Nonparametric estimation of time–dependent quantiles in a simulation model *

Felix Heimrich,[1] Michael Kohler[2,†] and Lisa Kristl[2]
[1] *Fachbereich Maschinenbau, Technische Universität Darmstadt, Otto-Bernd-Str. 2, 64287 Darmstadt, Germany, email: heimrich@dik.tu-darmstadt.de*
[2] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: {kohler, kristl}@mathematik.tu-darmstadt.de*

February 14, 2017

Abstract
The problem of estimating a time–dependent quantile at each time point $t \in [0,1]$, given independent samples of a stochastic process at discrete time points in $[0,1]$, is considered. It is assumed that the quantiles depend smoothly on t. Results concerning the rate of convergence of quantile estimates based on a local average estimate of the time dependent cumulative distribution functions are presented. In a simulation model importance sampling is applied to construct estimates which achieve better rates of convergences. The finite sample size performance of the estimates is illustrated by applying them to simulated data.

AMS classification: Primary 62G05; secondary 62G20.

Key words and phrases: Conditional quantile estimation, importance sampling, rate of convergence.

1 Introduction

Let $(Y_t)_{t\in[0,1]}$ be an $\mathbb{R}$–valued stochastic process. For equidistant time points $t_1, \ldots, t_n \in [0,1]$ we assume that we have given an independent data set

$$\mathcal{D}_n = \left\{ Y_1^{(t_1)}, \ldots, Y_n^{(t_n)} \right\}, \tag{1}$$

where

$$\mathbf{P}_{Y_k^{(t_k)}} = \mathbf{P}_{Y_{t_k}}.$$

Let $G_{Y_t}(y) = \mathbf{P}(Y_t \leq y)$ be the cumulative distribution function (cdf.) of Y_t, and for $\alpha \in (0,1)$ let

$$q_{Y_t,\alpha} = \inf\{y \in \mathbb{R} \,:\, G_{Y_t}(y) \geq \alpha\}$$

*Running title: *Estimation of time–dependent quantiles*
†Corresponding author: Tel. +49-6151-16-23382, Fax. +49-6151-16-23381

be the α–quantile of Y_t for $t \in [0,1]$. Given the data set $\mathcal{D}_n$ we are interested in constructing estimates $\hat{q}_{Y_t,\alpha} = \hat{q}_{Y_t,\alpha}(\mathcal{D}_n)$ of $q_{Y_t,\alpha}$ such that we have a "small" error

$$\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}|. \tag{2}$$

1.1 Main results

In our first result we use plug-in estimators of $q_{Y_t,\alpha}$ based on local averaging estimators of G_{Y_t} in order to define our quantile estimates. More precisely, let $K : \mathbb{R} \to \mathbb{R}$ be a nonnegative kernel function (e.g., the uniform kernel $K(z) = 1/2 \cdot \mathbb{1}_{[-1,1]}(z)$ or the Epanechnikov kernel $K(z) = 3/4 \cdot (1 - z^2) \cdot \mathbb{1}_{[-1,1]}(z)$). We estimate

$$G_{Y_t}(y) = \mathbf{P}(Y_t \leq y) = \mathbf{E}\{1_{(-\infty,y]}(Y_t)\}$$

by the local average estimator

$$\hat{G}_{Y_t}(y) = \frac{\sum_{i=1}^n 1_{(-\infty,y]}(Y_i^{(t_i)}) \cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)} \tag{3}$$

and use the following plug-in estimator of $q_{Y_t,\alpha}$:

$$\hat{q}_{Y_t,\alpha} = \inf\{y \in \mathbb{R} \,:\, \hat{G}_{Y_t}(y) \geq \alpha\}. \tag{4}$$

Under the assumptions that $G_{Y_t}(y)$ is Hölder smooth with exponent $p \in (0,1]$ (as a function of $t \in [0,1]$), and that a density of Y_t exists which is bounded away from zero and infinity in a neighborhood of $q_{Y_t,\alpha}$, we show that for suitable chosen bandwidth h_n and kernel $K : \mathbb{R} \to \mathbb{R}$ the supremum norm error (2) of this estimate converges to zero in probability with rate $(\log(n)/n)^{p/(2p+1)}$.

In our second result we show that this rate of convergence can be improved in a simulation model using importance sampling. Here we assume that Y_t is given by

$$Y_t = m(t, X_t),$$

where X_t is an $\mathbb{R}^d$–valued random variable with a density $f(t,\cdot) : \mathbb{R}^d \to \mathbb{R}$ and $m : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ is a costly to evaluate blackbox function. In this framework, we construct an importance sampling variant of the above plug-in quantile estimator, which is based on an initial quantile estimator and a suitably chosen estimation (surrogate) m_n of m. Our main result here is that under suitable assumptions on f and m we can achieve in case of Hölder smooth $G_{Y_t}(y)$ (with exponent $p \in (0,1]$ and as a function of $t \in [0,1]$) the rate of convergence $(\log(n)/n)^{p\cdot(4p+1)/(2p+1)^2)}$. This rate of convergence is achievable for an estimate which is based on at most n evaluations of m (as the estimate (4)).

The finite sample size performance of our estimates are illustrated by applying them to simulated data.

1.2 Discussion of related results

Time–dependent quantile estimation can be regarded as conditional quantile estimation for fixed design, where we condition on the time t. A short introduction into conditional quantile estimation is presented in Yu et al. (2003). Plug-in conditional quantile estimators have been considered already in quite a few papers. Stone (1977) showed consistency in probability, Stute (1986) proved asymptotic normality, and Bhattacharya and Gangopadhya (1990) used a Bahadur-type representation, c.f. Bahadur (1966), to show asymptotic normality. A double-kernel approach was presented by Yu and Jones (1998), who analyzed the mean squared error of their estimator.

Other conditional quantile estimation approaches are discussed for example in Koenker and Bassett (1978), who proposed a quantile regression estimator, and Mehra et al. (1991), who presented a smooth conditional quantile estimator, showed its asymptotic normality and analyzed its pointwise almost sure rate of convergence. Xiang (1996) also proposed a new kernel estimator of a conditional quantile and derived the same pointwise almost sure rate of convergence than Mehra et al. (1991) under weaker assumptions. Quantile regression estimators are also considered in Chaudhuri (1991), Fan et al. (1996) and Yu and Jones (1998). In contrast to the articles cited above, we analyze the rate of convergence in probability of the supremum norm error of our quantile estimates.

As an estimate m_n of m any kind of nonparametric regression estimate can be chosen. Amongst others a kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980), Stone (1977, 1982) or Devroye and Krzyżak (1989)), a partitioning regression estimate (cf., e.g., Györfi (1981) or Beirlant and Györfi (1998)), a nearest neighbor regression estimate (cf., e.g., Devroye (1982) or Devroye, Györfi, Krzyżak and Lugosi (1994)), an orthogonal series regression estimate (cf., e.g. Rafajlowicz (1987) or Greblicki and Pawlak (1985)), a least squares estimate (cf., e.g. Lugosi and Zeger (1995) or Kohler (2000)) or a smoothing spline estimate (cf., e.g., Wahba (1990) or Kohler and Krzyżak (2001)).

Importance sampling is a well-known variance reduction technique, which was originally introduced in order to improve the rate of convergence of estimates of expectations, cf., e.g., Glasserman (2004). The main idea in our setting is to consider, instead of Y_t, a real-valued random variable Z_t, where the distribution of Z_t is chosen such that Z_t is concentrated in a region of the sample space, which has a strong effect on the estimation of $q_{Y_t,\alpha}$. Quantile estimation based on importance sampling has been studied by Cannamela, Garnier and Ioos (2008), Egloff and Leippold (2010) and Morio (2012). However, only Egloff and Leippold (2010) derived theoretical properties about their estimate, such as consistency, but did not analyze the rate of convergence of their estimate. Kohler et al. (2014), Kohler and Tent (2015) as well as Kohler and Krzyżak (2016) studied rates of convergences of importance sampling quantile estimators based on surrogate models, but did not consider a time–dependent setting respectively conditional quantile estimation.

1.3 Outline

In Section 2 the rate of convergence of the first estimate is presented. In Section 3 a time-dependent simulation model is considered and the construction of a time-dependent importance sampling quantile estimate is described and its rate of convergence is analyzed. Section 4 illustrates the finite sample size behavior of the two presented estimates by applying them to simulated data and presents an application in a simulation model in mechanical engineering. Finally, the proofs are given in Section 5.

2 Estimation of time–dependent quantiles

Theorem 1 *Let $\alpha \in (0,1)$. Let $(Y_t)_{t\in[0,1]}$ be an $\mathbb{R}$-valued stochastic process and let G_{Y_t} be the cdf. of Y_t for $t \in [0,1]$. Let $q_{Y_t,\alpha}$ be the α-quantile of Y_t. Assume that Y_t has a density $g(t,\cdot) : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue-Borel measure, which is uniformly bounded away from zero in a neighborhood of $q_{Y_t,\alpha}$, i.e. for some $\epsilon > 0$ there exists a constant $c_1 > 0$ such that*

$$\inf_{t\in[0,1]} \inf_{u\in(q_{Y_t,\alpha}-\epsilon, q_{Y_t,\alpha}+\epsilon)} g(t,u) \geq c_1. \tag{5}$$

Assume further that the function $t \mapsto G_{Y_t}(y)$ is Hölder continuous with Hölder constant $C > 0$ and Hölder exponent $p \in (0,1]$ for $y \in \mathbb{R}$, i.e., assume

$$|G_{Y_s}(y) - G_{Y_t}(y)| \leq C|s-t|^p \quad \text{for all } s,t \in [0,1] \text{ and } y \in \mathbb{R}. \tag{6}$$

Let $n \in \mathbb{N}$ and set $t_k = k/n$ $(k = 1,\ldots,n)$. Let the estimator $\hat{q}_{Y_t,\alpha}$ be defined by (3) and (4) with a nonnegative kernel function $K : \mathbb{R} \to \mathbb{R}$, which is left-continuous on $\mathbb{R}_+$ and monotonically decreasing on $\mathbb{R}_+$, and satisfies

$$K(z) = K(-z) \quad (z \in \mathbb{R}) \tag{7}$$

and

$$c_2 \cdot \mathbb{1}_{[-\alpha,\alpha]}(z) \leq K(z) \leq c_3 \cdot \mathbb{1}_{[-\beta,\beta]}(z) \quad (z \in \mathbb{R}) \tag{8}$$

for some constants $\alpha, \beta, c_2, c_3 \in \mathbb{R}_+ \backslash \{0\}$. Let $h_n > 0$ be such that

$$h_n \to 0 \quad (n \to \infty), \tag{9}$$

$$\frac{n \cdot h_n}{\log(n)} \to \infty \quad (n \to \infty). \tag{10}$$

Then we have for a constant $c_4 > 0$

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_4 \cdot \left(\sqrt{\frac{\log(n)}{n \cdot h_n}} + h_n^p \right) \right) \to 0 \quad \text{for} \quad n \to \infty.$$

In particular, if we set $h_n = c_5 \cdot (\log(n)/n)^{1/(2p+1)}$ for a constant $c_5 > 0$, there exists a constant $c_6 > 0$, such that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_6 \cdot \left(\frac{\log(n)}{n} \right)^{p/(2p+1)} \right) \to 0 \quad \text{for} \quad n \to \infty.$$

Remark 1. In Theorem 1 we have shown the same rate of convergence as the optimal minimax rate of convergence for estimation of a Hölder continuous function (with exponent $p \in (0,1]$) on a compact subset of $\mathbb{R}$ in supremum norm derived in Stone (1982).

Remark 2. To apply the time–dependent quantile estimator in practice, the bandwidth h_n has to be selected in a data-driven way. We suggest to choose h_n in an optimal way concerning the estimation of the time–dependent cdf. G_{Y_t} by $\hat{G}_{Y_t}$ using a version of the well-known splitting the sample technique (cf., e.g., Chapter 7 in Györfi et al. (2002)). Assume that for each of the equidistant time–points t_k $(k = 1, \ldots, n)$ we have given an additional random variable $Y_{k,2}^{(t_k)}$, such that $Y_{k,1}^{(t_k)}$ and $Y_{k,2}^{(t_k)}$ are independent and identically distributed. Let y be the α-quantile of the empirical cdf. corresponding to the data $Y_{1,1}^{(t_1)}, \ldots, Y_{n,1}^{(t_n)}$ and define $\hat{G}_{Y_{t_k}}(y)$ by (3) using the data $Y_{1,1}^{(t_1)}, \ldots, Y_{n,1}^{(t_n)}$ and a bandwidth h_n for $k = 1, \ldots, n$. Then we choose the optimal bandwidth $h_n^\star$ from a finite set of possible bandwidths H_n by minimizing

$$\Delta_{h_n} = \frac{1}{n} \sum\nolimits_{k=1}^{n} \left| \mathbb{1}_{\{Y_{k,2}^{(t_k)} \leq y\}} - \hat{G}_{Y_{t_k}}(y) \right|^2 .$$

3 Application of importance sampling in a simulation model

Let $(X_t)_{t\in[0,1]}$ be an $\mathbb{R}^d$–valued stochastic process and assume that X_t has a density $f(t,\cdot) : \mathbb{R}^d \to \mathbb{R}$ with respect ot the Lebesgue-Borel measure. Let $m : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ be a function, which is costly to compute, and define Y_t by

$$Y_t = m(t, X_t).$$

In the sequel we will assume that we have given data sets $\mathcal{D}_{n,1}$ and $\mathcal{D}_{n,2}$ of the form

$$\begin{aligned} \mathcal{D}_{n,1} &= \left\{ \left(t_1, X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)}\right), \ldots, \left(t_n, X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)}\right) \right\}, \\ \mathcal{D}_{n,2} &= \left\{ \left(t_1, X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)}\right), \ldots, \left(t_n, X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)}\right) \right\}, \end{aligned} \tag{11}$$

where $t_k = k/n$ $(k = 1, \ldots, n)$,

$$\mathbf{P}_{\left(X_{k,i}^{(t_k)}, Y_{k,i}^{(t_k)}\right)} = \mathbf{P}_{\left(X_{t_k}, Y_{t_k}\right)}$$

for $i = 1, 2$, $k = 1, \ldots, n$ and where

$$\left(X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)}\right), \ldots, \left(X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)}\right), \left(X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)}\right), \ldots, \left(X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)}\right) \tag{12}$$

are independent. I.e., we have given two independent samples of (X_{t_k}, Y_{t_k}) at each time point t_k $(k = 1, \ldots, n)$. Furthermore, we assume that we have given independent random variables $X_{k,3}^{(t_k)}, X_{k,4}^{(t_k)}, \ldots$ distributed as X_{t_k} for $k = 1, \ldots, n$ and that we are allowed to

evaluate m at n additional time points. Let m_n be an estimate of m depending on the data set $\mathcal{D}_{n,2}$ and satisfying

$$\sup_{t\in[0,1],\ x\in K_n} |m_n(t,x) - m(t,x)| \leq \beta_n \tag{13}$$

for some $\beta_n > 0$ and some $K_n \subseteq \mathbb{R}^d$. Let $\hat{q}_{Y_t,\alpha}$ be an estimate of $q_{Y_t,\alpha}$ depending on the data set $\mathcal{D}_{n,1}$ and satisfying

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| \geq \eta_n\right) \to 0 \quad (n\to\infty), \tag{14}$$

for some sequence $(\eta_n)_{n\in\mathbb{N}} \in \mathbb{R}_+$, which converges to zero as n goes to infinity, e.g. the estimator $\hat{q}_{Y_{t,\alpha}}$ defined in (4) and $\eta_n = 2\cdot c_6 \cdot (\log(n)/n)^{p/(2p+1)}$ (cf., Theorem 1). Assume that

$$\mathbf{P}\left(\exists t\in[0,1]\ :\ X_t \notin K_n\right) = O\left(\beta_n + \eta_n\right). \tag{15}$$

Set

$$h(t,x) = \frac{1}{c_t}\cdot\left(\mathbb{1}_{\{x\in K_n\,:\,\hat{q}_{Y_t,\alpha}-3\beta_n-3\eta_n\leq m_n(t,x)\leq\hat{q}_{Y_t,\alpha}+3\beta_n+3\eta_n\}} + \mathbb{1}_{\{x\notin K_n\}}\right)\cdot f(t,x),$$

where

$$c_t = \int_{\mathbb{R}^d}\left(\mathbb{1}_{\{x\in K_n\,:\,\hat{q}_{Y_t,\alpha}-3\beta_n-3\eta_n\leq m_n(t,x)\leq\hat{q}_{Y_t,\alpha}+3\beta_n+3\eta_n\}} + \mathbb{1}_{\{x\notin K_n\}}\right)\cdot f(t,x)\,dx.$$

Set $t_k = k/n$ for $k = 1,\ldots,n$. Let Z_t be a random variable with density $h(t,\cdot)$, and let $Z_1^{(t_1)},\ldots,Z_n^{(t_n)}$ be independent random variables such that

$$\mathbf{P}_{Z_k^{(t_k)}} = \mathbf{P}_{Z_{t_k}}$$

for $k = 1,\ldots,n$. Define

$$\hat{G}_{Y_t}^{(IS)}(y) = \frac{\sum_{i=1}^n\left(c_{t_i}\cdot\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}} + b_{t_i}\right)\cdot K\left(\frac{t-t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)}, \tag{16}$$

where

$$b_t = \int_{\mathbb{R}^d}\mathbb{1}_{\{x\in K_n:\ m_n(t,x)<\hat{q}_{Y_t,\alpha}-3\beta_n-3\eta_n\}}\cdot f(t,x)dx$$

for $t\in[0,1]$ and define the plug-in importance sampling estimate of $q_{Y_t,\alpha}$ by

$$\hat{q}_{Y_t,\alpha}^{(IS)} = \inf\{y\in\mathbb{R}\ :\ \hat{G}_{Y_t}^{(IS)}(y)\geq\alpha\}. \tag{17}$$

Theorem 2 *Assume that $(X_t)_{t\in[0,1]}$ is an $\mathbb{R}^d$-valued stochastic process such that X_t has a density $f(t,\cdot)\colon\mathbb{R}^d\to\mathbb{R}$ with respect to the Lebesgue-Borel measure. Let $m\colon[0,1]\times\mathbb{R}^d\to\mathbb{R}$ be a measurable function and assume that Y_t is given by $Y_t = m(t,X_t)$. Let $\alpha\in(0,1)$*

and let $q_{Y_t,\alpha}$ be the α-quantile of Y_t for $t \in [0,1]$ and assume that the function $t \mapsto q_{Y_t,\alpha}$ is Hölder continuous with Hölder constant $C_1 > 0$ and Hölder exponent $q \in (0,1]$, i.e.

$$|q_{Y_{t_1},\alpha} - q_{Y_{t_2},\alpha}| \leq C_1 \cdot |t_1 - t_2|^q.$$

Let $G_{Y_t}(\cdot)$ be the cdf. of Y_t for $t \in [0,1]$. Assume that Y_t has a density $g(t,\cdot) : \mathbb{R} \to \mathbb{R}$, which is continuous as well as uniformly bounded away from zero in a neighborhood of $q_{Y_t,\alpha}$ and which is uniformly bounded from above, i.e. it is assumed that (5) holds and that there exists a constant $c_7 > 0$ such that

$$\sup_{t\in[0,1]} \sup_{u\in\mathbb{R}} g(t,u) \leq c_7. \tag{18}$$

Assume further that the function $t \mapsto G_{Y_t}(y)$ for $y \in \mathbb{R}$ is Hölder continuous with Hölder constant $C_2 > 0$ and Hölder exponent $p \in (0,1]$, i.e. it is assumed that (6) is satisfied. Let $n \in \mathbb{N}$ and set $t_k = k/n$ $(k = 1,\ldots,n)$. Let $K : \mathbb{R} \to \mathbb{R}_+$ be an on $\mathbb{R}_+$ left-continuous and on $\mathbb{R}_+$ monotone decreasing function, which satisfies (7) and (8) for some constants α,β,c_2,$c_3 \in \mathbb{R}_+ \setminus \{0\}$. Let the estimator $\hat{q}^{(IS)}_{Y_t,\alpha}$ be defined by (16) and (17) with $h_{n,1} > 0$, an estimate m_n of m, which satisfies (13) for some $\beta_n > 0$, and an estimate $\hat{q}_{Y_t,\alpha}$ of $q_{Y_t,\alpha}$, which satisfies (14) for some $\eta_n \in \mathbb{R}_+$. Additionally, assume

$$\eta_n \to 0 \quad \textit{for} \quad n \to \infty, \tag{19}$$

$$\beta_n \to 0 \quad \textit{for} \quad n \to \infty, \tag{20}$$

$$h_{n,1} \to 0 \quad \textit{for} \quad n \to \infty, \tag{21}$$

$$\frac{n \cdot h_{n,1}}{\log(n)} \to \infty \quad \textit{for} \quad n \to \infty \tag{22}$$

and for $r = \min\{p,q\}$

$$\frac{h^r_{n,1}}{\beta_n + \eta_n} \to 0 \quad \textit{for} \quad n \to \infty. \tag{23}$$

Furthermore, assume that (15) is satisfied. Then there exists a constant $c_8 > 0$ such that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_8 \cdot \left((\beta_n + \eta_n) \cdot \sqrt{\frac{\log(n)}{n \cdot h_{n,1}}} + h^p_{n,1} \right) \right) \to 0 \quad \textit{for} \quad n \to \infty.$$

In particular, if we set $h_{n,1} = c_9 \cdot (\beta_n + \eta_n)^{2/(2p+1)} \cdot \big(\log(n)/n\big)^{1/(2p+1)}$ for some constant $c_9 > 0$, there exists a constant $c_{10} > 0$ such that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_{10} \cdot (\beta_n + \eta_n)^{\frac{2p}{2p+1}} \cdot \left(\frac{\log(n)}{n} \right)^{\frac{p}{2p+1}} \right) \to 0 \quad \textit{for} \quad n \to \infty.$$

Remark 3. If $\eta_n = 2 \cdot c_6 \cdot (\log(n)/n)^{(p/2p+1)}$ as in Theorem 1 and $\beta_n \leq \eta_n$ is satisfied, we get $(\log(n)/n)^{p\cdot(4p+1)/(2p+1)^2}$ as rate of convergence in Theorem 2.

Remark 4. Observations of $Z_k^{(t_k)}$ for $k = 1, \ldots, n$ can be generated by a rejection method. For this purpose one uses several observations of $(t_k, X_k^{(t_k)})$ for each $k = 1, \ldots, n$ and selects the first observation that satisfies either the condition

$$X_k^{(t_k)} \in K_n \quad \text{and} \quad (|m_n(t_k, X_k^{(t_k)}) - \hat{q}_{Y_{t_k},\alpha}| \leq 3\beta_n + 3\eta_n) \quad \text{or} \quad (X_k^{(t_k)} \notin K_n).$$

Remark 5. Since the smoothness of the system m is unknown in practice, the approximation error β_n of the surrogate model m_n and the estimation error η_n of the initial quantile estimate are unknown. A data-driven method to select β_n and η_n will be presented in Section 4.

Remark 6. As for the first time–dependent estimator a bandwidth $h_{n,1}$ has to be selected in a data-driven way for any application of the importance sampling quantile estimator. To do this, we suggest to proceed as in Remark 2, but to use the importance sampling random variables. More precisely, assume that for each of the equidistant time points t_k $(k = 1, \ldots, n)$ a random variable $Z_{k,2}^{(t_k)}$, such that $Z_{k,1}^{(t_k)}$ and $Z_{k,2}^{(t_k)}$ are independent and identically distributed, as well as observations $m(t_k, Z_{k,2}^{(t_k)})$ for $k = 1, \ldots, n$ are available. Analogously to Remark 2 the bandwidth $h_{n,1}$ can be selected from a set of possible bandwidths $H_{n,1}$ by minimizing

$$\Delta_{h_{n,1}} = \frac{1}{n} \sum\nolimits_{k=1}^{n} \left| \mathbb{1}_{\{m(t_k, Z_{k,2}^{(t_k)}) \leq y\}} - \hat{G}_{Y_{t_k}}^{(IS)}(y) \right|^2$$

over all $h_{n,1} \in H_{n,1}$, where y is chosen as the α-quantile of the empirical cdf. corresponding to the data $m_n(t_1, Z_{1,1}^{(t_1)}), \ldots, m_n(t_n, Z_{n,1}^{(t_n)})$.

Remark 7. In Section 4 we will use Monte-Carlo simulation and additional data $(t_k, X_{k,3}^{(t_k)}), \ldots, (t_k, X_{k,N+2}^{(t_k)})$ for $k = 1, \ldots, n$ and some $N \in \mathbb{N}$ sufficiently large, e.g. $N = 10,000$, in order to approximate the integrals in c_{t_k} and b_{t_k} for $k = 1, \ldots, n$ by

$$\hat{c}_{t_k} = \frac{1}{N} \sum_{i=3}^{N+2} \left(\mathbb{1}_{\{u \in K_n:\ \hat{q}_{Y_{t_k},\alpha} - 3\beta_n - 3\eta_n \leq m_n(t_k,u) \leq \hat{q}_{Y_{t_k},\alpha} + 3\beta_n + 3\eta_n\}} \left(X_{k,i}^{(t_k)} \right) + \mathbb{1}_{\{u \notin K_n\}} \left(X_{k,i}^{(t_k)} \right) \right),$$

$$\hat{b}_{t_k} = \frac{1}{N} \sum_{i=3}^{N+2} \mathbb{1}_{\{u \in K_n:\ m_n(t_k,u) < \hat{q}_{Y_{t_k},\alpha} - 3\beta_n - 3\eta_n\}} \left(X_{k,i}^{(t_k)} \right).$$

4 Application to simulated data

Next, we examine the finite sample size behavior of the local average based time–dependent quantile estimator $\hat{q}_{Y_t,\alpha}$ defined in (4) and the importance sampling time–dependent quantile estimator $\hat{q}_{Y_t,\alpha}^{(IS)}$ defined in (17) by applying them to simulated data. Both estimators use the same number $3n$ of evaluations of m, which we archieve by using for the local average based quantile estimator $\hat{q}_{Y_t,\alpha}$ three independent copies of Y_{t_k} for each time point $t_k = k/n$ $(k = 1, \ldots, n)$. Here

$$\tilde{\mathcal{D}}_{n,1} = \{Y_{1,1}^{(t_1)}, Y_{1,2}^{(t_1)}, \ldots, Y_{n,1}^{(t_n)}, Y_{n,2}^{(t_n)}\}$$

is used for the main quantile estimation and

$$\tilde{\mathcal{D}}_{n,2} = \{Y_{1,3}^{(t_1)}, \ldots, Y_{n,3}^{(t_n)}\}$$

is used as testing data for the data–driven bandwidth selection method described in Remark 3, where for each $k = 1, \ldots, n$ we compare $Y_{k,1}^{(t_k)}$ and $Y_{k,2}^{(t_k)}$ with $Y_{k,3}^{(t_k)}$.
For the importance sampling estimator, we also use three evaluations of the function m at each time point $t_k = k/n$ $(k = 1, \ldots, n)$ as well as additional copies $X_{k,3}^{(t_k)}, X_{k,4}^{(t_k)}, \ldots$ of X_{t_k}, which are used for the generation of $Z_{k,1}^{(t_k)}$ and $Z_{k,2}^{(t_k)}$ for $k = 1, \ldots, n$ and for integral approximation by Monte-Carlo simulation in the estimation of c_{t_k} and b_{t_k} (cf. Remark 7) for $k = 1, \ldots, n$ and $N = 10,000$. To generate observations of $Z_{k,1}^{(t_k)}$ and $Z_{k,2}^{(t_k)}$ for $k = 1, \ldots, n$ by applying the rejection method presented in Remark 4, a surrogate model m_n of m as well as its approximation error β_n (see (13)) and an initial quantile estimation as well as its estimation error η_n (see (14)) are required. Although we use the in Section 2 investigated local average based time–dependent quantile estimator $\hat{q}_{Y_{t_k},\alpha}$ for the initial quantile estimation, η_n is unknown in reality because the Hölder exponent p of the smoothness condition in Theorem 1 is unknown.
A data set $\mathcal{D}_{n,1}$, as described in (11), is used to generate an initial quantile estimation by the local average based time–dependent quantile estimator $\hat{q}_{Y_t,\alpha}$. To determine η_n in a data-driven way, we suggest to use a bootstrap method and the data sets $\mathcal{D}_{n,1}$ and $\mathcal{D}_{n,2}$. For each time point t_k $(k = 1, \ldots, n)$, we choose $(t_k, Y_{k,1}^{(t_k)})$ or $(t_k, Y_{k,2}^{(t_k)})$ randomly from $\mathcal{D}_{n,1}$ or $\mathcal{D}_{n,2}$ as learning or testing data sets. We repeat the procedure 30 times to obtain multiple learning and testing data sets and to estimate $q_{Y_{t_k},\alpha}$ by $\hat{q}_{Y_{t_k},\alpha}$ for $k = 1, \ldots, n$ multiple times. For each time point t_k $(k = 1, \ldots, n)$, we estimate the interquartile range and choose η_n as the median of the interquartile ranges over all time points.
Next, a surrogate model m_n of m can be estimated by a smoothing spline estimator, such as the here applied routine *Tps()* in the statistic package *R*, on the data set $\mathcal{D}_{n,2}$. To estimate β_n in a data-driven way, we suggest a cross-validation method. First, we split $\mathcal{D}_{n,2}$ in five parts. Then for $j = 1, \ldots, 5$ we approximate $m_n^{(j)}$ of m using the data $\mathcal{D}_{n,2}$ without the j-th part and use the j-th part as testing data to compute the absolute error of $m_n^{(j)}$ for each time point t_k $(k = 1, \ldots, n)$. Finally, we determine the maximal absolut error of $m_n^{(j)}$ for each time point and choose β_n as the mean of these maximal errors.
Now, $Z_{1,1}^{(t_1)}, \ldots, Z_{n,1}^{(t_n)}$ and $Z_{1,2}^{(t_1)}, \ldots, Z_{n,2}^{(t_n)}$ can be generated according to Remark 4 for some K_n, where we suggest to use $K_n = [-\hat{c} \cdot \log(n), \hat{c} \cdot \log(n)]$ for some constant $\hat{c} > 0$ (c.f., Table 1).
We compare the two time–dependent quantile estimators on three different models. In all three models we consider first $n_1 = 50$, then $n_2 = 100$ and finally $n_3 = 200$ equidistant time-points in the time intervall $[0, 1]$, i.e. overall 150, 300 or 600 evaluations of the function m, and estimate the time–dependent 0.95-quantiles. Since it is not possible to compare the error in the supremum norm (2), we will compare the maximal absolute errors

$$\max_{t \in \{t_1, \ldots, t_n\}} |\hat{q}_{Y_{t,\alpha}} - q_{Y_{t,\alpha}}| \quad \text{to} \quad \max_{t \in \{t_1, \ldots, t_n\}} |\hat{q}_{Y_{t,\alpha}}^{(IS)} - q_{Y_{t,\alpha}}|. \tag{24}$$

		Model 1	Model 2	Model 3
H		$0.1, 0.2, 0.3, 0.4, 0.5$	$0.1, 0.2, 0.3, 0.4, 0.5$	$0.3, 0.4, 0.5, 0.6, 0.7$
$\hat{c}$		0.7	0.7	0.25
$\hat{q}_{Y_{t,\alpha}}$	$n_1 = 50$	0.65255 (0.51553)	0.42933 (0.25265)	0.13462 (0.13462)
	$n_2 = 100$	0.62070 (0.49037)	0.33727 (0.19845)	0.12204 (0.12204)
	$n_3 = 200$	0.61024 (0.48209)	0.29215 (0.17190)	0.11215 (0.11215)
$\hat{q}^{(IS)}_{Y_{t,\alpha}}$	$n_1 = 50$	0.63547 (0.50204)	0.46357 (0.27280)	0.14755 (0.14755)
	$n_2 = 100$	0.54891 (0.43366)	0.30981 (0.18229)	0.09100 (0.09100)
	$n_3 = 200$	0.45842 (0.36215)	0.24536 (0.14437)	0.07598 (0.07598)

Table 1: Results for the 0.95–quantile estimation by $\hat{q}_{Y_{t,\alpha}}$ and $\hat{q}^{(IS)}_{Y_{t,\alpha}}$. Reported are the maximal absolute errors (and in brackets the relative errors).

We repeat the estimation 100 times and compare the mean of these errors.

In our first model X_t is $\mathcal{N}(0, (1/2 \cdot t - t^2 + 1/2)^2)$ distributed and

$$m(t, x) = t \cdot \exp(x) \quad (t \in [0, 1],\ x \in \mathbb{R}).$$

For the second model X_t is $\mathcal{N}(0, (t^2 - t^4 + 1/2)^2)$ distributed and m is characterized by

$$m(t, x) = \sqrt{t + x^2} \quad (t \in [0, 1],\ x \in \mathbb{R}).$$

In our last model X_t is $\mathcal{N}(0, (3/2 \cdot t^4 - 3/2 \cdot t^2 + 1)^2)$ distributed and m is given by

$$m(t, x) = \begin{cases} 0 & \text{, for} \quad x \leq 0 \\ \sin(x) & \text{, for} \quad 0 < x < \frac{\pi}{2} \\ 1 & \text{, for} \quad x \geq \frac{\pi}{2} \end{cases} \quad (t \in [0, 1],\ x \in \mathbb{R}).$$

All three models satsify the assumptions of Theorem 1 and Theorem 2. The results for both estimators are presented in Table 1. Moreover, Table 1 shows the set of possible bandwidths H for both estimators and the applied constant $\hat{c}$ in the intervall K_n. As expected the results for the importance sampling time–dependent quantile estimation $\hat{q}^{(IS)}_{Y_{t,\alpha}}$ are better than the results for the local average based quantile estimation $\hat{q}_{Y_{t,\alpha}}$ as the sample size increases in the investigated models. More over, it can be seen that for both estimators the estimation becomes more accurate for higher sample sizes.

In contrast to the first three examples, where we have demonstrated the finite sample size behavior of the two presented estimators, we now consider a problem, which occured in the Collaborative Research Centre 805, to illustrate the usefullness of time–dependent quantile estimation. The German CRC 805 works on controlling uncertainty in mechanical structures such as aircraft landing gears. To test different approaches to control uncertainty the CRC 805 has designed a demonstrator model of a suspension strut, which is shown in Figure 1. The demonstrator model is designed in two versions, a

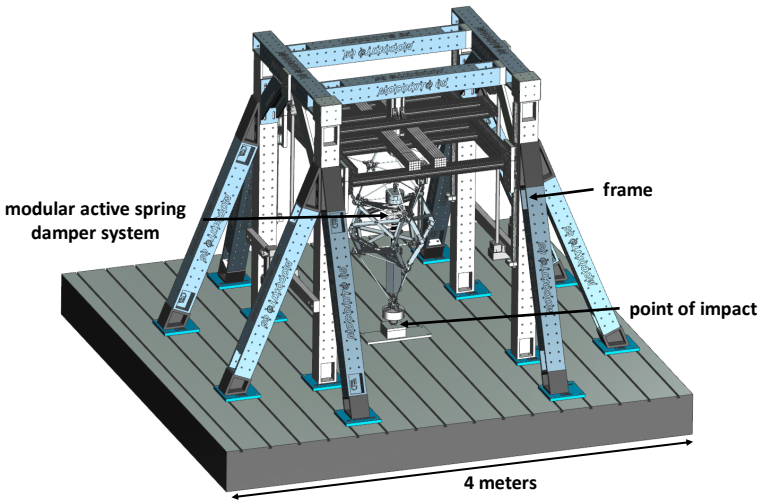

Figure 1: The suspension strut demonstrator of CRC 805.

virtual computer experiment and a real experimental setup. In the experiments a modular spring damper system is suspended on a frame and falls down on the base of the frame. In doing so sensors measure different parameters such as acceleration, absolute position of the modular active spring damper system and the force at the point of impact. Predicting this force is important to calculate the stress and its deviation in order to determine the correct load capacity for the usage phase of the product already in the development phase. We will investigate the impact of an aging spring, i.e. an over time decreasing spring stiffness X_t, on the force at the point of impact $Y_t = m(t, X_t)$ using the virtual demonstrator to generate time–dependent data. In a time–invariant system the spring stiffness is assumed to be normally distributed with expectation $\mu = 35000$ [N/m] and standard deviation $\sigma = 1166.67$ [N/m] (c.f. Schuëller (2007)). It seems reasonable that the spring will weaken over time, when it is used continuously. Therefore, in this academic example, we assumme that the spring constant deteriorates over time exponentially as Zill and Wright (2009) do in Chapter 3.8.1. More precisely, we assume that the spring stiffness X_t is normally distributed with expectation $\mu_t = 35000 * \exp(-0.005 * t)$ [N/m] and standard deviation $\sigma_t = 1166.67$ [N/m]. To generate the data of the force at the point of impact Y_t differential-algebraic equation systems have to be solved by the routine *RecurDyn* of the software *Siemens NX*. As before we use 3 observations of Y_t at $n = 100$ time points, i.e. 300 evaluations of the computer experiment m. Since the true quantiles are unknown, we only present the 0.95–quantiles estimated by the importance sampling estimator $\hat{q}^{(IS)}_{Y_t,0.95}$. The results are shown in Figure 2. It can be seen that less force acts on the point of impact, when the spring stiffness decreases over time.

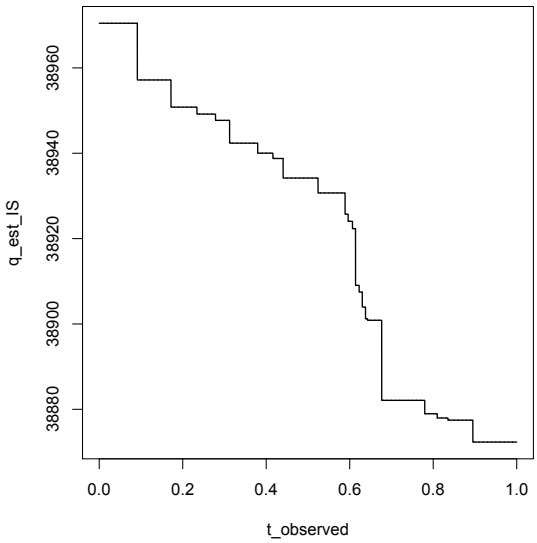

Figure 2: The 0.95-quantile of the force at the point of impact estimated by $\hat{q}^{(IS)}_{Y_{t,0.95}}$.

5 Proofs

5.1 Preliminaries to the proofs of Theorem 1 and 2

In the proofs of Theorem 1 and 2 we will need two auxiliary lemmas. In order to formulate our first auxiliary result, we need the notion of covering numbers. Denote by $\mathcal{N}_1(\epsilon, \mathcal{G}, x_1^n)$ the size of the smallest L_1 norm ϵ-cover of a set of functions $\mathcal{G}$ on $x_1^n = (x_1, \ldots, x_n) \in \mathbb{R}^d$, where a L_1 norm ϵ-cover is a finite collection of functions $g_1, \ldots, g_N : \mathbb{R}^d \to \mathbb{R}$ with the property that for every $g \in \mathcal{G}$ there exists a $j = j(g) \in \{1, \ldots, N\}$ such that

$$\frac{1}{n} \sum_{i=1}^n |g(x_i) - g_j(x_i)| < \epsilon.$$

Lemma 1 *Let $n \in \mathbb{N}$, let $Z_{t_1}, \ldots, Z_{t_n}$ be independent random variables with values in $\mathbb{R}^d$, $t_i = i/n$ for $i = 1, \ldots, n$ and some sequence $(\epsilon_n)_{n\in\mathbb{N}} \in \mathbb{R}_+ \setminus \{0\}$. Let $\mathcal{G}_n$ be a set of functions $g : [0,1] \times \mathbb{R}^d \to [0, B_n]$ such that*

$$\frac{1}{n} \sum_{i=1}^n g(t_i, x_i) \leq \nu_n \quad (g \in \mathcal{G}_n,\ (t_1, x_1), \ldots, (t_n, x_n) \in [0,1] \times \mathbb{R}^d) \tag{25}$$

for some sequences $(B_n)_{n\in\mathbb{N}}$, $(\nu_n)_{n\in\mathbb{N}} \in \mathbb{R}_+ \setminus \{0\}$. Set

$$(\bar{t}, \bar{Z}) = ((t_1, Z_{t_1}), (t_2, Z_{t_2}), \ldots, (t_n, Z_{t_n}))^T \in ([0,1] \times \mathbb{R}^d)^n.$$

Then $n \geq 8B_n\nu_n/\epsilon_n^2$ implies

$$\mathbf{P}\left\{\exists g \in \mathcal{G}_n : \left|\frac{1}{n}\sum_{i=1}^n g(t_i, Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g(t_i, Z_{t_i})\right\}\right| > \epsilon_n\right\}$$

$$\leq \quad 8 \cdot \sup_{(\bar{t},\bar{z})\in([0,1]\times\mathbb{R}^d)^n} \mathcal{N}_1\left(\frac{\epsilon_n}{8}, \mathcal{G}_n, (\bar{t},\bar{z})\right) \cdot e^{-\frac{n\cdot\epsilon_n^2}{128\cdot B_n\cdot\nu_n}}.$$

In Lemma 1 there may be some measurability problems because the supremum is taken over a possible uncountable set. In order to avoid that the notation becomes too complicated, we will ignore these problems and refer to van der Vaart and Wellner (1996), where such problems are handled very elegantly by using the notion of outer probability. In the proof we extend the arguments of the proof of Theorem 9.1 in Györfi et al. (2002).

Proof. Step 1: Symmetrization by a ghost sample.
Choose random variables $Z'_{t_1}, \ldots, Z'_{t_n}$, such that Z_{t_i}, Z'_{t_i} are identically distributed for $i = 1, \ldots, n$ and $Z_{t_1}, \ldots, Z_{t_n}, Z'_{t_1}, \ldots, Z'_{t_n}$ are independent. Set $\bar{Z}' = (\bar{Z}'_{t_1}, \ldots, \bar{Z}'_{t_n})$. Let g^* be a function $g \in \mathcal{G}_n$, such that

$$\left|\frac{1}{n}\sum_{i=1}^n g(t_i, Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g(t_i, Z_{t_i})\right\}\right| \; > \; \epsilon_n,$$

if there exists any such function, and let g^* be an arbitrary function in $\mathcal{G}_n$, if such a function does not exist. By Chebyshev's inequality we have

$$\begin{aligned}
& \mathbf{P}\left(\left|\mathbf{E}\left\{\frac{1}{n}\sum\nolimits_{i=1}^n g^*(t_i, Z'_{t_i}) \;\middle|\; Z_1^n\right\} - \frac{1}{n}\sum\nolimits_{i=1}^n g^*(t_i, Z'_{t_i})\right| > \frac{\epsilon_n}{2} \;\middle|\; Z_1^n\right) \\
\leq \quad & \frac{4}{\epsilon_n^2 n^2} \cdot \sum\nolimits_{i=1}^n \mathbf{Var}\{g^*(t_i, Z'_{t_i}) \mid Z_1^n\} \\
\leq \quad & \frac{4\cdot B_n}{\epsilon_n^2 n^2} \cdot \mathbf{E}\left\{\sum\nolimits_{i=1}^n g^*\left(t_i, Z'_{t_i}\right) \;\middle|\; Z_1^n\right\} \\
\leq \quad & \frac{4\cdot B_n\cdot\nu_n}{\epsilon_n^2\cdot n},
\end{aligned}$$

where we have used the independence of $Z'_{t_1}, \ldots, Z'_{t_n}$, the upper bound B_n of the functions $g \in \mathcal{G}_n$ and assumption (25). Consequently, we have for $n \geq 8B_n\nu_n/\epsilon_n^2$:

$$\begin{aligned}
& \mathbf{P}\left(\exists g \in \mathcal{G}_n : \left|\frac{1}{n}\sum_{i=1}^n g(t_i, Z_{t_i}) - \frac{1}{n}\sum_{i=1}^n g(t_i, Z'_{t_i})\right| > \frac{\epsilon_n}{2}\right) \\
\geq \quad & \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z'_{t_i}) \;\middle|\; Z_1^n\right\}\right| > \epsilon_n,\right. \\
& \qquad \left.\left|\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z'_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z'_{ti}) \;\middle|\; Z_1^n\right\}\right| \leq \frac{\epsilon_n}{2}\right) \\
= \quad & \mathbf{E}\left\{\mathbb{1}_{\left\{\left|\frac{1}{n}\sum_{i=1}^n g^*(t_i,Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g^*(t_i,Z'_{t_i}) \;\middle|\; Z_1^n\right\}\right| > \epsilon_n\right\}}\right. \\
& \qquad \left.\cdot\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z'_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^n g^*(t_i, Z'_{t_i}) \;\middle|\; Z_1^n\right\}\right| \leq \frac{\epsilon_n}{2} \;\middle|\; Z_1^n\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&\geq \quad \frac{1}{2}\cdot \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} g^*(t_i, Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g^*(t_i, Z'_{t_i}) \;\middle|\; Z_1^n\right\}\right| > \epsilon_n\right)\\
&= \quad \frac{1}{2}\cdot \mathbf{P}\left(\exists g \in \mathcal{G}_n : \left|\frac{1}{n}\sum_{i=1}^{n} g(t_i, Z_{t_i}) - \mathbf{E}\left\{\frac{1}{n}\sum_{i=1}^{n} g(t_i, Z_{t_i})\right\}\right| > \epsilon_n\right).
\end{aligned}
$$

Step 2 (introduction of additional randomness by random signs) and **Step 3** (conditioning and introduction of a covering) are analogously to Step 2 and Step 3 of the proof of Theorem 9.1 in Györfi et al. (2002). We will only state the results of these steps. For independent and uniformly over $\{-1,1\}$ distributed random variables $U_1,\ldots,U_n$, which are independent of $Z_{t_1},\ldots,Z_{t_n},Z'_{t_1},\ldots,Z'_{t_n}$, we have

$$
\begin{aligned}
&\mathbf{P}\left(\exists g \in \mathcal{G}_n : \left|\frac{1}{n}\sum\nolimits_{i=1}^{n} g(t_i, Z_{t_i}) - \frac{1}{n}\sum\nolimits_{i=1}^{n} g(t_i, Z'_{t_i})\right| > \frac{\epsilon_n}{2}\right)\\
\leq \quad & 2\cdot \sup_{(\bar{t},\bar{z})\in([0,1]\times\mathbb{R}^d)^n} \mathcal{N}_1\left(\frac{\epsilon_n}{8}, \mathcal{G}_n, (\bar{t},\bar{z})\right)\cdot \max_{g\in\mathcal{G}_{n,\frac{\epsilon_n}{8}}} \mathbf{P}\left(\left|\frac{1}{n}\sum\nolimits_{i=1}^{n} U_i\cdot g(t_i, z_{t_i})\right| > \frac{\epsilon_n}{8}\right),
\end{aligned}
$$

where $\mathcal{G}_{n,\frac{\epsilon_n}{8}}$ is an L_1 $\frac{\epsilon_n}{8}$-cover on $(\bar{t},\bar{z})$ of minimal size.

Step 4: Application of Hoeffding's inequality.
Since $U_1\cdot g(t_1,Z_{t_1}),\ldots,U_n\cdot g(t_n,Z_{t_n})$ are independent random variables with

$$
-g(t_i,z_{t_i}) \leq U_i\cdot g(t_i,z_{t_i}) \leq g(t_i,z_{t_i}) \quad \text{for } i=1,\ldots,n,
$$

we obtain by using Hoeffding's inequality, the upper bound of $g\in\mathcal{G}_n$ and (25)

$$
\begin{aligned}
\mathbf{P}\left(\left|\frac{1}{n}\sum\nolimits_{i=1}^{n} U_i g(t_i,z_{t_i})\right| > \frac{\epsilon_n}{8}\right) &\leq 2\cdot\exp\left(-\frac{2\cdot n\cdot\left(\frac{\epsilon_n}{8}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}|g(t_i,z_{t_i})-(-g(t_i,z_{t_i}))|^2}\right)\\
&\leq 2\cdot\exp\left(-\frac{2\cdot n\cdot\left(\frac{\epsilon_n}{8}\right)^2}{\frac{4B_n}{n}\cdot\sum_{i=1}^{n} g(t_i,z_{t_i})}\right) \leq 2\cdot\exp\left(-\frac{n\epsilon_n^2}{128B_n\nu_n}\right).
\end{aligned}
$$

All four steps considered, the assertion of the lemma is proven. □

Lemma 2 *Let*

$$
\begin{aligned}
\bar{\mathcal{G}}_n \;:=\; \Big\{&\bar{g} : [0,1]\times\mathbb{R}^d \to [0,d_n] :\; \bar{g}(u,x) = c_u\cdot \mathbb{1}_{\{m(u,x)\leq y\}}\cdot K\left(\frac{t-u}{h_{n,1}}\right)\\
&((u,x)\in[0,1]\times\mathbb{R}^d),\; t\in[0,1],\; y\in\mathbb{R}\Big\},
\end{aligned}
$$

where $c_u\in[0,d_n]$ for all $u\in[0,1]$ and $d_n\in\mathbb{R}_+$. Let the kernel K and m be defined as in Theorem 2. Then for any $(u_1^n,x_1^n)\in\mathbb{R}^n\times\mathbb{R}^n$ and $0<\epsilon_n<d_n\cdot\min\{1,K(0)/2\}$ it holds

$$
\mathcal{N}_1(\epsilon_n,\bar{\mathcal{G}}_n,(u_1^n,x_1^n)) \leq c_{11}\cdot n\cdot\left(\frac{d_n}{\epsilon_n}\right)^8
$$

for some constant $0<c_{11}<\infty$.

Proof. The proof is a straightforward extension of standard bounds for covering numbers (cf., e.g., Chapter 9 in Györfi et al. (2002) and Lemma 3.2 in Kohler et al. (2003)), and is available from the authors on request. □

5.2 Proof of Theorem 1

To prove Theorem 1, we need three auxiliary lemmas.

Lemma 3 *Assume that $G_{Y_t}(q_{Y_t,\alpha}) = \alpha$ and that the kernel K is defined as in Theorem 1. Furthermore, assume that (10) holds and that $t_1, \ldots, t_n$ are equidistant in $[0,1]$. Then we have on the event that $Y_1^{(t_1)}, \ldots, Y_n^{(t_n)}$ are pairwise disjoint that for any $t \in [0,1]$ it holds*

$$|G_{Y_t}(q_{Y_t,\alpha}) - \hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha})| \leq \frac{c_{12}}{n \cdot h_n}$$

for some constant $c_{12} > 0$ and $n \in \mathbb{N}$ sufficiently large.

Proof. On the event that $Y_1^{(t_1)}, \ldots, Y_n^{(t_n)}$ are pairwise disjoint $\hat{G}_{Y_t}$ is a cdf. with n jumps, and the jumps sizes are bounded from above by

$$\frac{K(0)}{\sum_{j=1}^n K\left(\frac{t_i - t_j}{h_n}\right)} \quad (i = 1, \ldots, n).$$

By assumption (8), Lemma 5 from Bott et al. (2017) and assumption (10), we have

$$\sum_{j=1}^n K\left(\frac{t - t_j}{h_n}\right) \geq c_{13} \cdot n \cdot h_n \quad (t \in [0,1]),$$

for some constant $c_{13} > 0$ and sufficiently large $n \in \mathbb{N}$. This implies

$$\alpha \leq \hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha}) \leq \alpha + \frac{c_{12}}{n \cdot h_n}$$

for some constant $c_{12} > 0$ and n large enough. Using $G_{Y_t}(q_{Y_t,\alpha}) = \alpha$ we get the assertion.

Lemma 4 *Assume that the kernel K is nonnegative and satisfies assumption (8) of Theorem 1. Assume further that the function $t \mapsto G_{Y_t}(y)$ for $y \in \mathbb{R}$ is Hölder continuous with Hölder constant $C > 0$ and Hölder exponent $p \in (0,1]$, i.e.*

$$|G_{Y_s}(y) - G_{Y_t}(y)| \leq C|s - t|^p \quad \textit{for all } s, t \in [0,1] \textit{ and all } y \in \mathbb{R},$$

and assume that

$$n \cdot h_n \to \infty \quad \textit{for} \quad n \to \infty. \tag{26}$$

Then for any $t \in [0,1]$ and equidistant $t_1, \ldots, t_n \in [0,1]$ we have

$$\sup_{y \in \mathbb{R}} \left| G_{Y_t}(y) - \mathbf{E}\{\hat{G}_{Y_t}(y)\} \right| \leq c_{14} \cdot h_n^p$$

for some constant $c_{14} > 0$ and sufficiently large $n \in \mathbb{N}$.

Proof. We have

$$\begin{aligned}
\sup_{y\in\mathbb{R}}\left|G_{Y_t}(y)-\mathbf{E}\{\hat{G}_{Y_t}(y)\}\right| &= \sup_{y\in\mathbb{R}}\left|G_{Y_t}(y)-\frac{\sum_{i=1}^n \mathbf{E}\left\{1_{(-\infty,y]}(Y_i^{(t_i)})\right\}\cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)}\right| \\
&= \sup_{y\in\mathbb{R}}\left|G_{Y_t}(y)-\frac{\sum_{i=1}^n G_{Y_{t_i}}(y)\cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)}\right| \\
&\leq \sup_{y\in\mathbb{R}}\frac{\sum_{i=1}^n |G_{Y_t}(y)-G_{Y_{t_i}}(y)|\cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)} \\
&\leq \frac{\sum_{i=1}^n C\cdot|t-t_i|^p\cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)} \\
&\leq c_{14}\cdot h_n^p
\end{aligned}$$

for some constant $c_{14}>0$ and $n\in\mathbb{N}$ sufficiently large. Here the case $0/0$ does not occur for $n\in\mathbb{N}$ sufficiently large, since we get with assumption (26)

$$0\leq \limsup_{n\to\infty}\sup_{t\in[0,1]}\min_{j=1,\dots,n}\frac{|t-t_j|}{h_n}\leq \limsup_{n\to\infty}\frac{1}{n\cdot h_n}\leq\alpha.$$

□

Lemma 5 *Assume that the kernel function K is defined as in Theorem 1. Let $t_1,\dots,t_n$ be equidistant in $[0,1]$. Assume further that (9) and (10) hold. Then there exist constants $c_{15},c_{16},c_{17}>0$ such that*

$$\mathbf{P}\left(\sup_{t\in[0,1],\ y\in\mathbb{R}}\left|\hat{G}_{Y_t}(y)-\mathbf{E}\{\hat{G}_{Y_t}(y)\}\right|>c_{15}\cdot\sqrt{\frac{\log(n)}{nh_n}}\right)\leq c_{16}\cdot n^9\cdot\exp(-c_{17}\cdot\log(n)).$$

Proof. By the definition of $\hat{G}_{Y_t}(y)$ and the fact that K is nonnegative, we get

$$\begin{aligned}
&\mathbf{P}\left(\sup_{t\in[0,1],\ y\in\mathbb{R}}\left|\hat{G}_{Y_t}(y)-\mathbf{E}\{\hat{G}_{Y_t}(y)\}\right|>c_{15}\cdot\sqrt{\frac{\log(n)}{nh_n}}\right) \\
&=\mathbf{P}\left(\sup_{t\in[0,1],\ y\in\mathbb{R}}\left|\frac{\sum_{i=1}^n\left(1_{(-\infty,y]}(Y_i^{(t_i)})-\mathbf{E}\left\{1_{(-\infty,y]}(Y_i^{(t_i)})\right\}\right)\cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)}\right|>c_{15}\cdot\sqrt{\frac{\log(n)}{nh_n}}\right) \\
&\leq\mathbf{P}\left(\sup_{t\in[0,1],\ y\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n\left(1_{(-\infty,y]}(Y_i^{(t_i)})-\mathbf{E}\left\{1_{(-\infty,y]}(Y_i^{(t_i)})\right\}\right)K\left(\frac{t-t_i}{h_n}\right)\right|\right. \\
&\qquad\left.>\inf_{t\in[0,1]}c_{15}\cdot\sqrt{\frac{\log(n)}{nh_n}}\cdot\frac{1}{n}\sum_{j=1}^n K\left(\frac{t-t_j}{h_n}\right)\right) \qquad (27)
\end{aligned}$$

for some constant $c_{15} > 0$. Using that K is bounded from below by an uniform kernel and Lemma 5 from Bott et al. (2017), we obtain

$$\inf_{t\in[0,1]} \sum_{i=1}^{n} K\left(\frac{t-t_i}{h_n}\right) \geq \inf_{t\in[0,1]} c_2 \sum_{i=1}^{n} \mathbb{1}_{[-\alpha,\alpha]}\left(\frac{t-t_i}{h_n}\right) \geq c_2 \cdot (\alpha n h_n - 2) \geq c_{16} \cdot n h_n \quad (28)$$

for some constant $c_{16} > 0$ and $n \in \mathbb{N}$ sufficiently large, where the last inequality follows from assumption (10). Hence, the probability on the right–hand side of (27) can be bounded from above by

$$\begin{aligned}
&\mathbf{P}\Bigg(\sup_{t\in[0,1],\ y\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} \left(1_{(-\infty,y]}(Y_i^{(t_i)}) - \mathbf{E}\left\{1_{(-\infty,y]}(Y_i^{(t_i)})\right\}\right) K\left(\frac{t-t_i}{h_n}\right)\right| \\
&\qquad > c_{15}\cdot c_{16}\cdot \sqrt{\frac{\log(n)}{n h_n}}\cdot h_n\Bigg) \\
&= \mathbf{P}\left(\exists g \in \mathcal{G}_n : \ \left|\frac{1}{n}\sum_{i=1}^{n} g\left(\bar{t}_{(i)}, \bar{Y}_{(i)}\right) - \mathbf{E}\left\{g\left(\bar{t}_{(i)}, \bar{Y}_{(i)}\right)\right\}\right| > c_{15}\cdot c_{16}\cdot\sqrt{\frac{\log(n)}{n h_n}}\cdot h_n\right)
\end{aligned} \quad (29)$$

for sufficiently large $n \in \mathbb{N}$,

$$(\bar{t}, \bar{Y}) = \left((t_1, Y_1^{(t_1)}), (t_2, Y_2^{(t_2)}), \ldots, (t_n, Y_n^{(t_n)})\right)^T \in ([0,1]\times\mathbb{R})^n$$

and

$$\begin{aligned}
\mathcal{G}_n = \quad &\Big\{g : \mathbb{R}\times\mathbb{R} \to [0, K(0)] : \quad g(u,x) = K\left(\frac{t-u}{h_n}\right) \mathbb{1}_{(-\infty,y]}(x) \quad ((u,x)\in\mathbb{R}\times\mathbb{R}), \\
&\qquad t \in [0,1],\ y\in\mathbb{R}\Big\}.
\end{aligned}$$

Next, we will apply Lemma 1 to the last probability in (29). The assumptions of Lemma 1 are satisfied for $\nu_n = c_{17}\cdot h_n$, $\epsilon_n = c_{15}\cdot c_{16}\cdot\sqrt{\frac{\log(n) h_n}{n}}$ and $B_n = K(0)$:
Since K satisfies (8) and $\lim_{n\to\infty} n h_n = \infty$ follows from assumption (10), we have according to Lemma 5 from Bott et al. (2017) for all $g \in \mathcal{G}_n$

$$\sum_{i=1}^{n} g\left(\bar{t}_{(i)}, \bar{Y}_{(i)}\right) \leq \sup_{t\in[0,1]} c_3 \cdot \sum_{i=1}^{n} \mathbb{1}_{[-\beta,\beta]}\left(\frac{t-t_i}{h_n}\right) \leq c_3\cdot(2\beta n h_n + 1) \leq c_{17}\cdot n\cdot h_n = n\cdot\nu_n,$$

for some constant $c_{17} > 0$ and sufficiently large $n \in \mathbb{N}$. Furthermore, we have

$$n \geq c_{19}\cdot\frac{n}{\log(n)} = \frac{8 B_n \nu_n}{\epsilon_n^2}$$

for some constant $c_{19} > 0$ and $n \in \mathbb{N}$ sufficiently large. By Lemma 1 we obtain

$$8 \cdot \sup_{(\bar{t},\bar{y})\in([0,1]\times\mathbb{R})^n} \mathcal{N}_1\left(\frac{\epsilon_n}{8}, \mathcal{G}_n, (\bar{t},\bar{y})\right) \cdot \exp\left(-\frac{n}{B_n}\cdot\frac{\epsilon_n^2}{128\nu_n}\right), \quad (30)$$

as an upper bound for the last probability in (29). Using Lemma 2 (with $c_u = 1 = d_n$) we can bound the covering number in (30) by

$$\sup_{(\bar{t},\bar{y})\in([0,1]\times\mathbb{R})^n} \mathcal{N}_1\left(\frac{\epsilon_n}{8}, \mathcal{G}_n, (\bar{t},\bar{y})\right) \leq c_{11}\cdot n\cdot\left(\frac{1}{\epsilon_n}\right)^8 = c_{20}\cdot n\cdot\left(\frac{n}{\log(n)\cdot h_n}\right)^4$$

for some constant $c_{20} > 0$ and sufficiently large $n \in \mathbb{N}$. Since assumption (10) implies $h_n > 1/n$ for sufficiently large $n \in \mathbb{N}$, this can be bounded further by

$$c_{20}\cdot n\cdot\left(\frac{n}{\log(n)\cdot h_n}\right)^4 \leq c_{20}\cdot n\cdot\left(\frac{n^2}{\log(n)}\right)^4 \leq c_{21}\cdot n^9$$

for some constant $c_{21} > 0$ and sufficiently large $n \in \mathbb{N}$. Therefore, the term on the right–hand side of (30) can be bounded further from above by

$$c_{21}\cdot n^9\cdot\exp\left(-\frac{n}{B_n}\cdot\frac{\epsilon_n^2}{128\nu_n}\right) = c_{21}\cdot n^9\cdot\exp\left(-\frac{c_{15}^2\cdot c_{16}^2}{c_{17}}\cdot\log(n)\right),$$

for some constant $c_{21} > 0$. If we choose in the beginning constant c_{15} such that $c_{15}^2\cdot c_{16}^2/c_{17} \geq 10$, the right–hand side converges to zero as n goes to infinity. □

Proof of Theorem 1.

In the first step of the proof we show for some constant $c_4 > 0$ that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |G_{Y_t}(\hat{q}_{Y_t,\alpha}) - G_{Y_t}(q_{Y_t,\alpha})| > \frac{c_1}{2}\cdot\left(\sqrt{\frac{\log(n)}{nh_n}} + h_n^p\right)\right) \to 0 \quad (n\to\infty) \tag{31}$$

implies

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_4\cdot\sqrt{\frac{\log(n)}{nh_n}} + h_n^p\right) \to 0 \quad (n\to\infty). \tag{32}$$

Set $\epsilon_n = c_4\cdot\left(\sqrt{\log(n)/(nh_n)} + h_n^p\right)$ for $n\in\mathbb{N}$ and assume that for some $t \in [0,1]$ it holds

$$|\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > \epsilon_n. \tag{33}$$

Because of assumption (9) and (10) we have

$$\sqrt{\frac{\log(n)}{nh_n}} + h_n^p \to 0 \quad (n\to\infty).$$

Since Y_t has a density with respect to the Lebesgue–Borel measure, the cdf. G_{Y_t} is differentiable on $\mathbb{R}$ for any $t \in [0,1]$. Inequality (33), the Mean-Value Theorem and assumption (5) ensue

$$|G_{Y_t}(\hat{q}_{Y_t,\alpha}) - G_{Y_t}(q_{Y_t,\alpha})| = g(t,\xi)\cdot|q_{Y_t,\alpha} + \epsilon_n - q_{Y_t,\alpha}| \geq \frac{c_1}{2}\cdot\epsilon_n, \tag{34}$$

for some $\xi \in (q_{Y_t,\alpha}, q_{Y_t,\alpha} + \epsilon_n)$. Thus, we have shown that (33) implies (34), which yields the assertion of the first step.

In the second step of the proof we show (31). Since we have

$$\begin{aligned}
& \sup_{t\in[0,1]} |G_{Y_t}(q_{Y_t,\alpha}) - G_{Y_t}(\hat{q}_{Y_t,\alpha})| \\
\leq \quad & \sup_{t\in[0,1]} |G_{Y_t}(q_{Y_t,\alpha}) - \hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha})| + \sup_{t\in[0,1]} \left|\hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha}) - \mathbf{E}\left\{\hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha})\right\}\right| \\
& + \sup_{t\in[0,1]} \left|\mathbf{E}\left\{\hat{G}_{Y_t}(\hat{q}_{Y_t,\alpha})\right\} - G_{Y_t}(\hat{q}_{Y_t,\alpha})\right| \\
= \quad & T_{1,N} + T_{2,n} + T_{3,n},
\end{aligned}$$

it suffices to show

$$\mathbf{P}\left(T_{i,n} > \frac{c_1}{6} \cdot \left(\sqrt{\frac{\log(n)}{nh_n}} + h_n^p\right)\right) \to 0 \quad (n \to \infty) \tag{35}$$

for $i = 1, 2, 3$. For $i = 1$ this follows directly from Lemma 3. Here $G_{Y_t}(q_{Y_t,\alpha}) = \alpha$ is guaranteed, since Y_t has a density with respect to the Lebesgue-Borel measure. For $i = 2$ the assertion (35) follows from Lemma 5 and for $i = 3$ this follows from Lemma 4. □

5.3 Proof of Theorem 2

Let C_n be the event that $\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| < \eta_n$.

In the first step of the proof we show for arbitrary $t \in [0, 1]$ that if $y \in \mathbb{R}$ satisfies

$$|y - q_{Y_t,\alpha}| \leq 2\beta_n + 2\eta_n, \tag{36}$$

then we have on the event C_n

$$\mathbf{E}_t^\star \left\{\mathbb{1}_{\{m(t,Z_t)\leq y\}}\right\} = \frac{1}{c_t} \cdot (G_{Y_t}(y) - b_t),$$

where in $\mathbf{E}_t^\star$ the expectation is computed with respect to $\mathbf{P}_{Z_t}$.

To do so, we modify arguments of the proofs of Lemma 1 and Lemma 2 in Kohler et al. (2014). Set

$$\begin{aligned}
A_n &= \{x \in K_n : m_n(t, x) < \hat{q}_{Y_t,\alpha} - 3\beta_n - 3\eta_n\}, \\
B_n &= \{x \in K_n : m_n(t, x) > \hat{q}_{Y_t,\alpha} + 3\beta_n + 3\eta_n\}
\end{aligned}$$

for $n \in \mathbb{N}$. Then $h(t, x)$ is given by

$$h(t, x) = \frac{1}{c_t} \cdot \mathbb{1}_{\{x\notin A_n\cup B_n\}} \cdot f(t, x).$$

Using (36) and assumption (13), we obtain for $x \in A_n$ on the event C_n

$$y \geq q_{Y_t,\alpha} - 2\beta_n - 2\eta_n > \hat{q}_{Y_t,\alpha} - 2\beta_n - 3\eta_n > m_n(t,x) + \beta_n \geq m(t,x)$$

which implies

$$\mathbb{1}_{\{m(t,x)\leq y\}} \cdot \mathbb{1}_{\{x\in A_n\}} = \mathbb{1}_{\{x\in A_n\}}.$$

Moreover, (36), (13) and $x \in B_n$ imply on the event C_n

$$y \leq q_{Y_t,\alpha} + 2\beta_n + 2\eta_n < \hat{q}_{Y_t,\alpha} + 2\beta_n + 3\eta_n < m_n(t,x) - \beta_n \leq m(t,x),$$

which implies

$$\mathbb{1}_{\{m(t,x)\leq y\}} \cdot \mathbb{1}_{\{x\in B_n\}} = 0.$$

Therefore, the assertion of Step 1 follows from

$$\begin{aligned}
& \mathbf{E}_t^\star \left\{ \mathbb{1}_{\{m(t,Z_t)\leq y\}} \right\} \\
= \quad & \int_{\mathbb{R}} \mathbb{1}_{\{m(t,x)\leq y\}} \mathbf{P}_{Z_t}(dz_t) \\
= \quad & \int_{\mathbb{R}} \mathbb{1}_{\{m(t,x)\leq y\}} \cdot h(t,x)dx \\
= \quad & \frac{1}{c_t} \cdot \int_{\mathbb{R}} \mathbb{1}_{\{m(t,x)\leq y\}} \cdot \left(1 - \mathbb{1}_{\{x\in A_n\}} - \mathbb{1}_{\{x\in B_n\}}\right) \cdot f(t,x)dx \\
= \quad & \frac{1}{c_t} \cdot \left(\int_{\mathbb{R}} \mathbb{1}_{\{m(t,x)\leq y\}} \cdot f(t,x)dx - \int_{\mathbb{R}} \mathbb{1}_{\{x\in A_n\}} \cdot f(t,x)dx \right) \\
= \quad & \frac{1}{c_t} \cdot \left(G_{Y_t}(y) - b_t \right).
\end{aligned}$$

In the second step of the proof we show that we have on the event C_n

$$\inf_{t\in[0,1]} c_t \geq c_{22} \cdot (\beta_n + \eta_n), \tag{37}$$

$$\sup_{t\in[0,1]} c_t \leq c_{23} \cdot (\beta_n + \eta_n) \tag{38}$$

for some constants $c_{22} > 0$, $c_{23} > 0$ and $n \in \mathbb{N}$ sufficiently large.

First, we show (37) using (13), the definition of the event C_n, assumption (5) and the fact that β_n and η_n go to zero as n goes to infinity

$$\begin{aligned}
\inf_{t\in[0,1]} c_t \quad & \geq \inf_{t\in[0,1]} \int_{\mathbb{R}^d} \left(\mathbb{1}_{\{\hat{q}_{Y_t,\alpha} - 3\beta_n - 3\eta_n \leq m_n(t,x) \leq \hat{q}_{Y_t,\alpha} + 3\beta_n + 3\eta_n\}} \right) \cdot f(t,x)dx \\
& \geq \inf_{t\in[0,1]} \int_{\mathbb{R}^d} \left(\mathbb{1}_{\{q_{Y_t,\alpha} - 2\beta_n - 2\eta_n \leq m(t,x) \leq q_{Y_t,\alpha} + 2\beta_n + 2\eta_n\}} \right) \cdot f(t,x)dx \\
& \geq \inf_{t\in[0,1]} \mathbf{P}\big(m(t,X_t) \in (q_{Y_t,\alpha} - 2\beta_n - 2\eta_n,\ q_{Y_t,\alpha} + 2\beta_n + 2\eta_n)\big) \\
& \geq \inf_{t\in[0,1]} \left(\inf_{u\in E_{n,t}} g(t,u) \right) \cdot (4\beta_n + 4\eta_n)
\end{aligned}$$

$$\geq c_{22} \cdot (\beta_n + \eta_n)$$

for $E_{t,n} = (q_{Y_t,\alpha} - 2\beta_n - 2\eta_n,\ q_{Y_t,\alpha} + 2\beta_n + 2\eta_n)$, some constant $c_{22} > 0$ and $n \in \mathbb{N}$ sufficiently large. Analogously, one can prove inequality (38) using assumption (18) instead of (5).

For $t \in [0,1]$ define the sets

$$\begin{aligned} H_{t,n} &= \{y \in \mathbb{R} :\ |y - q_{Y_t,\alpha}| \leq \beta_n + \eta_n\}, \\ I_{t,n} &= \{y \in \mathbb{R} :\ |y - q_{Y_t,\alpha}| \leq 2\beta_n + 2\eta_n\}. \end{aligned}$$

In the third step of the proof we prove that on the event C_n we have

$$\sup_{t \in [0,1],\ y \in H_{t,n}} \left| \mathbf{E}^{\star}_{t_1,\ldots,t_n} \left\{ \hat{G}^{(IS)}_{Y_t}(y) \right\} - G_{Y_t}(y) \right| \leq C_2 \cdot \beta^p \cdot h^p_{n,1}$$

for large enough $n \in \mathbb{N}$, where the expectation $\mathbf{E}^{\star}_{t_1,\ldots,t_n}$ is defined with respect to $\mathbf{P}_{Z_{t_1},\ldots,Z_{t_n}}$.

First, we observe that by the Theorem of Fubini and the independence of $Z_{t_1}, \ldots, Z_{t_n}$ we have

$$\mathbf{E}^{\star}_{t_1,\ldots,t_n} \left\{ c_{t_i} \cdot \mathbb{1}_{\{m(t_i, Z_{t_i}) \leq y\}} + b_{t_i} \right\} = c_{t_i} \cdot \mathbf{E}^{\star}_{t_i} \left\{ \mathbb{1}_{\{m(t_i, Z_{t_i}) \leq y\}} \right\} + b_{t_i}. \tag{39}$$

Next, we observe that $y \in H_{t,n}$ yields $y \in I_{t_i,n}$ for every $i \in \{1, \ldots, n\}$ that satisfies $K\left(\frac{t_i - t}{h_{n,1}}\right) \neq 0$ (which implies $|t_i - t| \leq \beta \cdot h_{n,1}$ because of assumption (8)) for $n \in \mathbb{N}$ sufficiently large, since

$$\begin{aligned} |y - q_{Y_{t_i},\alpha}| &\leq |y - q_{Y_t,\alpha}| + |q_{Y_t,\alpha} - q_{Y_{t_i},\alpha}| \\ &\leq \beta_n + \eta_n + C_1 \cdot |t - t_i|^q \\ &\leq \beta_n + \eta_n + C_1 \cdot \beta^q \cdot h^q_{n,1} \\ &\leq 2\beta_n + 2\eta_n, \end{aligned}$$

for $n \in \mathbb{N}$ sufficiently large, where we have used that the function $t \mapsto q_{Y_t,\alpha}$ is Hölder continuous and that assumption (23) holds. Thus, (39) and Step 1 yield for $y \in I_{t,n}$ and for $n \in \mathbb{N}$ sufficiently large

$$\begin{aligned} \mathbf{E}^{\star}_{t_1,\ldots,t_n} \left\{ \hat{G}^{(IS)}_{Y_t}(y) \right\} &= \frac{\sum_{i=1}^n \left(c_{t_i} \cdot \mathbf{E}^{\star}_{t_i} \left\{ \mathbb{1}_{\{m(t_i, Z_{t_i}) \leq y\}} \right\} + b_{t_i} \right) \cdot K\left(\frac{t - t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t - t_j}{h_{n,1}}\right)} \\ &= \frac{\sum_{i=1}^n G_{Y_{t_i}}(y) \cdot K\left(\frac{t - t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t - t_j}{h_{n,1}}\right)}. \end{aligned}$$

Here the case $0/0$ does not occur for $n \in \mathbb{N}$ sufficiently large, since

$$0 \leq \limsup_{n \to \infty} \sup_{t \in [0,1]} \min_{j=1,\ldots,n} \frac{|t - t_j|}{h_{n,1}} \leq \limsup_{n \to \infty} \frac{1}{n \cdot h_{n,1}} \leq \alpha,$$

where the last step holds because of (22), and thus $\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right) > 0$ for equidistant $t_1, \ldots, t_n \in [0,1]$ and $n \in \mathbb{N}$ large enough. Using this, the fact that K is nonnegative and satisfies (8) and that the function $t \mapsto G_{Y_t}(\cdot)$ is Hölder continuous, we get

$$\begin{aligned}
& \sup_{t\in[0,1],\ y\in H_{t,n}} \left| \mathbf{E}^\star_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\} - G_{Y_t}(y)\right| \\
= & \sup_{t\in[0,1],\ y\in H_{t,n}} \left| \frac{\sum_{i=1}^n G_{Y_{t_i}}(y)\cdot K\left(\frac{t-t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)} - G_{Y_t}(y)\right| \\
\leq & \sup_{t\in[0,1],\ y\in H_{t,n}} \frac{\sum_{i=1}^n \left|G_{Y_{t_i}}(y) - G_{Y_t}(y)\right| \cdot K\left(\frac{t-t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)} \\
\leq & \sup_{t\in[0,1],\ y\in H_{t,n}} \frac{\sum_{i=1}^n C_2\cdot |t_i - t|^p \cdot K\left(\frac{t-t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)} \\
\leq & \sup_{t\in[0,1],\ y\in H_{t,n}} \frac{\sum_{i=1}^n C_2\cdot \beta^p \cdot h_{n,1}^p \cdot K\left(\frac{t-t_i}{h_{n,1}}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)} \\
= & C_2\cdot \beta^p \cdot h_{n,1}^p
\end{aligned}$$

for $n \in \mathbb{N}$ sufficiently large, which yields the assertion of the third step.

In the fourth step of the proof we observe that because of assumption (14) we have

$$\mathbf{P}(C_n) \to 1 \quad \text{for} \quad n \to \infty.$$

In the fifth step of the proof we show for some constant $c_{24} > 1$ the convergence

$$\begin{aligned}
\mathbf{P}\Bigg(\Bigg\{ & \sup_{y\in\mathbb{R},\ t\in[0,1]} \left|\hat{G}^{(IS)}_{Y_t}(y) - \mathbf{E}^\star_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\}\right| \\
& > c_{24}\cdot(\beta_n + \eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}}\Bigg\} \cap C_n\Bigg) \to 0 \qquad \text{for} \quad n\to\infty.
\end{aligned}$$

Using (39), assumption (8) as well as the nonnegativeness of the kernel K and Lemma 5 of Bott et al. (2017), we get for $\delta_n = c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}}$

$$\begin{aligned}
& \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]} \left|\hat{G}^{(IS)}_{Y_t}(y) - \mathbf{E}^\star_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\}\right| > \delta_n\right\}\cap C_n\right) \\
= & \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]} \left|\frac{\sum_{i=1}^n c_{t_i} K\left(\frac{t-t_i}{h_{n,1}}\right)\cdot\left[\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}} - \mathbf{E}^\star_{t_i}\left\{\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}\right\}\right]}{\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)}\right| > \delta_n\right\}\cap C_n\right)
\end{aligned}$$

$$
\begin{aligned}
\leq \quad & \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]}\left|\sum_{i=1}^n c_{t_i}K\left(\frac{t-t_i}{h_{n,1}}\right)\cdot\left[\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}-\mathbf{E}^\star_{t_i}\left\{\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}\right\}\right]\right|\right.\\
& \qquad \left.> \inf_{t\in[0,1]}\sum_{j=1}^n K\left(\frac{t-t_j}{h_{n,1}}\right)\cdot\delta_n\right\}\cap C_n\right)\\
\leq \quad & \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]}\left|\sum_{i=1}^n c_{t_i}K\left(\frac{t-t_i}{h_{n,1}}\right)\cdot\left[\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}-\mathbf{E}^\star_{t_i}\left\{\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}\right\}\right]\right|\right.\\
& \qquad \left.> c_2\cdot(\alpha\cdot n\cdot h_{n,1}-2)\cdot\delta_n\right\}\cap C_n\right)\\
\leq \quad & \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]}\left|\sum_{i=1}^n c_{t_i}K\left(\frac{t-t_i}{h_{n,1}}\right)\cdot\left[\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}-\mathbf{E}^\star_{t_i}\left\{\mathbb{1}_{\{m(t_i,Z_{t_i})\leq y\}}\right\}\right]\right|\right.\\
& \qquad \left.> \frac{1}{2}\cdot c_2\cdot\alpha\cdot n\cdot h_{n,1}\cdot\delta_n\right\}\cap C_n\right) \qquad (40)
\end{aligned}
$$

for sufficiently large $n\in\mathbb{N}$, where we have used assumption (22), which implies that $n\cdot h_{n,1}$ goes to infinity as n goes to infinity, for the last inequality. In order to apply Lemma 1, we define a set

$$
\begin{aligned}
\bar{\mathcal{G}}_n \quad := \quad & \left\{g:[0,1]\times\mathbb{R}^d\to[0,c_{23}\cdot(\beta_n+\eta_n)\cdot K(0)]:\right.\\
& \qquad g(u,x)=c_u\cdot\mathbb{1}_{\{|c_u|\leq c_{23}\cdot(\beta_n+\eta_n)\}}\cdot\mathbb{1}_{\{m(u,x)\leq y\}}\cdot K\left(\frac{t-u}{h_{n,1}}\right)\\
& \qquad \left.((u,x)\in[0,1]\times\mathbb{R}^d),\ t\in[0,1],\ y\in\mathbb{R}\right\},
\end{aligned}
$$

where on the event C_n the inequality $|c_u|\leq c_{23}\cdot(\beta_n+\eta_n)$ is satisfied for all $u\in[0,1]$ and $n\in\mathbb{N}$ sufficiently large according to Step 2, set

$$
(\bar{t},\bar{Z})=((t_1,Z_{t_1}),(t_2,Z_{t_2}),\ldots,(t_n,Z_{t_n}))^T\in([0,1]\times\mathbb{R}^d)^n
$$

and rewrite the probability on the right–hand side of (40) as

$$
\begin{aligned}
& \mathbf{P}\left(\sup_{g\in\bar{\mathcal{G}}_n}\left|\frac{1}{n}\sum_{i=1}^n g(\bar{t}_{(i)},\bar{Z}_{(i)})-\mathbf{E}^\star_{t_i}\left\{g(\bar{t}_{(i)},\bar{Z}_{(i)})\right\}\right|>\frac{1}{2}\cdot c_2\cdot\alpha\cdot h_{n,1}\cdot\delta_n\right)\\
= \quad & \mathbf{P}\left(\sup_{g\in\bar{\mathcal{G}}_n}\left|\frac{1}{n}\sum_{i=1}^n g(\bar{t}_{(i)},\bar{Z}_{(i)})-\mathbf{E}^\star_{t_i}\left\{g(\bar{t}_{(i)},\bar{Z}_{(i)})\right\}\right|>\epsilon_n\right), \qquad (41)
\end{aligned}
$$

for $\epsilon_n=\frac{1}{2}\cdot c_2\cdot c_{24}\cdot\alpha\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)\cdot h_{n,1}}{n}}$ and $n\in\mathbb{N}$ sufficiently large. Next, we show that for this ϵ_n, $\nu_n=c_{25}\cdot(\beta_n+\eta_n)\cdot h_{n,1}$ and $B_n=c_{23}\cdot(\beta_n+\eta_n)\cdot K(0)$ with some

constants $c_{25}, c_{23} > 0$, the assumptions of Lemma 1 hold:
Since (8) holds, we obtain by Lemma 5 of Bott et al. (2017)

$$\begin{aligned}\frac{1}{n}\sum\nolimits_{i=1}^{n} g(t_i, x_i) &\leq \sup_{t\in[0,1]} c_{23}\cdot(\beta_n+\eta_n)\cdot c_3\cdot\frac{1}{n}\sum\nolimits_{i=1}^{n}\mathbb{1}_{[-\beta,\beta]}\left(\frac{t_i-t}{h_{n,1}}\right)\\ &\leq c_{25}\cdot(\beta_n+\eta_n)\cdot h_{n,1}=\nu_n\end{aligned}$$

for arbitrary $x_1,\ldots,x_n\in\mathbb{R}$, some constants $c_{23}, c_{25} > 0$ and $n\in\mathbb{N}$ sufficiently large, where we have used assumption (22) in the last inequality. Furthermore, we have

$$\frac{8\cdot B_n\cdot\nu_n}{\epsilon_n^2}=\frac{c_{26}\cdot(\beta_n+\eta_n)^2\cdot h_{n,1}}{\left[(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)\cdot h_{n,1}}{n}}\right]^2}=\frac{c_{26}\cdot n}{\log(n)}\leq n$$

for some constant $c_{26} > 0$ and $n\in\mathbb{N}$ sufficiently large. Thus, assumption (25) and $n\geq 8\cdot B_n\cdot\nu_n/\epsilon_n^2$ are satisfied. By Lemma 1, we get

$$8\cdot\sup_{(\bar{t},\bar{z})\in([0,1]\times\mathbb{R}^d)^n}\mathcal{N}_1\left(\frac{\epsilon_n}{8},\bar{\mathcal{G}}_n,(\bar{t},\bar{z})\right)\cdot\exp\left(-\frac{n\cdot\epsilon_n^2}{128\cdot B_n\cdot\nu_n}\right)\tag{42}$$

as an upper bound for (41). The covering number can be bounded by

$$\sup_{(\bar{t},\bar{z})\in([0,1]\times\mathbb{R}^d)^n}\mathcal{N}_1\left(\frac{\epsilon_n}{8},\bar{\mathcal{G}}_n,(\bar{t},\bar{z})\right)\leq c_{27}\cdot n\cdot\left(\frac{\beta_n+\eta_n}{\epsilon_n}\right)^8\tag{43}$$

for some constant $c_{27} > 0$, using Lemma 2. Using (40) to (43), we obtain

$$\begin{aligned}&\mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]}\left|\hat{G}^{(IS)}_{Y_t}(y)-\mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\}\right|>c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{n\cdot h_{n,1}}}\right\}\cap C_n\right)\\ \leq\ & c_{28}\cdot n\cdot\left(\frac{\beta_n+\eta_n}{\epsilon_n}\right)^8\cdot\exp\left(-\frac{c_2^2c_{24}^2}{512\cdot K(0)\cdot c_{23}\cdot c_{25}}\cdot\log(n)\right)\\ \leq\ & c_{29}\cdot n\cdot\left(\sqrt{\frac{n}{\log(n)\cdot h_{n,1}}}\right)^8\cdot\exp\left(-\frac{c_2^2c_{24}^2}{512\cdot K(0)\cdot c_{23}\cdot c_{25}}\cdot\log(n)\right)\\ \leq\ & c_{30}\cdot n^9\cdot\exp\left(-\frac{c_2^2c_{24}^2}{512\cdot K(0)\cdot c_{23}\cdot c_{25}}\cdot\log(n)\right)\leq c_{30}\cdot n^9\cdot\exp\left(-10\cdot\log(n)\right)\end{aligned}\tag{44}$$

for constants $c_{28}, c_{29}, c_{30} > 0$ and n large enough, where we have used that (22) implies $h_{n,1} > 1/n$ for n large enough and where c_{24} was chosen at the beginning of Step 5 large enough. Since the right–hand side of (44) goes to 0 as n goes to infinity, Step 5 is shown.

Let J_n be the event that

$$\sup_{t\in[0,1]}\left|\hat{q}^{(IS)}_{Y_t,\alpha}-q_{Y_t,\alpha}\right|\leq\frac{1}{2}\cdot(\beta_n+\eta_n).$$

In the sixth step of the proof we prove that

$$\mathbf{P}(J_n \cap C_n) \to 1 \quad \text{for} \quad n \to \infty.$$

Let K_n be the event that

$$\sup_{t\in[0,1]} \hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) < \alpha$$

and L_n be the event that

$$\inf_{t\in[0,1]} \hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} + \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) \geq \alpha.$$

We observe that on the event $K_n \cap L_n$ we have for all $t \in [0,1]$

$$\hat{q}^{(IS)}_{Y_t,\alpha} \in \Big[q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n),\ q_{Y_t,\alpha} + \frac{1}{2}\cdot(\beta_n+\eta_n)\Big].$$

Thus, the event $K_n \cap L_n \cap C_n$ implies the event $J_n \cap C_n$ for sufficiently large $n \in \mathbb{N}$. In the following we will show that on the event C_n, we have

$$K_n \cap L_n \supseteq \left\{ \sup_{t\in[0,1],\ y\in\mathbb{R}} \left| \hat{G}^{(IS)}_{Y_t}(y) - \mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\}\right| \leq c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{n\cdot h_{n,1}}} \right\} \quad (45)$$

for $n \in \mathbb{N}$ sufficiently large, which implies the assertion by Step 4 and Step 5. To show (45), we first observe that on the event C_n the inequality

$$\begin{aligned} & \sup_{t\in[0,1]} \left| \mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}\left(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\right)\right\} - G_{Y_t}\left(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\right)\right| \\ \leq \quad & C_2\cdot\beta^p\cdot h^p_{n,1} \end{aligned} \quad (46)$$

holds by Step 3, since $q_{Y_t,\alpha} - 1/2\cdot(\beta_n+\eta_n) \in H_{t,n}$. Additionally, we obtain by the Mean-Value Theorem for an arbitrary $t \in [0,1]$

$$\begin{aligned} G_{Y_t}(q_{Y_t,\alpha}) - G_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) &= g(t,\psi_t)\cdot\Big(q_{Y_{t,\alpha}} - \Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Big) \\ &> \frac{c_1}{2}\cdot(\beta_n+\eta_n) \end{aligned} \quad (47)$$

for $\psi_t \in [q_{Y_t,\alpha} - 1/2\cdot(\beta_n+\eta_n), q_{Y_t,\alpha}]$, some constant $c_1 > 0$ and $n \in \mathbb{N}$ sufficiently large, where we have used assumption (5) and that β_n and η_n converge to zero as n goes to infinity. Using $\alpha = G_{Y_t}(q_{Y_t,\alpha})$, which holds because Y_t has a density which is bounded away from zero in a neighborhood of $q_{Y_t,\alpha}$, the inequalities (46) and (47) as well as the assumptions (22) and (23), we get on the event C_n

$$K_n = \left\{ \sup_{t\in[0,1]} \hat{G}^{(IS)}_{Y_t}\left(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\right) < \alpha \right\}$$

$$
\begin{aligned}
&\supseteq \Bigg\{ \sup_{t\in[0,1]} \Bigg(\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) - \mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Big\}\Bigg) \\
&\qquad + \sup_{t\in[0,1]} \Bigg(\mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Big\} - G_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Bigg) \\
&\qquad + \sup_{t\in[0,1]} \Bigg(G_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) - G_{Y_t}(q_{Y_t,\alpha})\Bigg) < 0 \Bigg\} \\
&\supseteq \Bigg\{ \sup_{t\in[0,1]} \Bigg(\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) - \mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Big\}\Bigg) \\
&\qquad < \frac{c_1}{2}\cdot(\beta_n+\eta_n) - C_2\cdot\beta^p\cdot h_{n,1}^p \Bigg\} \\
&\supseteq \Bigg\{ \sup_{t\in[0,1]} \Bigg(\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) - \mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} - \frac{1}{2}\cdot(\beta_n+\eta_n)\Big)\Big\}\Bigg) \\
&\qquad < c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}} \Bigg\} \\
&\supseteq \Bigg\{ \sup_{t\in[0,1],\ y\in\mathbb{R}} \Big| \hat{G}^{(IS)}_{Y_t}(y) - \mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}(y)\Big\}\Big| < c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}} \Bigg\} \qquad (48)
\end{aligned}
$$

for some $c_{24} > 1$ and $n \in \mathbb{N}$ large enough. Analogously one can show on the event C_n

$$
\begin{aligned}
L_n &= \Bigg\{ \inf_{t\in[0,1]} \hat{G}^{(IS)}_{Y_t}\Big(q_{Y_t,\alpha} + \frac{1}{2}\cdot(\beta_n+\eta_n)\Big) \geq \alpha \Bigg\} \\
&\supseteq \Bigg\{ \sup_{t\in[0,1],\ y\in\mathbb{R}} \Big| \mathbf{E}^{\star}_{t_1,\ldots,t_n}\Big\{\hat{G}^{(IS)}_{Y_t}(y)\Big\} - \hat{G}^{(IS)}_{Y_t}(y)\Big| \leq c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}} \Bigg\}. \qquad (49)
\end{aligned}
$$

Since (48) and (49) imply (45) for n large enough, we have shown the assertion of Step 6.

In the seventh step of the proof we show the assertion of the theorem.

First, we observe that on the event J_n by the Mean-Value Theorem and (5)

$$
\Big|\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}\Big| = \frac{1}{g(t,\psi_t)}\cdot\Big|G_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big) - G_{Y_t}(q_{Y_t,\alpha})\Big| \leq c_{31}\cdot\Big|G_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big) - G_{Y_t}(q_{Y_t,\alpha})\Big|
$$

holds, for some $\psi_t \in (q_{Y_t,\alpha} - 1/2\cdot(\beta_n+\eta_n), q_{Y_t,\alpha} + 1/2\cdot(\beta_n+\eta_n))$ and some constant $c_{31} > 0$. Let $\theta > 0$ be arbitrary. Using the definition of $\hat{q}^{(IS)}_{Y_t,\alpha}$ the right–hand side of the above inequality can be bounded further from above by

$$
\begin{aligned}
& c_{31}\cdot\Big|G_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big) - G_{Y_t}(q_{Y_t,\alpha})\Big| \\
\leq\ & c_{31}\cdot\Big|G_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big) - \hat{G}^{(IS)}_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big)\Big| + c_{31}\cdot\Big|\hat{G}^{(IS)}_{Y_t}\Big(\hat{q}^{(IS)}_{Y_t,\alpha}\Big) - \alpha\Big|
\end{aligned}
$$

$$
\begin{aligned}
= \quad & c_{31} \cdot \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - \hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right)\right| + c_{31} \cdot \left(\hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - \alpha\right) \\
\leq \quad & c_{31} \cdot \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - \hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right)\right| + c_{31} \cdot \left(\hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - \hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha} - \theta\right)\right) \\
\leq \quad & c_{31} \cdot \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - \hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right)\right| + c_{31} \cdot \left|\hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right)\right| \\
& + c_{31} \cdot \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha} - \theta\right)\right| \\
& + c_{31} \cdot \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha} - \theta\right) - \hat{G}^{(IS)}_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha} - \theta\right)\right|. \qquad (50)
\end{aligned}
$$

Since $G_{Y_t}(\cdot)$ is Lipschitz-continuous with Lipschitz constant c_7 for all $t \in [0,1]$ on $H_{t,n}$ for $n \in \mathbb{N}$ sufficiently large, which follows from assumption (18), we have

$$
\sup_{t\in[0,1]} \left| G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha}\right) - G_{Y_t}\left(\hat{q}^{(IS)}_{Y_t,\alpha} - \theta\right)\right| \leq c_7 \cdot \theta.
$$

Using this, inequality (50) and the fact that on the event J_n we have $\hat{q}^{(IS)}_{Y_t,\alpha} \in H_{t,n}$ as well as $\hat{q}^{(IS)}_{Y_t,\alpha} - \frac{1}{2} \cdot (\beta_n + \eta_n) \in H_{t,n}$, we get on the event J_n

$$
\sup_{t\in[0,1]} \left|\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}\right| \leq \sup_{y\in H_{t,n},\ t\in[0,1]} 3c_{31} \cdot \left| G_{Y_t}(y) - \hat{G}^{(IS)}_{Y_t}(y)\right|.
$$

for $n \in \mathbb{N}$ sufficiently large. Therefore, we obtain for

$$
s_n = 3c_{31} \cdot c_{24} \cdot \left((\beta_n + \eta_n) \cdot \sqrt{\frac{\log(n)}{nh_{n,1}}} + C_2 \cdot \beta^p \cdot h^p_{n,1} \right)
$$

the following inequality

$$
\begin{aligned}
& \mathbf{P}\left(\sup_{t\in[0,1]} \left|\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}\right| > s_n \right) \\
\leq \quad & \mathbf{P}(\{J_n \cap C_n\}^C) + \mathbf{P}\left(\left\{ \sup_{t\in[0,1]} \left|\hat{q}^{(IS)}_{Y_t,\alpha} - q_{Y_t,\alpha}\right| > s_n \right\} \cap \{J_n \cap C_n\} \right) \\
\leq \quad & \mathbf{P}(\{J_n \cap C_n\}^C) + \mathbf{P}\left(\left\{ \sup_{y\in H_{t,n},\ t\in[0,1]} 3c_{31} \cdot \left| G_{Y_t}(y) - \hat{G}^{(IS)}_{Y_t}(y) \right| > s_n \right\} \cap \{J_n \cap C_n\} \right).
\end{aligned}
$$

By applying Step 3, Step 5 and Step 6 the right–hand side can be bounded by

$$
\begin{aligned}
& \mathbf{P}\left(\{J_n \cap C_n\}^C\right) \\
& + \mathbf{P}\left(\left\{ \sup_{y\in\mathbb{R},\ t\in[0,1]} \left| \mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\} - \hat{G}^{(IS)}_{Y_t}(y) \right| \right.\right. \\
& \qquad \left.\left. + \sup_{y\in H_{t,n},\ t\in[0,1]} \left| \mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\} - G_{Y_t}(y) \right| > \frac{s_n}{3c_{31}} \right\} \cap C_n \right)
\end{aligned}
$$

$$\leq \quad \mathbf{P}\left(\left\{\sup_{y\in\mathbb{R},\ t\in[0,1]}\left|\mathbf{E}^{\star}_{t_1,\ldots,t_n}\left\{\hat{G}^{(IS)}_{Y_t}(y)\right\}-\hat{G}^{(IS)}_{Y_t}(y)\right|>c_{24}\cdot(\beta_n+\eta_n)\cdot\sqrt{\frac{\log(n)}{nh_{n,1}}}\right\}\cap C_n\right)$$
$$\mathbf{P}\left(\{J_n\cap C_n\}^C\right) \quad \to 0 \quad \text{for } n\to\infty$$

for some constant $c_{24} > 1$ and sufficiently large $n \in \mathbb{N}$. The proof is complete. □

6 Acknowledgment

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805.

References

[1] Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of the Mathematical Statistics*, **37**, pp. 577-580.

[2] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2-error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, **71**, pp. 93–107.

[3] Bott, A. K., Felber, T., Kohler, M. and Kristl, L. (2017). Estimation of a time-dependent density. *Journal of Statistical Planning and Inference*, **180**, pp. 81–107.

[4] Bhattacharya, P. K. and Gangopadhya, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* **18**, pp. 1400–1415.

[5] Cannamela, C., Garnier, J. and Ioos, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, vol. 2, no. 4, pp. 1554–1580.

[6] Chaudhuri, P. (1991). Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation. *The Annals of Statistics*, **2**, pp. 760-777.

[7] Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, bf 61, pp. 467–481.

[8] Devroye, L., Györfi, L., Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, **22**, pp. 1371–1385.

[9] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, **23**, pp. 71–82.

[10] Devroye, L. and Wagner, T.J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, **8**, pp. 231–239.

[11] Egloff, D. and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, vol. 38, no. 2, pp. 1244–1278.

[12] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of Conditional Densities and Sensitivity Measures. *Biometrika*, **83**, pp. 189-206.

[13] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, Springer Verlag, New York.

[14] Greblicki, W. and Pawlak, M. (1985). Fourier and Hermite series estimates of regression functions. *Annals of the Institute of Statistical Mathematics*, **37**, pp. 443–454.

[15] Györfi, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems of Control and Information Theory*, **10**, pp. 43–52.

[16] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics, Springer-Verlag, New York.

[17] Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, **46**, pp. 33-50.

[18] Kohler, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *Journal of Statistical Planning and Inference*, **89**, pp. 1–23.

[19] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory*, **47**, pp. 3054–3058.

[20] Kohler, M., Krzyżak, A., Tent, R. and Walk, H. (2014). Nonparametric quantile estimation using importance sampling. To appear in *Annals of the Institute of Statistical Mathematics.*

[21] Kohler, M. and Tent, R. (2015). Nonparametric quantile estimation using surrogate models and importance sampling. *Submitted for publication.*

[22] Kohler, M., Krzyżak, A. and Walk, H. (2003). Strong consistency of automatic kernel regression estimates. *Annals of the Institute of Statistical Mathematics*, **55**, pp. 287–308.

[23] Kohler, M. and Krzyżak, A. (2016). Estimation of extreme quantiles in a simulation model. *Submitted for publication.*

[24] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory*, **41**, pp. 677–687.

[25] Mehra, K. L., Rao, M. S. and Upadrasta, S. P. (1991). A Smooth Conditional Quantile Estimator and Related Applications of Conditional Empirical Processes. *Journal of Multivariate Analysis*, **37**, pp. 151-179.

[26] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, pp. 141–142.

[27] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications*, **15**, pp. 134-137.

[28] Morio, J. (2012). Extreme quantile estimation with nonparametric adaptive importance sampling. *Simulation Modelling Practice and Theory*, **27**, pp. 76–89.

[29] Rafajlowicz, E. (1987). Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in L2. *Statistics and Probability Letters*, **5**, pp. 219–224.

[30] Schuëller, G. (2007). On the treatment of uncertainties in structural mechanics and analysis. *Computers and Structures*, **85**, pp. 235–243.

[31] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, pp. 595–645.

[32] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, pp. 1040–1053.

[33] Stute, W. (1986). Conditional empirical processes. *Annals of Statistics*, **14**, pp. 638–647.

[34] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York.

[35] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia, PA.

[36] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, **26**, pp. 359-372.

[37] Xiang, X. (1996). A Kernel Estimator of a Conditional Quantile. *Journal of Multivariate Analysis*, **59**, pp. 206-216.

[38] Yu, K. and Jones, M. C. (1998). Local Linear Quantile Regression. *Journal of the American Statistical Association*, **93**, pp. 228-237.

[39] Yu, K., Lu, Z., and Stander, J. (2003). Quantile Regression: Applications and Current Research Areas. *Journal of the Royal Statistical Society: Series D*, **52**, pp. 331-350.

[40] Zill, D. G., and Wright, W. S. (2009). *Advanced Engineering Mathematics*, Jones & Bartlett Publishers, 4 edition, Sudburry.