_{t_{00}}^{t} \frac{dA(s)}{B(s)}$ is Hadamard differentiable on this set [see Anderson et al. (1993),page 113], the delta method [see Gill (1989)] yields

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

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

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

and thus it remains to prove that the second term in this sum is of order $o_P(1/\sqrt{nh})$. From Lemma B.2 in the Appendix B we obtain the bound: $\sup_{t_{00} \geq t \geq W_{(2)}(x,n)} \frac{1}{F_{L,n}(s-|x)-H_n(s-|x)} = O_P(1)$, where $W_{(2)}(x,n)$ is defined in the proof of theorem 3.5, and it follows

$$\int_{[W_{(2)}(x,n),t_{00})} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x) - H_n(s-|x)} \leq H_{0,n}(t_{00}|x)O_P(1).$$

Standard arguments yield the estimate $H_{0,n}(t_{00}|x) = o_P(1/\sqrt{nh})$ and thus it remains to derive an estimate for the integral $\int_{[0,W_{(2)}(x,n))} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x)-H_n(s-|x)}$. For this purpose we consider two cases. The first one appears if the δ_i corresponding to $W_{(1)}(x,n)$ equals 0. In this case there is positive mass at the point $W_{(1)}(x,n)$ but at the same time $F_{L,n}(s|x) = F_{L,n}(W_{(2)}(x,n)|x)$ for all $s \in [0, W_{(2)}(x,n))$ and hence $\int_{[0,t_{00})} \frac{H_{0,n}(ds|x)}{F_{L,n}(s-|x)-H_n(s-|x)} \leq H_{0,n}(t_{00}|x)O_P(1)$. For all other values of the corresponding δ_i the mass of $H_{0,n}(ds|x)$ at the point $W_{(1)}(x,n)$ equals zero and thus the integral vanishes. Now the proof of the theorem is complete. □

Proof of Theorem 3.9: Note that the estimator $F_{T,n}^{IP}(.|x)$ is nondecreasing by construction. The assertion for $\hat{q}^{IP}(.|x)$ now follows from the Hadamard differetiability of the inversion mapping tangentially to the space of continuous functions [see Proposition 1 in Gill (1989)], the continuity of $F_T(.|x)$ and the weak uniform consistency of $F_{T,n}^{IP}(.|x)$ on the interval $[0,\tau]$. The corresponding result for the estimator $\hat{q}(.|x)$ follows from the convergence $\mathrm{P}\Big(\hat{q}^{IP}(.|x) \equiv \hat{q}(.|x)\Big) \rightarrow 1$ [see the discussion after Lemma 3.4]. □

Proof of Theorem 3.10: Observe that the estimator $F_{T,n}^{IP}(.|x)$ is nondecreasing by construction and that Theorem 3.8 yields $\sqrt{n}(F_{T,n}^{IP}(.|x) - F^T(.|x)) \Rightarrow W(.)$ 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 $F_S(.|x)$ and $F_T^{-1}(.|x)$ at τ which implies

$F_S(F_T^{-1}(\tau+\alpha|x)|x) < 1$ for some $\alpha > 0$. By the same arguments $f_T(.|x) \geq \delta > 0$ on the interval $[\varepsilon-\alpha, \tau+\alpha]$ if we choose α sufficiently small. Thus Proposition 1 from Gill (1989) together with the delta method yield the weak convergence of the process for $\hat{q}^{IP}(.|x)$. The corresponding result for $\hat{q}(.|x)$ follows from the fact that $\mathrm{P}\Big(\hat{q}^{IP}(.|x) \equiv \hat{q}(.|x)\Big) \to 1$. □

<u>Proof of Theorem 3.11</u>: By the delta method [Gill (1989)], formula (3.6), and the Hadamard differentiability of the product-limit mapping [Anderson et al. (1993)] it suffices to verify the weak convergence of $\sqrt{nh}(\Lambda_{D,n}^-(t|x) - \Lambda_D^-(t|x))_t$ 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 converegnce to $D([0,t_0])$ it suffices to establish condition (A.8) [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 nonvanishing weight $\tilde{W}_i(x)$ i.e.

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

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

$$\begin{aligned}
\Lambda_{D,n}^-(y-|x) &= \sum_{i=1}^n \int_{[0,y)} \frac{d\Big(\tilde{W}_i(x) I_{\{Z_i \leq t, \Delta_i = 1\}}\Big)}{\sum_{j=1}^n \tilde{W}_j(x) I_{\{Z_j \geq t\}}} \\
&= \sum_{i=1}^n \int_{[0,y)} \frac{\tilde{W}_i(x) I_{\{Z_i \geq t\}} d\left(I_{\{Z_i \leq t, \Delta_i = 1\}}\right)}{\sum_{j=1}^n \tilde{W}_j(x) I_{\{Z_j \geq t\}}} \\
&= \sum_{i=1}^n \int_{[0,y)} C_i(x,t) I_{\big\{1 - F_{Z,n}(t-|x) > 0\big\}} dN_i(t)
\end{aligned}$$

for the plug-in estimator of $\Lambda_D^-(.|x)$, where

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

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

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

we obtain the decomposition

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

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

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

where we set the supremum over the empty set to zero. Hence assertion (A.8) can be obtained from the statements

(A.10) $$\sqrt{nh}\sup_{\sigma\le t<t_0}\Lambda^-_D((U\wedge t,t]|x) \xrightarrow{P} 0$$

(A.11) $$\sqrt{nh}\sup_{\sigma\le t\le U\wedge t_0}|(\hat\Lambda^-_{D,n} - \Lambda^-_{D})((\sigma,t]|x)| \xrightarrow{P} 0$$

(A.12) $$\limsup_{n\to\infty} P\left(\sqrt{nh}\sup_{\sigma\le t<U\wedge t_0}|(\Lambda^-_{D,n} - \hat\Lambda^-_{D,n})((\sigma,U\wedge t]|x)| > \delta\right) < \varepsilon/2,$$

which will be shown separately.

Proof of (A.10) For a proof of (A.10) note that

$$\Lambda^-_D((U\wedge t,t]|x) = \begin{cases} 0 & , \quad U\ge t\\ \Lambda^-_D((U,t]|x) & , \quad U<t\end{cases}$$

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.10) can be bounded by

$$(A) \qquad \sup_{\sigma\le t<t_0}\Lambda^-_D((U\wedge t,t]|x) \le \Lambda^-_D((U\wedge t_0,t_0]|x).$$

Observing (R2) we have $F_D([t_0,\infty]|x)>0$ [note that $\Lambda^-_D(dt|x) = \frac{F_D(dt|x)}{1-F_D(t-|x)}$] and obtain

$$(B) \qquad \Lambda^-_D((U\wedge t_0,t_0]|x) \le \int_{(U\wedge t_0,t_0]}\frac{F_D(dt|x)}{F_D([t_0,\infty]|x)} = \frac{F_D((U\wedge t_0,t_0]|x)}{F_D([t_0,\infty]|x)}.$$

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

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

[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

$$\begin{aligned}
P\left(\sqrt{nh}F_D((U\wedge t_0,t_0]|x) > \alpha\right) &\leq \mathrm{E}\left[I_{\{U\wedge t_0<u_n^\alpha\}}\right] = \mathrm{E}\left[\mathrm{E}\left[I_{\{U\wedge t_0<u_n^\alpha\}}\middle|X_1,...,X_n\right]\right] \\
&\leq \mathrm{E}\Big[\mathrm{E}\Big[\prod_{j=1}^n \Big\{1 - I_{\{Z_j\geq u_n^\alpha\}}I_{\{\tilde W_i(x)\neq 0\}}\Big\}\Big|X_1,...,X_n\Big]\Big] \\
&\leq \mathrm{E}\Big[\prod_{j=1}^n \big\{1-\mathrm{E}\left[I_{\{Z_j\geq u_n^\alpha\}}\middle|X_j\right] I_{\{|X_j-x|\leq c_n\}}\big\}\Big] \\
&= \mathrm{E}\Big[\prod_{j=1}^n \big\{1-F_Z([u_n^\alpha,\infty]|X_j)I_{\{|X_j-x|\leq c_n\}}\big\}\Big] \\
&\leq \mathrm{E}\Big[\prod_{j=1}^n \big\{1-F_Z([u_n^\alpha,\infty]|X_j)I_{\{X_j\in[x,x+c_n]\}}\big\}\Big] \\
&\overset{(*)}{\leq} \mathrm{E}\Big[\prod_{j=1}^n \big\{1-CF_Z([u_n^\alpha,\infty]|x)I_{\{X_j\in[x,x+c_n]\}}\big\}\Big] \\
&= \prod_{j=1}^n \big\{1-CF_D([u_n^\alpha,\infty]|x)F_B([u_n^\alpha,\infty]|x)\mathrm{E}\left[I_{\{X_j\in[x,x+c_n]\}}\right]\big\} \\
&\leq \prod_{j=1}^n \big\{1-CF_D([u_n^\alpha,t_0)|x)F_B([u_n^\alpha,\infty]|x)\mathrm{E}\left[I_{\{X_j\in[x,x+c_n]\}}\right]\big\} \\
&= \prod_{j=1}^n \{1-CF_D([u_n^\alpha,t_0)|x)F_B([u_n^\alpha,\infty]|x)\left(c_nf_X(x)+o(c_n)\right)\} \\
&\leq \prod_{j=1}^n \left\{1-CO(1)\frac{\alpha^2}{nh}\frac{F_B([u_n^\alpha,\infty]|x)}{F_D([u_n^\alpha,t_0)|x)}\left(c_nf_X(x)+o(c_n)\right)\right\} \\
&= \left(1-C\frac{\alpha^2}{n}\frac{F_B([u_n^\alpha,\infty]|x)}{F_D([u_n^\alpha,t_0)|x)}f_X(x)(1+o(1))\right)^n,
\end{aligned}$$

where the inequality $(*)$ follows from (R5), the last inequality follows from the definition of u_n^α, and the last equality is a consequence of the fact that the estimate $o(h)$ holds uniformly in j. Now we have

$$\begin{aligned}
\frac{F_D([u_n^\alpha,t_0)|x)}{F_B([u_n^\alpha,\infty]|x)} &\leq \int_{[u_n^\alpha,t_0)}\frac{F_D(ds|x)}{F_B((s,\infty]|x)} \leq \int_{[u_n^\alpha,t_0)}\frac{F_D(ds|x)}{F_B((s,\infty]|x)F_D((s,\infty]|x)F_D([s,\infty]|x)} \\
&= \int_{[u_n^\alpha,t_0)}\frac{\Lambda_D^-(ds|x)}{F_Z((s,\infty]|x)} \longrightarrow 0,
\end{aligned}$$

by (R2) [note that $u_n^\alpha \to t_0$ if $n\to\infty$] and hence the proof of (A.10) is complete.

Proof of (A.11) For fixed $\sigma \le s \le U \wedge t_0$ and sufficiently small h we have

$$\begin{aligned}
|(\hat{\Lambda}^-_{D,n} - \Lambda^-_D)((\sigma, s]|x)| &= \Big|\int_\sigma^s \sum_{i=1}^n C_i(x,t)(\lambda_D(t|X_i) - \lambda_D(t|x))dt\Big| \\
&= \Big|\int_\sigma^s \sum_{i=1}^n C_i(x,t)\left(\partial_x \lambda_D(t|x)(x - X_i) + \frac{1}{2}\partial_x^2 \lambda_D(t|\xi_i)(x - X_i)^2\right) dt\Big| \\
&\le \Big|\int_\sigma^s \sum_{i=1}^n C_i(x,t)(x - X_i)\partial_x \lambda_D(t|x)dt\Big| + \int_\sigma^s \sum_{i=1}^n C_i(x,t)(x - X_i)^2 \frac{C}{2} dt,
\end{aligned}$$

with some positive constant C, where we used (R4) in the last inequality. The second term in the above inequality can be bounded as follows

$$\frac{C}{2}\int_\sigma^s \sum_{i=1}^n C_i(x,t)(x - X_i)^2 dt \le \frac{C}{2}\int_\sigma^s \sum_{i=1}^n C_i(x,t)O(h^2)dt \le \frac{C}{2}(t_0 - \sigma)O(h^2) = O(h^2) = o\left(\frac{1}{\sqrt{nh}}\right),$$

where the last inequality holds uniformly in $s \in [\sigma, t_0]$. Thus it remains to consider the first term, which can be represented as follows

$$\begin{aligned}
R_n &:= \Big|\int_\sigma^s \frac{\sum_{i=1}^n \tilde{V}_i(x) I_{\{Z_i \ge t\}}(x - X_i)}{\sum_{j=1}^n \frac{\tilde{V}_j(x)}{\sum_{k=1}^n \tilde{V}_k(x)} I_{\{Z_j \ge t\}}} \frac{1}{\sum_{k=1}^n \tilde{V}_k(x)} \partial_x \lambda_D(t|x)dt\Big| \\
&= \Big|\frac{1}{\sum_{k=1}^n \tilde{V}_k(x)} \int_\sigma^s \sum_{i=1}^n \tilde{V}_i(x) I_{\{Z_i \ge t\}}(x - X_i)\left(\frac{1 - F_Z(t - |x)}{1 - F_{Z,n}(t - |x)}\right) \frac{\partial_x \lambda_D(t|x)}{1 - F_Z(t - |x)} dt\Big|.
\end{aligned}$$

Now, from condition (W1)(3) and (W1)(4) $\frac{1}{\sum_{k=1}^n \tilde{V}_k(x)} = O_P(1)$, $\sum_{i=1}^n \tilde{V}_i(x) I_{\{Z_i \ge t\}}(x - X_i) = o_P(1/\sqrt{nh})$ uniformly in $t \in (\sigma, U \wedge t_0)$, (R3) and $\frac{1 - F_Z(t-x)}{1 - F_{Z,n}(t-|x)} = O_P(1)$ uniformly in $t \in (\sigma, U \wedge t_0)$ [see Lemma $B.3$ in the Appendix B] we obtain

$$R_n = o_P(1/\sqrt{nh})\left|\int_\sigma^s \frac{\partial_x \lambda_D(t|x)}{1 - F_Z(t - |x)} dt\right| \le o_P(1/\sqrt{nh}) \int_\sigma^{t_0} \frac{|\partial_x \lambda_D(t|x)|}{1 - F_Z(t - |x)} dt = o_P(1/\sqrt{nh})$$

uniformly in $s \in [\sigma, t_0]$, and hence assertion (A.11) is established.

Proof of (A.12) Observe that $|(\Lambda^-_{D,n} - \hat{\Lambda}^-_{D,n})((\sigma, U \wedge t_0]|x)| \le |D_1(U \wedge t_0) - D_1(\sigma)|$, where we have used the notation $M_i(t) := N_i(t) - \int_0^t I_{\{Z_i \ge s\}} \Lambda^-_D(ds|X_i)$ and

(A.13) $$D_1(t) := \sum_{i=1}^n \int_{[0,t]} C_i(x,t) I_{\{1 - F_{Z,n}(t - |x) > 0\}} dM_i(t).$$

Define $\mathcal{F}_t := \sigma(X_i, I_{\{Z_i \le t, \Delta_i = 1\}}, I_{\{Z_i \le t, \Delta_i = 0\}} : i = 1, ..., n)$ and note that M_i are independent locally bounded martingales with respect to $(\mathcal{F}_t)_t$ [see Theorem 2.3.2 p. 61 in Fleming and Harrington

(1991)]. Moreover, $I_{\{1-F_{Z,n}(t-|x)>0\}}$, $I_{\{Z_j \geq t\}}$ and $\tilde{V}_i(x)$ [and with them $C_i(x,t)$] are measurable with respect to $\mathcal{F}_t$ and leftcontinuous, hence predictable. The structure of the 'weights' C_i also implies their boundedness.

Thus for $t < t_0$ $D_1(t)$ is a locally bounded right continuous martingale with predictable variation given by

$$\text{(A.14)} \qquad \begin{aligned} \langle D_1, D_1 \rangle (t) &= \int_{[0,t]} \sum_{i=1}^{n} C_i^2(x,s) I_{\{1-F_{Z,n}(t-|x)>0\}} d \langle M_i, M_i \rangle (s) \\ &= \int_{[0,t]} \sum_{i=1}^{n} C_i^2(x,s) I_{\{1-F_{Z,n}(t-|x)>0\}} \Lambda_D^-(ds|X_i). \end{aligned}$$

Note that with D_1, $D_1(t) - D_1(\sigma)$ is also a locally bounded martingale for $t \in [\sigma, t_0]$ with predictable variation $\langle D_1, D_1 \rangle (t) - \langle D_1, D_1 \rangle (\sigma)$. Hence from a version Lenglart's inequality [see Shorack and Wellner (1986), p. 893, Example 1] we obtain

$$\text{(A.15)} \qquad P\Big(\sup_{\sigma \leq t \leq U \wedge t_0} nh (D_1(t) - D_1(\sigma))^2 \geq \varepsilon \Big) \leq \frac{\eta}{\varepsilon} + P\left(\mathcal{D}_n \geq \eta\right),$$

where $\mathcal{D}_n = nh \left(\langle D_1, D_1 \rangle (U \wedge t_0) - \langle D_1, D_1 \rangle (\sigma) \right)$. If σ is sufficiently close to t_0 it follows

$$\begin{aligned} \mathcal{D}_n &= nh \int_{[\sigma, U \wedge t_0]} \sum_{i=1}^{n} C_i^2(x,t) \Lambda_D^-(dt|X_i) \\ &= nh \int_{[\sigma, U \wedge t_0]} \sum_{i=1}^{n} \frac{\tilde{V}_i^2(x) I_{\{Z_i \geq t\}}}{\left(\sum_{j=1}^{n} \tilde{V}_j(x) I_{\{Z_j \geq t\}} \right)^2} \Lambda_D^-(dt|X_i) \\ &\leq nh \sup_j \tilde{V}_j(x) \int_{[\sigma, U \wedge t_0]} \sum_{i=1}^{n} \frac{C_i(x,t)}{(1 - F_{Z,n}(t-|x))} \frac{1}{\sum_{k=1}^{n} \tilde{V}_k(x)} \Lambda_D^-(dt|X_i) \\ &\overset{(*)}{=} O_P(1) \int_{[\sigma, U \wedge t_0]} \sum_{i=1}^{n} \frac{C_i(x,t)}{(1 - F_{Z,n}(t-|x))} \lambda_D(t|x) dt (1 + o_P(1)) \\ &= O_P(1) \int_{[\sigma, U \wedge t_0]} \frac{\lambda_D(t|x)}{1 - F_{Z,n}(t-|x)} dt (1 + o_P(1)) \\ &= O_P(1) \int_{[\sigma, U \wedge t_0]} \frac{\lambda_D(t|x)}{1 - F_Z(t-|x)} \frac{1 - F_Z(t-|x)}{1 - F_{Z,n}(t-|x)} dt (1 + o_P(1)) \\ &= O_P(1) \int_{[\sigma, U \wedge t_0]} \frac{\lambda_D(t|x)}{1 - F_Z(t-|x)} dt \end{aligned}$$

where we have used (R6), (W1)(1) and (W1)(3) in equality $(*)$ [note that the $(1 + o_P(1))$ holds uniformly in i and t] and Lemma B.3 in the last equality. Now we obtain from (R2) the a.s. convergence $\int_{[\sigma, U \wedge t_0]} \frac{\lambda_D(t|x)}{1-F_Z(t-|x)} dt \overset{\sigma \to t_0}{\longrightarrow} 0$ and hence assertion $(A.12)$ ist established [first choose η in

$(A.15)$ small enough to make η/ε small and then choose σ close enough to t_0].
Summarizing these considerations, we have established (A.10)-(A.12) and the proof of the theorem is complete. □

B Auxiliary results: technical details

Lemma B.1 *Let M be a locally bounded, rightcontinuous martingale on $[0, \infty)$ and denote by $\langle M, M\rangle$ the predictable variation of M. Then we have for any stopping time U with $P(U < \infty) = 1$ and all $\eta, \varepsilon > 0$*

$$P\Big(\sup_{t \leq U} M^2(t) \geq \varepsilon\Big) \leq \frac{\eta}{\varepsilon} + P\left(\langle M, M\rangle\,(U) \geq \eta\right)$$

Proof: In fact this Lemma is a specific version of Lenglart's inequality [see Fleming and Harrington (1991), Theorem 3.4.1]. To be precise note that it suffices to prove that for any a.s. finite stopping time T

$$\text{(B.1)} \qquad \mathrm{E}[M^2(T)] \leq \mathrm{E}[\langle M, M\rangle(T)].$$

Let τ_k denote a localizing sequence such that $M(. \wedge \tau_k) \leq k$ and $M^2(t \wedge \tau_k) - \langle M, M\rangle(t \wedge \tau_k)$ is a martingale. Define the processes

$$X_k(t) := M^2(t \wedge \tau_k), \quad Y_k(t) := \langle M, M\rangle(t \wedge \tau_k).$$

Note that by Theorem 2.2.2 in Fleming and Harrington (1991) $(X_k - Y_k)(t \wedge T)$ is a martingale and hence for all t:

$$\text{(B.2)} \qquad \mathrm{E}[X_k(t \wedge T)] = \mathrm{E}[Y_k(t \wedge T)].$$

Moreover, $k \geq X_k(t \wedge T) \stackrel{t \to \infty}{\longrightarrow} X_k(T)$ *a.s.*, and hence we obtain by the Dominated Convergence Theorem

$$\mathrm{E}[X_k(T)] = \lim_{t \to \infty} \mathrm{E}[X_k(t \wedge T)].$$

Since the process $\langle M, M\rangle$ is increasing, we also have

$$\langle M, M\rangle(t \wedge T) \uparrow \langle M, M\rangle(T) \ a.s.$$

and by the Monotone Convergence Theorem

$$\mathrm{E}[Y_k(T)] = \lim_{t \to \infty} \mathrm{E}[Y_k(t \wedge T)].$$

Combining this and $(B.2)$ we obtain the identity $\mathrm{E}[X_k(T)] = \mathrm{E}[Y_k(T)]$ for all a.s. finite stopping times T. Hence we can apply Lenglart's inequality to the process X_k dominated by Y_k which leads to:

$$P_{1,k} := P\left(\sup_{t\le U} M^2(t\wedge\tau_k) \ge \varepsilon\right) \le \frac{\eta}{\varepsilon} + P\left(\langle M, M\rangle(U\wedge\tau_k) \ge \varepsilon\right) =: \frac{\eta}{\varepsilon} + P_{2,k}.$$

Finally, from $\sup_{t\le U} M^2(t\wedge\tau_k) = \sup_{t\le U\wedge\tau_k} M^2(t) \uparrow \sup_{t\le U} M^2(t)$ and $\langle M, M\rangle(U\wedge\tau_k) \uparrow \langle M, M\rangle(U)$ a.s. as k tends to infinity we obtain the desired result. □

Lemma B.2 *Assume that conditions (D2) and (D12) hold. Denote by $W_1(x,n), ..., W_k(x,n)$ those values of $Y_1, ..., Y_n$, whose weights fulfill $W_i(x) \neq 0$ and by $W_{(1)}(x,n), ..., W_{(k)}(x,n)$ 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)-(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 $b < r$:*

$$\sup_{b\ge s\ge W_{(2)}(x,n)} \frac{1}{F_{L,n}(s-|x) - H_n(s-|x)} = O_P(1).$$

Proof: 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

$$\sup_{w\le s\le V_x} \frac{1}{1 - F_{D,n}(s|x) - (1 - F_{Z,n}(s|x))} = O_P(1).$$

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

$$\frac{1}{1 - F_{D,n}(s|x) - (1 - F_{Z,n}(s|x))} = \frac{1}{(1 - F_{D,n}(s|x))F_{B,n}(s|x)}$$

[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 $1/(1 - F_{D,n}(s|x))$. Such a bound can be derived by exploiting the underlying martingale structure of the estimator $\Lambda^-_{D,n}(t)$ of the hazard measure. More precisely, using exactly the same arguments as given in the proof of Theorem 3.6 and the same notation we obtain $\Lambda^-_{D,n}(t\wedge V_x|x) - \hat\Lambda^-_{D,n}(t\wedge V_x|x) = D_1(t\wedge V_x)$, where $D_1(t)$ is defined in (A.13) and is a locally bounded continuous martingale on $[0,\infty)$ with predictable variation given in (A.14). The martingale property of $D_1(t)$ implies that $|D_1(t)|$ is a nonnegative submartingale and from Doob's submartigale inequality we obtain for any $\beta > 0$ and sufficiently large n

$$\mathrm{P}\Big(\sup_{t\le V_x} |D_1(t)| \ge \frac{1}{\beta}\Big) \le \beta \mathrm{E}|D_1(V_x)| \le \beta\sqrt{\mathrm{E}|D_1(V_x)|^2} \le \beta\sqrt{\mathrm{E}\langle D_1, D_1\rangle\,(V_x)} \le \beta\sqrt{\sup_{y\in U_\varepsilon(x)} \Lambda^-_D(V_x|y)},$$

where we have used the inequality (B.1) from the proof of Lemma B.1 and the fact that the weights C_i are positive and sum up to one. Note that the expression $\sqrt{\sup_{y \in U_\varepsilon(x)} \Lambda_D^-(V_x|y)}$ is finite. This follows from condition (D12), which now reads $\sup_{y \in U_\varepsilon(x)} 1 - F_D(\tilde{\tau}_T(y)|y) < 1$ since we have reversed time, and the relation $\Lambda_D^-(t|x) = -\log(1 - F_D(t|x))$. Thus we have obtained the estimate $\sup_{t \le V_x} |D_1(t)| = O_P(1)$.

From the definition of $\hat{\Lambda}_{D,n}^-(t|x)$ we can derive the bound $\sup_t \hat{\Lambda}_{D,n}^-(t|x) \le \sup_{y \in U_\varepsilon(x)} \Lambda_D^-(V_x|y)$, and thus obtain

$$\sup_{t \le V_x} \Lambda_{D,n}^-(t|x) \le \sup_{t \le V_x} |D_1(t)| + \sup_{t \le V_x} \hat{\Lambda}_{D,n}^-(t|x) = O_P(1). \tag{B.3}$$

Finally, we note that the estimator $F_{D,n}(s|x)$ can be expressed in terms of the statistic $\Lambda_{D,n}^-(t|x)$ by using the product limit map as $1 - F_{D,n}(t|x) = \prod_{[0,t]} \left(1 - \Lambda_{D,n}^-(ds|x)\right)$. By exactly the same arguments as given in the proof of Lemma 6 in Gill and Johansen (1990) we obtain the inequality

$$1 - F_{D,n}(t|x) \ge \exp\left(-c(\eta)\Lambda_{D,n}^-(t|x)\right) \qquad a.s.$$

whenever $0 < t \le V_x$, where $1 - 2\eta$ is the size of the largest atom of $\Lambda_{D,n}^-$ on the interval $(0, V_x]$ and $c(\eta) := -\log(\eta)/(1-\eta) < \infty$ [note that, whenever all observations take distinct values, the size of the largest atom of $\Lambda_{D,n}^-$ on $(0, V_x]$ is less or equal to the largest possible value of $\sum_i W_i(x) I_{\{Z_i = V_x, \Delta_i = 1\}} / \sum_i W_i(x) I_{\{Z_i \ge V_x\}}$ which can in turn be bounded by $\bar{c}/(\bar{c} + \underline{c}) < 1$ uniformly in n and thus $\eta > 0$]. The desired bound for $1/(1 - F_{D,n}(s|x))$ now follows from the above inequality together with (B.3) and thus the proof is complete. □

Lemma B.3 *Let $(X_1, Y_1), ..., (X_n, Y_n)$ 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 $nh^{2+\varepsilon} \to \infty$, $h \to 0$ 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 an interval of positive length with $x \in I$.*
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 $U := \max\{Y_i : V_j(x) \neq 0\}$, we have for $n \to \infty$

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

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

$$\frac{1 - F(y - |x)}{1 - \hat{F}_n(y - |x)} = \frac{1 - \bar{F}_n(y - |x)}{1 - \hat{F}_n(y - |x)} \frac{1 - F(y - |x)}{1 - \bar{F}_n(y - |x)}.$$

We now will derive bounds for both ratios on the right hand side. For the first factor we note that the interval I from condition 2. contains either $(x - \varepsilon, x]$ or $[x, x + \varepsilon)$ for some $\varepsilon > 0$. We only treat the first case. We have for sufficiently small h for all $t \in (t_0 - \delta, t_0]$

$$X_i \in (x - h, x] \Rightarrow 1 - F(t - |X_i) > C(1 - F(t - |x))$$

This implies

$$\begin{aligned}
&\sup_{t \in (t_0 - \delta, t_0]} \frac{1 - F(t - |x)}{1 - \bar{F}_n(t - |x)} \\
= &\sup_{t \in (t_0 - \delta, t_0]} \frac{1 - F(t - |x)}{\sum_i I_{\{X_i \in (x - h, x]\}}(1 - F(t - |X_i))} \frac{\sum_i I_{\{X_i \in (x - h, x]\}}(1 - F(t - |X_i))}{\sum_i I_{\{|x - X_i| \le h\}}(1 - F(t - |X_i))} \sum_i I_{\{|x - X_i| \le h\}} \\
\le &\ \frac{1}{C} \frac{\sum_i I_{\{|x - X_i| \le h\}}}{\sum_i I_{\{X_i \in (x - h, x]\}}} = \frac{1}{C} \frac{G_n(x + h) - G_n(x - h)}{G_n(x) - G_n(x - h-)},
\end{aligned}$$

where G_n denotes the empirical distribution function of $X_1, ..., X_n$.
It is a well known fact that $n^\alpha \|G_n - G\|_\infty \stackrel{n \to \infty}{\longrightarrow} 0 \; \forall \, \alpha < 1/2$ almost surely. Since G has a continuous density g with $g(x) > 0$, we obtain

$$\frac{G_n(x + h) - G_n(x - h)}{G_n(x) - G_n(x - h-)} = \frac{\frac{1}{h}(G(x + h) - G(x - h)) + o_P(1)}{\frac{1}{h}(G(x) - G(x - h-)) + o_P(1)} \stackrel{P}{\longrightarrow} 2,$$

which yields

$$\mathrm{P}\Big(\sup_{t \in [t_0 - \delta, t_0]} \frac{1 - F(t - |x)}{1 - \bar{F}_n(t - |x)} > \frac{2}{C} + \varepsilon\Big) \longrightarrow 0 \quad \forall \, \varepsilon > 0.$$

It now remains to consider the interval $[0, t_0 - \delta]$. Observe that condition 3. implies $1 - F(t_0 - \delta - |X_i) \ge 0.5(1 - F(t_0 - \delta - |x))$ if $|X_i - x|$ is sufficiently small, which yields

$$\frac{1 - F(t - |x)}{1 - \bar{F}_n(t - |x)} \le \frac{1 - F(t - |x)}{1 - \bar{F}_n(t_0 - \delta - |x)} \le 2 \frac{1 - F(t - |x)}{1 - F(t_0 - \delta - |x)} < \infty$$

for sufficiently large n. Summarizing, we have obtained the estimate

$$\sup_{0\le y\le t_0} \frac{1-\bar F_n(y-|x)}{1-F(y-|x)} = O_P(1).$$

Thus it remains to consider the ratio $(1-\bar F_n(y-|x))/(1-\hat F_n(y-|x))$. For this purpose note that

$$\text{(B.4)}\qquad \begin{aligned} 1-\hat F(y-|x) &= \sum_i \frac{V_i(x)(1-I_{\{Y_i<y\}})}{\sum_j V_j(x)} = \frac{1+o_P(1)}{C(x)}\sum_i V_i(x)(1-I_{\{Y_i<y\}}) \\ &\ge \underline{c}\frac{1+o_P(1)}{C(x)}\frac{1}{nh}\sum_i I_{\{V_i(x)\neq 0\}}(1-I_{\{Y_i<y\}}) \\ &\ge \underline{c}\frac{1+o_P(1)}{C(x)}\frac{1}{nh}\sum_i I_{\{|x-X_i|\le h\}}(1-I_{\{Y_i<y\}}) \\ &= \underline{c}f_X(x)\frac{1+o_P(1)}{C(x)}\frac{\sum_i I_{\{|x-X_i|\le h\}}(1-I_{\{Y_i\le y\}})}{\sum_j I_{\{|x-X_j|<h\}}}, \end{aligned}$$

uniformly in y. In (B.4) the last equality follows from $\frac{1}{nh}\sum_j I_{\{|x-X_j|\le h\}} = f_X(x)(1+o_P(1))$, the second equality is a consequence of (W1)(3) and the two inequalities follow from (W1)(1) and (W1)(2), respectively. Note that the quantity $\sum_i I_{\{|x-X_i|\le 0\}}(1-I_{\{Y_i<y\}})/\sum_j I_{\{|x-X_j|\le h\}}$ equals $1-\hat F^{NW}(y-|x)$ where $\hat F^{NW}$ is the Nadaraya-Watson estimator of F with rectangular kernel. Thus it remains to find a bound for $(1-\hat F^{NW}(y-|x))/(1-\bar F_n(y-|x))$. Conditionally on $X_1,...,X_n$, this is simply the ratio between $1-F_n$ and $1-\bar F$ where F_n is the empirical distribution function of the sample $\{Y_i : |x-X_i|\le h\}$ with sample size $\sum_j I_{\{|x-X_j|\le h\}}$ and $\bar F$ is the averaged distribution function of the corresponding Y_i. Since the random variables Y_i are independent conditionally on X_i, we can apply the results from van Zuijlen (1978) to obtain the bound

$$\mathrm{P}\left(1-\hat F^{NW}(t-|x) < \beta(1-\bar F_n(t-|x))\ \forall t\le U \Big| X_1,...,X_n\right) \le \frac{2\pi^2}{3}\frac{\beta^2}{(1-\beta)^4}.$$

Since the right hand side of the last inequality does not depend on any random quantities or their distributions, this result also holds unconditionally, and thus the proof is complete. □

References

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

Anderson, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer, New York.

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

Bennett, C. and Sharpley, R. C. (1988). Interpolation of operators. In *Pure and Applied Mathematics*, volume 129. Academic Press.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, University of California, Berkeley.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.*, 18(1):391–404.

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(4):1536–1547.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics*, 19:760–777.

Chernozhukov, V., Fernandéz-Val, I., and Galichon, A. (2006). Quantile and probability curves without crossing. *arXiv:0704.3649*.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics*, 14:181–197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics*, 17:1157–1167.

Dabrowska, D. M. (1992a). Nonparametric quantile regression with censored data. *Sankhyā Ser. A*, 54(2):252–259.

Dabrowska, D. M. (1992b). Variable bandwidth conditional Kaplan-Meier estimate. *Scand. J. Stat.*, 19(4):351–361.

Dette, H., Neumeyer, N., and Pilz, K. (2005). A note on nonparametric estimation of the effective dose in quantal bioassay. *Journal of the American Statistical Association*, 100:503–510.

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

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society, Series B*, 70:609–627.

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

El Ghouch, A. and Van Keilegom, I. (2008). Non-parametric regression with dependent censored data. *Scand. J. Statist.*, 35(2):228–247.

Fitzenberger, B. (1997). A guide to censored quantile regressions. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics: Robust Inference*, volume 15, pages 405–437. North-Holland, Amsterdam.

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

Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Stat.*, 11:49–58.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scandinavian Journal of Statistics*, 16:97–128.

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Annals of Statistics*, 18:1501–1555.

Gu, M. G. and Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.*, 21(2):611–624.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1988). *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Reprint of the 1952 edition.

Koenker, R. (2005). Quantile regression. In *Economic Society Monographs*. Cambridge University Press.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, 27.

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

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association*, 96:458–468.

Li, G. and Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Ann. Inst. Stat. Math.*, 53(4):708–729.

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Annals of Statistics*, 23:787–823.

Lorentz, G. G. (1953). An inequality for rearrangements. *Amer. Math. Monthly*, 60:176–179.

McKeague, I. W. and Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.*, 18(3):1172–1187.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076.

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

Patilea, V. and Rolin, J. (2006). Product-limit estimators of the survival function with twice censored data. *Annals of Statistics*, 34:925–938.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103:637–649.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

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

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics*, 32:143–155.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley & Sons, Chichester.

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

van Zuijlen, M. C. A. (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann. Probability*, 6(2):250–266.

Wang, J. G. (1987). A note on the uniform consistency of the Kaplan-Meier estimator. *Ann. Statist.*, 15(3):1313–1316.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94:137–145.

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

Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, 25:159–166.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93:228–237.