Statistica Sinica **30** (2020), 135-151
doi:https://doi.org/10.5705/ss.202017.0276

NONPARAMETRIC ESTIMATION OF TIME-DEPENDENT QUANTILES IN A SIMULATION MODEL

Felix Heimrich, Michael Kohler and Lisa Kristl

Technische Universität Darmstadt

Abstract: This study considers 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]. It is assumed that the quantiles depend smoothly on t. Here we present the rate of convergence of quantile estimates based on a local average estimate of the time-dependent cumulative distribution functions. Then we apply importance sampling in a simulation model to construct estimates that achieve better rates of convergence. Lastly, we demonstrate the finite-sample performance of our estimates by applying them to simulated data.

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 the data set

$$\mathcal{D}_n = \left\{ Y_1^{(t_1)}, \ldots, Y_n^{(t_n)} \right\}, \tag{1.1}$$

where $Y_1^{(t_1)}, \ldots, Y_n^{(t_n)}$ are independent and

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

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

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

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,

$$\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}|. \tag{1.2}$$

Before we describe the construction of estimates for the above estimation problem, we illustrate the practical relevance of this problem using an example. Here

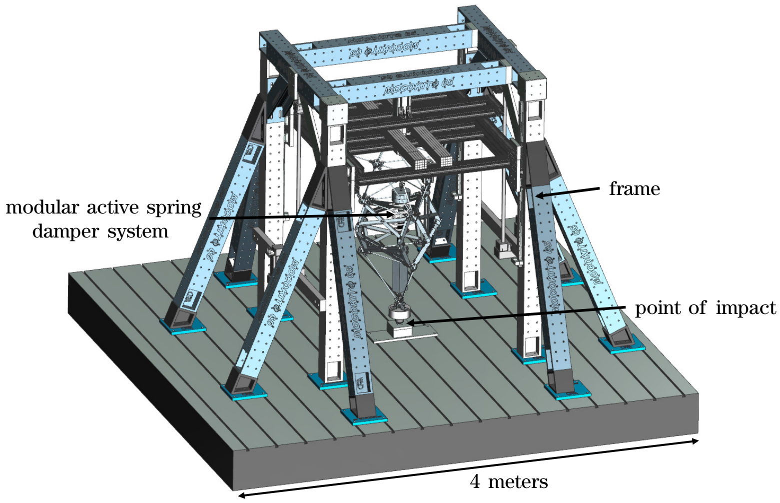

Figure 1. The suspension strut demonstrator of CRC 805.

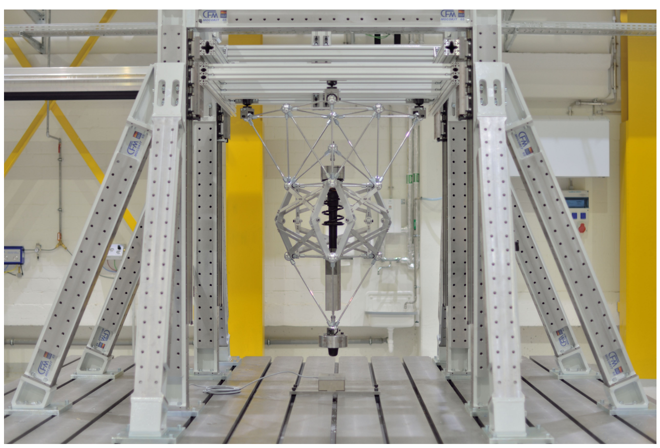

Figure 2. Picture of the experimental setup (picture taken from Mallapur and Platz (2017)).

we consider a problem that occured in the Collaborative Research Centre 805. The German CRC 805 examines ways in which uncertainty can be controlled in load-carrying structures in mechanical engineering. To test such approaches, the CRC 805 has designed a demonstrator model, which is shown in Figure 1.

This demonstrator model shows an example of a suspension strut, such as an aircraft's landing gear. It is designed in two versions: a virtual computer experiment and a real experimental setup. Figure 2 shows a photo of the real experimental setup. In the experiments, a modular spring damper system is suspended on a frame and falls onto the base of the frame. Then sensors measure various parameters, such as acceleration, the absolute position of the modular spring damper system, and the force at the point of impact. Predicting this force is important when calculating the stress and its deviation in order to determine the correct load capacity for the usage phase of a product that is already in its development phase. If one component of a suspension strut is time-dependent and uncertain, this force is affected and thus should be investigated before building a prototype. In this context, mechanical engineers need information about a complete time interval to, for example, guarantee the operational stability of the suspension strut. Here, a time-dependent quantile estimation can be used to estimate the α-quantiles for arbitrary time points in the considered interval.

We investigate the impact of an aging spring (i.e., the spring stiffness X_t decreases over time) on the force at the point of impact $Y_t = m(t, X_t)$. To do so, we use simulated data generated by the virtual demonstrator. Because we are describing how to support mechanical engineers in the product-development process, we assume there are no measured input data for an aging spring in this system. In a time-invariant system, the spring stiffness is assumed to be normally distributed with expectation $\mu = 35{,}000$ [N/m] and standard deviation $\sigma = 1{,}166.67$ [N/m] (cf., Schuëller (2007)). It seems reasonable that the spring will weaken over time with continuous use. Therefore, we assume that the spring constant deteriorates over time exponentially, as in Zill and Wright (2009) in Chapter 3.8.1. More precisely, we assume that the spring stiffness X_t is normally distributed with expectation $\mu_t = 35{,}000 * \exp(-0.5 * t)$ [N/m] and standard deviation $\sigma_t = 1{,}166.67$ [N/m].

For a given spring stiffness, we can use physical principles to model the force at the point of impact over time using partial differential equations which can be solved numerically. We do this implicitely using the routine *RecurDyn* of the software package *Siemens NX*. As such, we can compute the force at the point of impact Y_t for various (randomly) chosen values for the spring stiffness (chosen according to the distribution described above) and various time points t. In Subsection 4.2, we generate $n = 300$ values for the force at the point of impact Y_t in this way. Then, we use these values to estimate the time-dependent 0.95-quantile q_t of the force at the point of impact continuously over time t.

To compute a single value of Y_t, we repeat the computation of the force using independent data. This ensures that the data set (1.1) is independent in our example. Furthermore, because we can choose the time points for our simulation, the choice of equidistant time points does not pose any problems in this practical application. Finally, we need to consider the estimation error (1.2), because this allows us to make statistical inferences about the force at the point of impact for the overall time interval.

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} to define our quantile estimates. In particular, let $K : \mathbb{R} \to \mathbb{R}$ be a non-negative 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 the cdf of Y_t,

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

using the local average estimator

$$\hat{G}_{Y_t}(y) = \frac{\sum_{i=1}^{n} 1_{(-\infty,y]}(Y_i^{(t_i)}) \cdot K\left((t - t_i)/h_n\right)}{\sum_{j=1}^{n} K\left((t - t_j)/h_n\right)}. \tag{1.3}$$

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

$$\hat{q}_{Y_t,\alpha} = \inf\{y \in \mathbb{R} \,:\, \hat{G}_{Y_t}(y) \geq \alpha\}. \tag{1.4}$$

Here, we assume 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 that is bounded away from zero and infinity in a neighborhood of $q_{Y_t,\alpha}$. Then, we show that for a suitably chosen bandwidth h_n and kernel $K : \mathbb{R} \to \mathbb{R}$, the supremum norm error (1.2) of this estimate converges to zero in probability with rate $(\log(n)/n)^{p/(2p+1)}$.

In our second result, we use a simulation model to show that this rate of convergence can be improved 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 density $f(t, \cdot) : \mathbb{R}^d \to \mathbb{R}$ and the function $m : [0, 1] \times \mathbb{R}^d \to \mathbb{R}$ is a blackbox function that is costly to evaluate. In this framework, we construct an importance-sampling variant of the above plug-in quantile estimator. This variant is based on an initial quantile estimator and a suitably chosen estimator (surrogate) m_n of m. Our main result is that, under suitable assumptions on f and m, for Hölder smooth $G_{Y_t}(y)$ (with exponent $p \in (0, 1]$ and as a function of $t \in [0, 1]$), we can achieve the rate of convergence $(\log(n)/n)^{p\cdot(4p+1)/(2p+1)^2}$. This rate of convergence is achievable for an estimate based on at most n evaluations of m (as in (1.4)). We demonstrate the finite-sample performance of our estimates by applying them to simulated data.

1.2. Discussion of related results

A time-dependent quantile estimation can be regarded as a conditional quantile estimation for a fixed design, where we condition on the time t. A short introduction to conditional quantile estimations is presented in Yu, Lu and Stander (2003). Several studies have considered plug-in conditional quantile estimators. For example, Stone (1977) showed their consistency in probability, Stute (1986) proved their asymptotic normality, and Bhattacharya and Gangopadhya (1990) used a Bahadur-type representation, following Bahadur (1966), to show their 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 include those of Koenker and Bassett (1978), who proposed a quantile regression estimator, and Mehra, Rao and Upadrasta (1991), who presented a smooth conditional quantile estimator, showing its asymptotic normality and analyzing its pointwise almost sure rate of convergence. Xiang (1996) also proposed a new kernel estimator for a conditional quantile and derived the same pointwise almost sure rate of convergence as that shown in Mehra, Rao and Upadrasta (1991), but under weaker assumptions. Additional results on quantile regression estimators can be found in in Chaudhuri (1991), Fan, Yao and Tong (1996), Yu and Jones (1998), Li, Liu and Zhu (2007), and Plumlee and Tuo (2014), and in the literature cited therein. In contrast to the aforementioned works, 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; see, for example, Györfi et al. (2002).

Importance sampling is a well-known variance-reduction technique that was originally introduced to improve the rate of convergence of estimates of expectations; see, for example, Glasserman (2004). The main idea in our setting is as follows. Instead of Y_t, we consider 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 estimations based on importance sampling have been studied by Cannamela, Garnier, and Ioos (2008), Egloff and Leippold (2010), and Morio (2012). Of these works only Egloff and Leippold (2010) derived the theoretical properties for their estimate, such as consistency. However, they did not analyze the rate of convergence of their estimate. Kohler et al. (2018) studied the rates of convergence of importance-sampling quantile estimators based on surrogate models.

However, they did not consider a time-dependent setting for a conditional quantile estimation.

1.3. Outline

In Section 2, we present the rate of convergence of the first estimate. In Section 3, we consider a time-dependent simulation model and construct a time-dependent importance-sampling quantile estimate. Here, we also analyze the rate of convergence of the estimate. Section 4 illustrates the finite-sample performance of the two estimates by applying them to simulated data and to the application discussed in the introduction.

2. Estimation of Time-Dependent Quantiles

In this section, we analyze the rate of convergence of our local averaging plug-in quantile estimate. It is well known that the derivation of any nontrivial rate of convergence result in nonparametric curve estimations requires that various regularity conditions on the data hold.

As a result, we make the following assumptions on the data used to estimate the quantile.

(A1) Let $(Y_t)_{t\in[0,1]}$ be an $\mathbb{R}$-valued stochastic process, let $t_1,\ldots,t_n \in [0,1]$ be equidistant, and set

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

where $Y_1^{(t_1)},\ldots,Y_n^{(t_n)}$ are independent and

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

(A2) Assume that Y_t has 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 $q_{Y_t,\alpha}$: that is, for some $\epsilon>0$, there exists a constant $c_1>0$ such that

$$\inf_{t\in[0,1]}\ \inf_{u\in(q_{Y_t,\alpha}-\epsilon,q_{Y_t,\alpha}+\epsilon)} g(t,u) \geq c_1. \tag{2.1}$$

(A3) Assume 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 $y\in\mathbb{R}$: that is assume

$$|G_{Y_s}(y)-G_{Y_t}(y)| \leq C|s-t|^p \quad \text{for all } s,t\in[0,1] \text{ and } y\in\mathbb{R}. \tag{2.2}$$

Here $(A3)$ is the standard smoothness assumption, which is necessary for the derivation of a nontrivial result on the rate of convergence. Assumption $(A2)$ is needed because we want to analyze the error of our estimate in the supremum norm. Assumption $(A1)$ is used to simplify our proofs. As explained in the introduction, we can ensure that this condition holds by generating our data in a proper way.

Next we describe the assumptions needed for the parameters of our estimate, that is, the assumptions on the kernel and the bandwidth.

(EST1) Let K be a non-negative kernel function $K : \mathbb{R} \to \mathbb{R}$ that is left-continuous on $\mathbb{R}_+$, monotonically decreasing on $\mathbb{R}_+$, and satisfies

$$K(z) = K(-z) \quad (z \in \mathbb{R}), \tag{2.3}$$

$$c_2 \cdot \mathbb{1}_{[-\alpha,\alpha]}(z) \leq K(z) \leq c_3 \cdot \mathbb{1}_{[-\beta,\beta]}(z) \quad (z \in \mathbb{R}) \tag{2.4}$$

for some constants $\alpha, \beta, c_2, c_3 \in \mathbb{R}_+ \setminus \{0\}$.

(EST2) Let $h_n > 0$ be such that

$$h_n \to 0 \quad (n \to \infty), \tag{2.5}$$

$$\frac{n \cdot h_n}{\log(n)} \to \infty \quad (n \to \infty). \tag{2.6}$$

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. Let $n \in \mathbb{N}$ and set $t_k = k/n$ $(k = 1,\ldots,n)$. Assume that* (A1), (A2)*, and* (A3) *hold. Let the estimator $\hat{q}_{Y_t,\alpha}$ be defined by* (1.3) *and* (1.4)*, where the kernel and the bandwidth satisfy (EST1) and (EST2), respectively. Then, for a constant $c_4 > 0$, we have*

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

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

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

Remark 1. To derive a nontrivial statement about the rate of convergence, smoothness assumptions, such as (2.2), are required (cf. Devroye (1982)). In Theorem 1, we have shown the same rate of convergence as the optimal minimax rate of convergence for the estimation of a Hölder continuous function (with exponent $p \in (0,1]$) on a compact subset of $\mathbb{R}$ in the 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 choosing h_n in an optimal way for 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 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 (i.i.d.) 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 (1.3) 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

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

Remark 3. It follows from the proof of Theorem 1 that the result also holds when $t_k \in [0,1]$ are chosen such that

$$\frac{c_7}{n} \leq \left| t_k - \frac{k}{n} \right| \leq \frac{c_8}{n}$$

for some constants $c_7 > 0$ and $c_8 > 0$.

3. Application of Importance Sampling in a Simulation Model

In this section, we use additional assumptions on the given data and importance-sampling to construct an estimate that achieves (for essential the same smoothness of the underlying cdf) a better rate of convergence than that in Theorem 1. Here, we assume (as in our application in the introduction) that Y_t is given by some $m(t, X_t)$, where we know the distribution of X_t and can observe X_t and $m(t, X_t)$. Then, we change the density of X_t in such a way that it is concentrated on a set (with a small measure), which is important for the estimation of the quantile of $m(t, X_t)$ (cf., the density $h(t, x)$, defined below).

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

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

In the following we assume we have independent data sets $\mathcal{D}_{n,1}$ and $\mathcal{D}_{n,2}$ of the form

$$\mathcal{D}_{n,1} = \left\{ \left(t_1, X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)} \right), \ldots, \left(t_n, X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)} \right) \right\},$$

$$\mathcal{D}_{n,2} = \left\{ \left(t_1, X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)}\right), \dots, \left(t_n, X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)}\right) \right\}, \tag{3.1}$$

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

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

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

$$\big(X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)}\big), \dots, \big(X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)}\big), \big(X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)}\big), \dots, \big(X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)}\big) \tag{3.2}$$

are independent. That is, we have two independent samples of (X_{t_k}, Y_{t_k}) at each time point t_k $(k = 1, \dots, n)$. As stated previously this independence assumption can be justified in our application by generating the data properly.

Furthermore, we assume we have independent random variables $X_{k,3}^{(t_k)}, X_{k,4}^{(t_k)}$, $\dots$, distributed as X_{t_k} for $k = 1, \dots, n$, 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

$$\mathbf{P}\left(\sup_{t \in [0,1],\ x \in K_n} |m_n(t,x) - m(t,x)| \geq \beta_n \right) \to 0 \quad (n \to \infty) \tag{3.3}$$

for some sequence $(\beta_n)_{n \in \mathbb{N}} \in \mathbb{R}_+$ 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

$$\mathbf{P}\left(\sup_{t \in [0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| \geq \eta_n \right) \to 0 \quad (n \to \infty), \tag{3.4}$$

for some sequence $(\eta_n)_{n \in \mathbb{N}} \in \mathbb{R}_+$, which converges to zero as n tends to infinity, for example the estimator $\hat{q}_{Y_{t,\alpha}}$ defined in (1.4) and $\eta_n = 2 \cdot c_6 \cdot (\log(n)/n)^{p/(2p+1)}$ (cf., Theorem 1). Assume that

$$\mathbf{P}\left(\exists t \in [0,1] \,:\, X_t \notin K_n\right) = O\left(\beta_n + \eta_n\right). \tag{3.5}$$

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

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

for $k = 1, \dots, n$. Define (for some $h_{n,1} > 0$)

$$\hat{G}^{(IS)}_{Y_t}(y) = \frac{\sum_{i=1}^n \left(c_{t_i} \cdot \mathbb{1}_{\{m(t_i, Z_{t_i}) \leq y\}} + b_{t_i} \right) \cdot K\left((t-t_i)/h_{n,1}\right)}{\sum_{j=1}^n K\left((t-t_j)/h_{n,1}\right)}, \tag{3.6}$$

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

$$\hat{q}^{(IS)}_{Y_t,\alpha} = \inf\{y \in \mathbb{R} \,:\, \hat{G}^{(IS)}_{Y_t}(y) \geq \alpha\}. \tag{3.7}$$

In order to analyze the rate of convergence of this estimate, we assume (A3), as well as the following.

(A4) Assume that Y_t has density $g(t,\cdot) : \mathbb{R} \to \mathbb{R}$ that is continuous and uniformly bounded away from zero in a neighborhood of $q_{Y_t,\alpha}$, and that is uniformly bounded from above. That is, we assume that (2.1) holds and that there exists a constant $c_9 > 0$ such that

$$\sup_{t \in [0,1]} \sup_{u \in \mathbb{R}} g(t,u) \leq c_9. \tag{3.8}$$

(A5) For $\alpha \in (0,1)$, 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 $q \in (0,1]$, that is

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

Here, (A4) is a slightly stronger version of (A2), and (A5) is an additional smoothness assumption on the quantiles.

With regard to the parameters of our estimate, that is, the kernel, the bandwidth, the estimate of m, and the original quantile estimate, we assume ($EST1$), as well as the following.

(EST3) The estimate m_n of m satisfies (3.3) for some $\beta_n > 0$, where

$$\beta_n \to 0 \quad \text{for} \quad n \to \infty. \tag{3.9}$$

(EST4) The estimate $\hat{q}_{Y_t,\alpha}$ of $q_{Y_t,\alpha}$ satisfies (3.4) for some $\eta_n \in \mathbb{R}_+$, where

$$\eta_n \to 0 \quad \text{for} \quad n \to \infty. \tag{3.10}$$

(EST5)

$$h_{n,1} \to 0 \quad \text{for} \quad n \to \infty, \tag{3.11}$$

and

$$\frac{n \cdot h_{n,1}}{\log(n)} \to \infty \quad \text{for} \quad n \to \infty \tag{3.12}$$

(EST6) For $r = \min\{p, q\}$, we have

$$\frac{h_{n,1}^r}{\beta_n + \eta_n} \to 0 \quad \text{for} \quad n \to \infty. \tag{3.13}$$

Here, (EST3) and (EST4) mean that the errors of the estimate of m and our initial quantile estimate must vanish asymptotically, and ($EST6$) requires that the bandwidth $h_{n,1}$ is not too small in comparison with these errors. These assumptions require rates of convergence for the estimate of m and our initial quantile estimate. For the latter estimate, the result is given by Theorem 1. For the estimate of m, the result requires a smoothness assumption on m (see, e.g., Györfi et al. (2002) for related rates of convergence for various estimates).

Theorem 2. *Assume that $(X_t)_{t\in[0,1]}$ is an $\mathbb{R}^d$-valued stochastic process such that X_t has 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 $\alpha \in (0,1)$, let $q_{Y_t,\alpha}$ be the α-quantile of Y_t for $t \in [0,1]$, and assume that* (A3), (A4), *and* (A5) *hold. Let $n \in \mathbb{N}$ and set $t_k = k/n$ $(k = 1,\dots,n)$. Assume that the kernel K satisfies (EST1). Let the estimator $\hat{q}_{Y_t,\alpha}^{(IS)}$ be defined by* (3.6) *and* (3.7) *with $h_{n,1} > 0$, and assume that (EST3), (EST4), (EST5), and (EST6) hold. Furthermore, assume that* (3.5) *is satisfied. Then, there exists a constant $c_{10} > 0$ such that*

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

for $\quad n \to \infty.$

In particular, if we set $h_{n,1} = c_{11}\cdot(\beta_n+\eta_n)^{2/(2p+1)}\cdot\big(\log(n)/n\big)^{1/(2p+1)}$ for some constant $c_{11} > 0$, there exists a constant $c_{12} > 0$ such that

$$\mathbf{P}\left(\sup_{t\in[0,1]} |\hat{q}_{Y_t,\alpha}^{(IS)} - q_{Y_t,\alpha}| > c_{12}\cdot(\beta_n+\eta_n)^{2p/(2p+1)}\cdot\left(\frac{\log(n)}{n}\right)^{p/(2p+1)}\right) \to 0$$

for $\quad n \to \infty.$

Remark 4. If $\eta_n = 2\cdot c_6\cdot(\log(n)/n)^{(p/2p+1)}$, as in Theorem 1, and $\beta_n \le \eta_n$ is satisfied, we get $(\log(n)/n)^{p\cdot(4p+1)/(2p+1)^2}$ as the rate of convergence in Theorem 2. Hence, in this case, the rate of convergence in Theorem 2 is better than that in Theorem 1, although it requires stronger conditions than those in Theorem 1. However, the basic smoothness assumption on the cdf (in particular, the value of p) is the same in both theorems.

Remark 5. Observations of $Z_k^{(t_k)}$ for $k = 1,\dots,n$ can be generated using the rejection method. Here, we use several observations of $(t_k, X_k^{(t_k)})$ for each $k = 1,\ldots,n$, and then select the first observation that satisfies either

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

Remark 6. Because 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 for selecting β_n and η_n is presented in Section 4.

Remark 7. As for the first time-dependent estimator, a bandwidth $h_{n,1}$ has to be selected in a data-driven way to apply the importance-sampling quantile estimator. To do this, we suggest proceeding as in Remark 2, except that we use the importance-sampling random variables. In particular, 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 i.i.d., and 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

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

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 8. In Section 4, we use a 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. $N = 10,000$), to approximate the integrals in c_{t_k} and b_{t_k} for $k = 1,\ldots,n$ by

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

4. The Finite-sample Performance of the Estimates

4.1. Application to simulated data

Next, we examine the finite-sample performance of the local average-based time-dependent quantile estimator $\hat{q}_{Y_t,\alpha}$ defined in (1.4) and the importance-sampling time-dependent quantile estimator $\hat{q}^{(IS)}_{Y_t,\alpha}$ defined in (3.7) by applying them to the simulated data. Both estimators use the same number $3n$ of evaluations of m. For the local average-based quantile estimator $\hat{q}_{Y_t,\alpha}$, we derive these using three independent copies of Y_{t_k} for each time point $t_k = k/n$ $(k = 1, \ldots, n)$. Here,

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

is used for the main quantile estimation, and

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

is used as test data for the data-driven bandwidth selection method described in Remark 4, where for each $k = 1, \ldots, n$, we compare $Y^{(t_k)}_{k,1}$ and $Y^{(t_k)}_{k,2}$ with $Y^{(t_k)}_{k,3}$.

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 $X^{(t_k)}_{k,3}, X^{(t_k)}_{k,4}, \ldots$ of X_{t_k}. These copies are used for the generation of $Z^{(t_k)}_{k,1}$ and $Z^{(t_k)}_{k,2}$ for $k = 1, \ldots, n$ and for the integral approximation by the Monte Carlo simulation in the estimation of c_{t_k} and b_{t_k} (cf., Remark 8) for $k = 1, \ldots, n$ and $N = 10,000$. To generate observations of $Z^{(t_k)}_{k,1}$ and $Z^{(t_k)}_{k,2}$ for $k = 1, \ldots, n$ by applying the rejection method presented in Remark 5, we need a surrogate model m_n of m and its approximation error β_n (see (3.3)), and an initial quantile estimation and its estimation error η_n (see (3.4)). The local average-based time-dependent quantile estimator $\hat{q}_{Y_{t_k},\alpha}$ described in Section 2 is used for the initial quantile estimation. Here, η_n is unknown, because the Hölder exponent p of the smoothness condition in Theorem 1 is unknown.

A data set $\mathcal{D}_{n,1}$, as described in (3.1), 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 using 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 $(t_k, Y^{(t_k)}_{k,1})$ or $(t_k, Y^{(t_k)}_{k,2})$ randomly from $\mathcal{D}_{n,1}$ or $\mathcal{D}_{n,2}$ as learning or test data sets. We repeat the procedure 30 times to obtain multiple learning and test data sets and to estimate $q_{Y_{t_k},\alpha}$ by $\hat{q}_{Y_{t_k},\alpha}$ $k = 1, \ldots, n$ 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 using a smoothing spline estimator (here we use the routine *Tps()* in the statistics package *R*) on the data set $\mathcal{D}_{n,2}$. To estimate β_n in a data-driven way, we suggest employing a cross-validation method. First, we split $\mathcal{D}_{n,2}$ into five parts. Then, for $j = 1, \ldots, 5$,

Table 1. Results for the 0.95-quantile estimation by $\hat{q}_{Y_{t,\alpha}}$ and $\hat{q}^{(IS)}_{Y_{t,\alpha}}$. The values are the maximal absolute errors, with the relative errors shown in paratheses.

		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)

we approximate $m_n^{(j)}$ of m using the data $\mathcal{D}_{n,2}$ without the jth part, and use the jth part as test 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 absolute error of $m_n^{(j)}$ for each time point and choose β_n as the mean of these maximal errors.

Now, $Z^{(t_1)}_{1,1}, \ldots, Z^{(t_n)}_{n,1}$ and $Z^{(t_1)}_{1,2}, \ldots, Z^{(t_n)}_{n,2}$ can be generated according to Remark 5 for some K_n, where we suggest using $K_n = [-\hat{c} \cdot \log(n), \hat{c} \cdot \log(n)]$ for some constant $\hat{c}>0$ (cf., Table 1).

We compare the two time-dependent quantile estimators for three different models. In all three models, we consider $n_1 = 50$, $n_2 = 100$, and $n_3 = 200$ equidistant time points in the time interval [0, 1] (i.e., 150, 300 or 600 evaluations of the function m, overall) and estimate the time-dependent 0.95-quantiles. Because it is not possible to compare the errors in the supremum norm (1.2), we compare the maximal absolute errors

$$\max_{t\in\{t_1,\ldots,t_n\}} |\hat{q}_{Y_{t,\alpha}} - q_{Y_{t,\alpha}}| \quad \text{to} \quad \max_{t\in\{t_1,\ldots,t_n\}} |\hat{q}^{(IS)}_{Y_{t,\alpha}} - q_{Y_{t,\alpha}}|. \tag{4.1}$$

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

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

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

In the second model, X_t follows the distribution $\mathcal{N}(0, (t^2 - t^4 + 1/2)^2)$ and m is given by

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

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

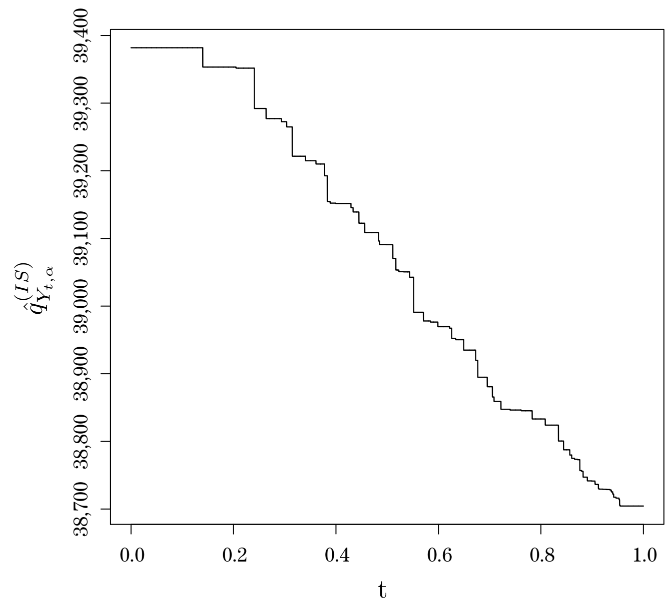

Figure 3. The 0.95-quantile of the force at the point of impact estimated by $\hat{q}^{(IS)}_{Y_{t,0.95}}$.

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

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 chosen constant $\hat{c}$ in the interval K_n. As expected, the importance-sampling time-dependent quantile estimation $\hat{q}^{(IS)}_{Y_{t,\alpha}}$ outperforms the local average-based quantile estimation $\hat{q}_{Y_t,\alpha}$ as the sample size increases. Moreover, for both estimators, the estimation becomes more accurate as the sample size increases. A comparison of the errors within the three different models shows that the relative error in Model 1 is larger than those in the other models. We believe this is because the distributions in Model 1 have much larger tails than those of the other models, which makes estimating the quantiles more difficult.

4.2. Analysis of the effect of an aging spring on the force at the point of impact

Finally, we apply the proposed estimation methods to the practical problem described in the introduction. As before, we use three observations of Y_t at

$n = 100$ time points, that is, 300 evaluations of the computer experiment m. As in the previous subsection, for the importance-sampling estimator, a surrogate model m_n of the underlying function m is estimated using a smoothing spline estimator. The bandwidth h is chosen as described in Remark 7. Because the true quantiles are unknown, we present only the 0.95-quantiles estimated by the importance-sampling estimator $\hat{q}^{(IS)}_{Y_{t,0.95}}$. The results are shown in Figure 3. The figure shows that less force acts on the point of impact as the spring stiffness decreases over time.

Supplementary Material

The Supplementary Material contains the proofs of Theorem 1 and Theorem 2.

Acknowledgments

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805 (Projektnummer 57157498 - SFB 805). Furthermore, the authors would like to thank the Associate Editor and two anonymous referees for their helpful comments on an earlier version of the manuscript.

References

Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of the Mathematical Statistics* **37**, 577–580.

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

Cannamela, C., Garnier, J. and Ioos, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics* **2**, 1554–1580.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local bahadur representation. *The Annals of Statistics* **2**, 760–777.

Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* **61**, 467–481.

Egloff, D. and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics* **38**, 1244–1278.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures. *Biometrika* **83**, 189–206.

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

Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Non-*

parametric Regression. Springer Series in Statistics, Springer-Verlag, New York.

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

Kohler, M., Krzyżak, A., Tent, R. and Walk, H. (2018). Nonparametric quantile estimation using importance sampling. *Annals of the Institute of Statistical Mathematics* **70**, 439–465.

Li, Y., Liu, Y. and Zhu, J. (2007). Qunatile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association* **102**, 255–268.

Mallapur, S. and Platz, R. (2017). Quantification and evaluation of uncertainty in the mathematical modelling of a suspension strut using bayesian model validation approach. In: IMAC 2017, 30.01.-02.02.2017, Garden Grove, California USA.

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**, 151–179.

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

Plumlee, M. and Tuo, R. (2014). Building accurate emulators for stochastic simulations via quantile Kriging. *Technometrics* **56**, 466–473.

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

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics* **5**, 595–645.

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

Stute, W. (1986). Conditional empirical processes. *The Annals of Statistics* **14**, 638–647.

Xiang, X. (1996). A kernel estimator of a conditional quantile. *Journal of Multivariate Analysis* **59**, 206–216.

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

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**, 331–350.

Zill, D. G. and Wright, W. S. (2009). *Advanced Engineering Mathematics*, 4th Edition. Jones & Bartlett Publishers, Sudburry.

Fachbereich Maschinenbau, Technische Universität Darmstadt, Otto-Bernd-Str. 2, 64287 Darmstadt, Germany.

E-mail: heimrich@dik.tu-darmstadt.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany.

E-mail: kohler@mathematik.tu-darmstadt.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany.

E-mail: kristl@mathematik.tu-darmstadt.de

(Received February 2017; accepted March 2018)