

# Conditional Quantile Estimation based on Optimal Quantization: from Theory to Practice

Isabelle Charlier, Davy Paindaveine, Jérôme Saracco

► **To cite this version:**

Isabelle Charlier, Davy Paindaveine, Jérôme Saracco. Conditional Quantile Estimation based on Optimal Quantization: from Theory to Practice. Computational Statistics and Data Analysis, Elsevier, 2015. hal-01108504v2

**HAL Id: hal-01108504**

**<https://hal.inria.fr/hal-01108504v2>**

Submitted on 13 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Conditional Quantile Estimation based on Optimal Quantization: From Theory to Practice<sup>1</sup>

Isabelle Charlier<sup>a,b</sup>, Davy Paindaveine<sup>a,\*</sup>, Jérôme Saracco<sup>b</sup>

<sup>a</sup>*Université Libre de Bruxelles, ECARES and Département de Mathématique, 50 Avenue F.D. Roosevelt, CP114/04, B-1050, Brussels, Belgium.*

<sup>b</sup>*Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 and INRIA Bordeaux Sud-Ouest, équipe CQFD, 351 Cours de la Libération, 33405 Talence, France.*

---

## Abstract

Small-sample properties of a nonparametric estimator of conditional quantiles based on optimal quantization, that was recently introduced (J. Statist. Plann. Inference, 156, 14–30, 2015), are investigated. More precisely, (i) the practical implementation of this estimator is discussed (by proposing in particular a method to properly select the corresponding smoothing parameter, namely the number of quantizers) and (ii) its finite-sample performances are compared to those of classical competitors. Monte Carlo studies reveal that the quantization-based estimator competes well in all cases and sometimes dominates its competitors, particularly when the regression function is quite complex. A real data set is also treated. While the main focus is on the case of a univariate covariate, simulations are also conducted in the bivariate case.

*Keywords:* Conditional quantiles, Optimal quantization, Nonparametric regression.

---

## 1. Introduction

Quantile regression is used to quantify the relationship between a univariate response variable  $Y$  and a  $d$ -dimensional covariate  $X$ . Since the conditional mean only models the average relationship between these, standard (mean) regression is way less informative

---

<sup>1</sup>For the interpretation of the references to color in some of the figures, the electronic version of the paper should be consulted.

\*Corresponding author

*Email addresses:* [ischarli@ulb.ac.be](mailto:ischarli@ulb.ac.be) (Isabelle Charlier), [dpaindav@ulb.ac.be](mailto:dpaindav@ulb.ac.be) (Davy Paindaveine), [Jerome.Saracco@math.u-bordeaux1.fr](mailto:Jerome.Saracco@math.u-bordeaux1.fr) (Jérôme Saracco)

than quantile regression, that provides a more complete picture of conditional distributions: quantile regression may indeed capture heteroscedasticity, conditional asymmetry, etc. Conditional quantiles allow to construct reference hypersurfaces (curves if the covariate is univariate) and conditional prediction intervals, hence are widely used in many different areas, including medicine, economics or lifetime analysis.

Quantile regression was first introduced by [Koenker and Bassett \(1978\)](#) in a linear framework. Throughout this paper, we consider quantile regression in the more general nonparametric regression setting. Many methods have been proposed to estimate conditional quantiles in this context. In particular, nearest-neighbor estimators were investigated in [Bhattacharya and Gangopadhyay \(1990\)](#), and local linear and double-kernel approaches were considered in [Yu and Jones \(1998\)](#). [Koenker et al. \(1994\)](#) and [He et al. \(1998\)](#) developed spline-based methods, while [Lee \(2003\)](#), [Horowitz and Lee \(2005\)](#) or [Belloni et al. \(2011\)](#) focused on global nonparametric series methods. Many other estimators have also been investigated; we refer to, e.g., [Chaudhuri \(1991\)](#), [Fan et al. \(1994\)](#), [Hallin et al. \(2009\)](#), [Gannoun et al. \(2002\)](#) or [Yu et al. \(2003\)](#), and to [Koenker \(2005\)](#) for a review.

In this work, we consider the problem of estimating, in a nonparametric way, the conditional quantiles of a scalar response variable  $Y$  with respect to a  $d$ -dimensional covariate  $X$ . The theoretical properties of the method we will focus on were actually derived in a more stringent setup under which

$$Y = m(X, \varepsilon), \tag{1.1}$$

where  $\varepsilon$  is a random error term that is independent of  $X$ . It is further assumed that both  $X$  and  $\varepsilon$  are absolutely continuous and that the function  $(x, z) \mapsto m(x, z)$  is of the form  $m_1(x) + m_2(x)z$ , where  $m_1(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $m_2(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}_0^+$  are Lipschitz functions. In practice, both this link function  $m$  and the distribution of  $\varepsilon$  are unspecified, so that the conditional distribution of  $Y$  given  $X = x$  is unknown. Therefore it needs to be estimated by using a random sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  from (1.1). All consistent estimators of conditional quantiles in this context require some localization or penalization technique; four of the most classical estimators, that are used as competitors to our own approach in the sequel, are described in Section 4.1.

The method we will focus was introduced in [Charlier, Paidaveine, and Saracco \(2015\)](#)

— hereafter, [CPS15](#) — and performs localization through *optimal quantization*, a concept that projects the covariates  $X_i$ ,  $i = 1, \dots, n$  onto a finite grid of  $N$  *quantizers*. This projection of the  $X_i$ 's proceeds according to their belongings to appropriate *Voronoi cells* that are partitioning the covariate space  $\mathbb{R}^d$  (see [Pagès \(1998\)](#) for more details on this Voronoi partition). The [CPS15](#) estimator can therefore be viewed as a nonparametric partitioning regression estimator. Partitioning ideas were already used in the construction of least square estimators; see, e.g., [Györfi et al. \(2002\)](#) for a complete overview and [Kohler et al. \(2006\)](#) or [Cattaneo and Farrell \(2013\)](#) for more recent results.

Now, the emphasis in [CPS15](#) was almost exclusively on theoretical aspects; the main results there (i) are quantifying how well quantization allows to approximate conditional quantiles as  $N$  grows and (ii) are providing the convergence in probability of a sample quantization-based conditional quantile to its population version (see [\(2.3\)](#) and [\(2.4\)](#) below, respectively). Practical implementation, however, was barely touched in [CPS15](#), and finite-sample performances were not investigated. The goal of the present work is to extensively cover these crucial aspects.

The outline of the paper is as follows. For the sake of completeness, we describe in [Section 2](#) the concept of optimal quantization and the quantization-based estimators from [CPS15](#). Then we discuss in [Section 3](#) the choice of the number  $N$  of quantizers, that, in the nonparametric setup considered, plays the role of a smoothing parameter. In [Section 4](#), we compare our estimator with well-known nonparametric estimators of conditional quantiles, namely (see [Section 4.1](#)) quantile regression splines,  $k$  nearest-neighbor ( $k$ NN) estimators, and (kernel) local linear/constant estimators. The comparison is based on visual inspection of estimated conditional quantile curves ([Section 4.2](#)) and on empirical integrated squared errors ([Section 4.3](#)). We provide a real data example in [Section 5](#) and investigate, in [Section 6](#), the case where the number  $d$  of covariates is larger than one. We conclude with a summary and some final comments in [Section 7](#).

## 2. Conditional quantiles through optimal quantization

### 2.1. Quantization-based approximations of conditional quantiles

Let  $X$  be a random  $d$ -vector defined on a probability space  $(\Omega, \mathcal{F}, P)$ , such that the  $L_2$ -norm  $\|X\| = (\mathbb{E}[|X|^2])^{1/2}$  is finite (here,  $|\cdot|$  denotes the Euclidean norm). For any fixed

positive integer  $N$ , quantization replaces  $X$  with the discrete vector  $\tilde{X}^{\gamma^N} = \text{Proj}_{\gamma^N}(X)$  obtained by projecting  $X$  onto an  $N$ -grid  $\gamma^N$  of  $\mathbb{R}^d$  (so that  $\gamma^N \in (\mathbb{R}^d)^N$ ). Optimal quantization is achieved if the grid minimizes the *quantization error*  $\|X - \tilde{X}^{\gamma^N}\|$ . In the sequel,  $\tilde{X}^N$  will denote the approximation obtained by projecting  $X$  onto an arbitrary optimal  $N$ -grid (unlike existence, unicity does not always hold; see [Pagès \(1998\)](#)). Of course, it is expected that this approximation becomes arbitrarily precise as  $N$  increases to infinity, which is confirmed by the fact that

$$\|X - \tilde{X}^N\| = O(N^{-1/d}) \quad \text{as } N \rightarrow \infty; \quad (2.1)$$

see, e.g., [Graf and Luschgy \(2000\)](#).

Denoting by  $\mathbb{I}_A$  the indicator function of the set  $A$ , consider the well-known *check function*  $z \mapsto \rho_\alpha(z) = z(\alpha - \mathbb{I}_{[z < 0]}) = -(1 - \alpha)z\mathbb{I}_{[z < 0]} + \alpha z\mathbb{I}_{[z \geq 0]}$ ; see [Koenker and Bassett \(1978\)](#). In a regression context involving the scalar response  $Y$  and the (continuous) covariate  $d$ -vector  $X$ , this function allows to characterize the conditional  $\alpha$ -quantile  $q_\alpha(x)$  of  $Y$  given  $X = x$  as  $q_\alpha(x) = \arg \min_{a \in \mathbb{R}} \mathbb{E}[\rho_\alpha(Y - a) | X = x]$ . In this context, [CPS15](#) proposed approximating  $q_\alpha(x)$  by

$$\tilde{q}_\alpha^N(x) = \arg \min_{a \in \mathbb{R}} \mathbb{E}[\rho_\alpha(Y - a) | \tilde{X}^N = \tilde{x}], \quad (2.2)$$

where  $\tilde{x}$  denotes the projection of  $x$  onto the optimal  $N$ -grid  $\gamma^N$  defining  $\tilde{X}^N$ ; here,  $N$  is a fixed integer whose choice in practice will be discussed below. [CPS15](#) proved that, under mild regularity assumptions, there exists a constant  $C$  such that  $\|\tilde{q}_\alpha^N(X) - q_\alpha(X)\| \leq C\|\tilde{X}^N - X\|^{1/2}$  for  $N$  sufficiently large, which, in view of (2.1), implies that

$$\|\tilde{q}_\alpha^N(X) - q_\alpha(X)\| = O(N^{-1/2d}) \quad \text{as } N \rightarrow \infty. \quad (2.3)$$

Since pointwise results are also appealing in quantile regression, they further showed that  $\tilde{q}_\alpha^N(x) \rightarrow q_\alpha(x)$  as  $N \rightarrow \infty$ , uniformly in  $x$  (in the support  $S_X$  of  $X$ ).

## 2.2. Quantization-based estimators of conditional quantiles

In real data applications, independent copies  $(X_1, Y_1), \dots, (X_n, Y_n)$  of  $(X, Y)$  are available, and we have to define a sample version,  $\hat{q}_\alpha^{N,n}(x)$  say, of the quantization-based population quantile  $\tilde{q}_\alpha^N(x)$ . As described in [CPS15](#), this can be achieved in the following two steps :

- (S1) First, an optimal grid is obtained by making use of the CLVQ algorithm described below, with  $n$  iterations (based on  $\xi^t = X_t$ ,  $t = 1, \dots, n$ ) and an initial  $N$ -grid  $X^0$  that results from sampling randomly among the  $X_i$ 's without replacement under the constraint that an  $x$ -value cannot be picked more than once;
- (S2) letting  $\widehat{X}_i^{N,n} = \text{Proj}_{\widehat{\gamma}^{N,n}}(X_i)$ ,  $i = 1, \dots, n$ , and  $\widehat{x}^{N,n} = \text{Proj}_{\widehat{\gamma}^{N,n}}(x)$ , where  $\widehat{\gamma}^{N,n} = (\widehat{x}_1^{N,n}, \dots, \widehat{x}_N^{N,n})$  is the optimal grid obtained in (S1), the proposed empirical conditional quantile  $\widehat{q}_\alpha^{N,n}(x)$  is the sample  $\alpha$ -quantile of the  $Y_i$ 's whose corresponding  $\widehat{X}_i^{N,n}$  is equal to  $\widehat{x}^{N,n}$ . Throughout, we will use the sample quantile concept that is associated with the default value 7 of the `type` argument in the R function `quantile`.

It was shown in [CPS15](#) that, for any fixed  $N$  and  $x$ ,

$$\widehat{q}_\alpha^{N,n}(x) \xrightarrow{P} \widetilde{q}_\alpha^N(x), \quad \text{as } n \rightarrow \infty. \quad (2.4)$$

In other words, the empirical conditional quantile  $\widehat{q}_\alpha^{N,n}(x)$  is weakly consistent for the — as we have seen, arbitrarily good — approximation  $\widetilde{q}_\alpha^N(x)$  of the true conditional quantile  $q_\alpha(x)$ . In principle, the convergence  $\widetilde{q}_\alpha^N(x) - q_\alpha(x) \rightarrow 0$  as  $N \rightarrow \infty$  could be combined with (2.4) to provide an asymptotic result stating that, if  $N = N_n$  goes to infinity in an appropriate way, then  $\widehat{q}_\alpha^{N,n}(x) - q_\alpha(x) \xrightarrow{P} 0$  as  $n \rightarrow \infty$ . Proving formally such a result, however, is extremely difficult, since all convergence results for the CLVQ algorithm below are as  $n \rightarrow \infty$  *with  $N$  fixed*. In any case, this is of academic interest only; in practice, indeed, the sample size  $n$  is of course fixed and the user needs to choose appropriately the size  $N$  of the quantization grid. We later on propose a method that allows to choose this smoothing parameter  $N$  by achieving a bias-variance trade-off (see the discussion of [Figure 1](#) below).

The so-called *Competitive Learning Vector Quantization* (CLVQ) algorithm, that provides an optimal grid in (S1) above, is a stochastic gradient algorithm that proceeds as follows. Let  $(\xi^t)_{t \in \mathbb{N}_0}$ , with  $\mathbb{N}_0 = \{1, 2, \dots\}$ , be a sequence of mutually independent random  $d$ -vectors that are all distributed as the vector  $X$  subject to quantization, and let  $(\delta_t)_{t \in \mathbb{N}_0}$  be a deterministic sequence in  $(0, 1)$  such that  $\sum_{t=1}^\infty \delta_t = +\infty$  and  $\sum_{t=1}^\infty \delta_t^2 < +\infty$ . Starting from an initial  $N$ -grid  $X^0 = x^0$ , with pairwise distinct components, the CLVQ

algorithm recursively defines the grid  $X^t$ ,  $t \geq 1$ , as

$$X^t = X^{t-1} - \frac{\delta t}{2} \nabla_x d_N^2(X^{t-1}, \xi^t), \quad (2.5)$$

where  $\nabla_x d_N^2(x, \xi)$  stands for the gradient with respect to the  $x$ -argument of the so-called local quantization error  $d_N^2(x, \xi) = \min_{1 \leq i \leq N} |x_i - \xi|^2$ , with  $x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$  and  $\xi \in \mathbb{R}^d$ . For any  $\xi$ , the  $i$ th component of this gradient is

$$(\nabla_x d_N^2(x, \xi))_i = 2(x_i - \xi) \mathbb{I}_{[x_i = \text{Proj}_x(\xi)]}, \quad i = 1, \dots, N.$$

The grids  $X^{t-1}$  and  $X^t$  thus only differ by one point, namely the one corresponding to the non-zero component of the gradient above; see, e.g., [Pagès \(1998\)](#) or [Pagès and Printems \(2003\)](#) for details.

In the sequel, we will write throughout that the grids provided by the CLVQ algorithm are optimal. Strictly speaking, though, the grid  $\hat{\gamma}^{N,n}$  resulting from this algorithm, for any fixed  $n$ , does not minimize the quantization error, hence is not optimal. Yet it is shown in [Pagès \(1998\)](#) that, under mild assumptions,  $\hat{\gamma}^{N,n}$  converges to an optimal grid as the number of iterations (here, the sample size  $n$ ) tends to infinity (with  $N$  fixed). Clearly, only moderate-to-large values of  $n$  are expected to provide good approximations of genuine optimal  $N$ -grids.

We end this section by quickly evaluating the performances of the CLVQ algorithm, both for  $d = 1$  and  $d = 2$ . To do so, we generated, for the various combinations of  $n$  and  $N$  considered in [Table 1](#), 50 mutually independent random samples of size  $n$  from the uniform distribution over  $[-2, 2]^d$  and performed in each case the CLVQ algorithm with the corresponding value of  $N$ . For each  $(n, N)$ , this leads to 50 CPU times to completion and 50 sample quantization errors

$$e = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \text{Proj}_{\hat{\gamma}^{N,n}}(X_i))^2}. \quad (2.6)$$

[Table 1](#) reports the corresponding averaged CPU times, along with the averages and standard deviations of the various collections of 50 quantization errors. Since the number of iterations of the algorithm is equal to  $n$ , it is no surprise that the error decreases with  $n$  while the computation time increases with  $n$ .

	$d = 1$ and $N = 15$			$d = 2$ and $N = 30$		
	$n = 100$	$n = 1,000$	$n = 10,000$	$n = 200$	$n = 5,000$	$n = 20,000$
CPU	0.012	0.116	1.145	0.556	0.883	3.468
AVE[ $e$ ]	0.132	0.0993	0.0843	0.364	0.322	0.315
SD[ $e$ ]	0.0196	0.0115	0.00657	0.0315	0.0130	0.00835

Table 1: Averaged computation times (CPU, in seconds) dedicated to the CLVQ algorithm with grid size  $N$ , along with the averages (AVE[ $e$ ]) and standard deviations (SD[ $e$ ]) of the sample quantization errors in (2.6), obtained from 50 independent samples of size  $n$ ; see Section 2.2 for details.

### 2.3. Bootstrapped quantization-based estimators of conditional quantiles

As explained in CPS15, it is more appropriate in practice to use a bootstrap version of the estimator  $\hat{q}_\alpha^{N,n}(x)$ , particularly so for small to moderate values of  $n$ . Obtaining bootstrapped quantization-based estimators can be achieved in the following two steps :

- ( $\bar{S}1$ ) We first generate  $B(\in \mathbb{N}_0)$  samples of size  $n$  from the original sample with replacement. Performing (S1) on each of them leads to  $B$  optimal grids;
- ( $\bar{S}2$ ) each grid obtained in ( $\bar{S}1$ ) then yields, through (S2), an estimator  $\hat{q}_\alpha^{(b)}(x) = \hat{q}_\alpha^{(b),N,n}(x)$ ,  $b = 1, \dots, B$ , obtained from the projection of the original sample  $\{(X_i, Y_i), i = 1, \dots, n\}$  on the corresponding grid.

The resulting bootstrap estimator  $\bar{q}_{\alpha,B}^{N,n}(x)$  is

$$\bar{q}_{\alpha,B}^{N,n}(x) = \frac{1}{B} \sum_{b=1}^B \hat{q}_\alpha^{(b)}(x). \quad (2.7)$$

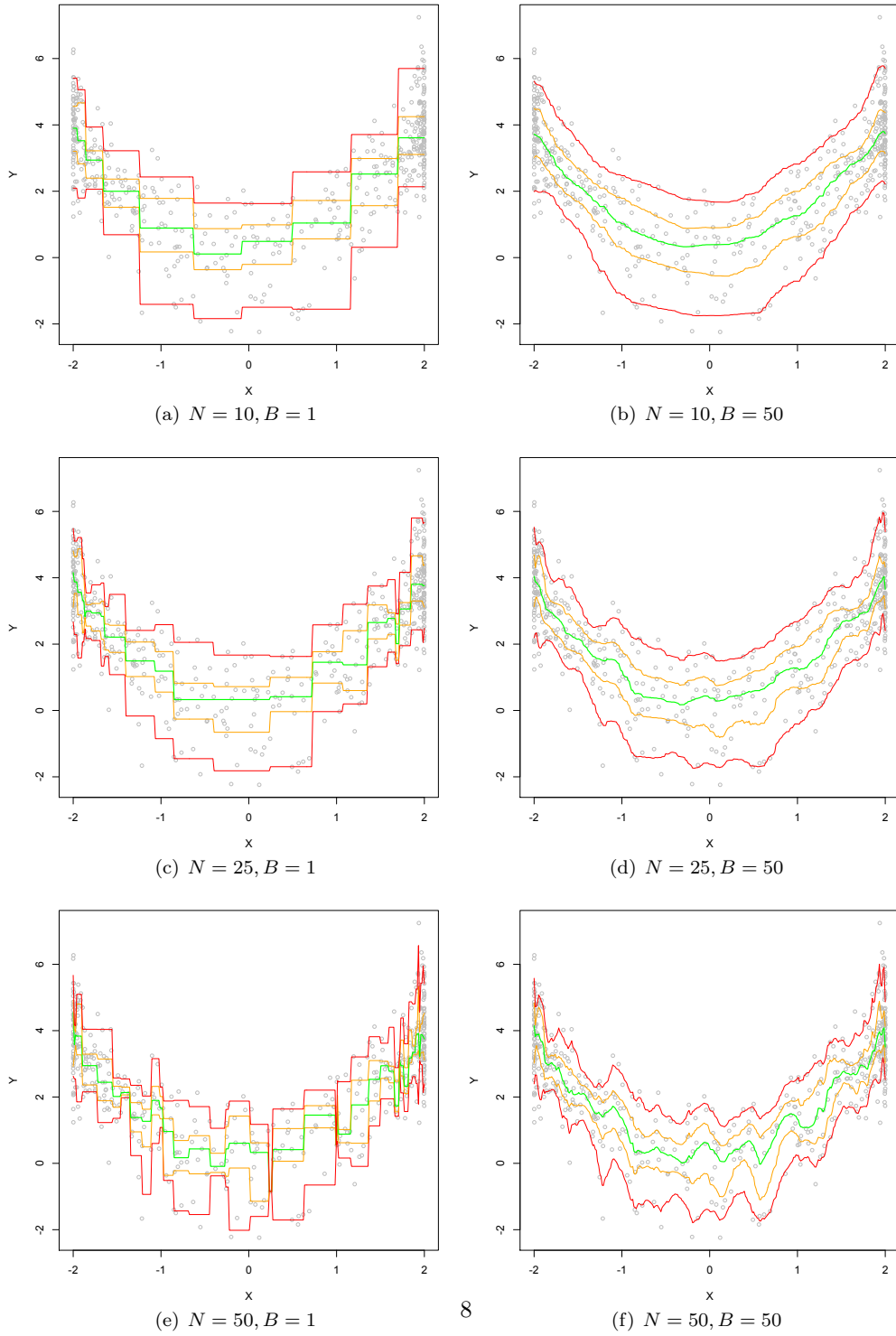
In the sequel, the notation  $B = 1$  means that no bootstrap is performed.

Figure 1 provides an illustration of the original estimator  $\hat{q}_\alpha^{N,n}(x)$  and its bootstrap version  $\bar{q}_{\alpha,B}^{N,n}(x)$  (with  $B = 50$ ). Based on a random sample of size  $n = 500$  from the model

$$Y = X^2 + \varepsilon, \quad (2.8)$$

where  $X = 4Z - 2$  (with  $Z \sim \text{Beta}(0.3, 0.3)$ ) and  $\varepsilon \sim \mathcal{N}(0, 1)$  are independent, we evaluated both  $\hat{q}_\alpha^{N,n}(x)$  and  $\bar{q}_{\alpha,B}^{N,n}(x)$  for  $N = 10, 25$ , and 50 (actually, these were evaluated only at 300 equispaced points between the minimum and maximum values of  $X_1, \dots, X_n$ ). Clearly, the bootstrap significantly smooths all curves. This happens provided that  $B$  is





8

Figure 1: Estimated conditional quantile curves  $x \mapsto \hat{q}_{\alpha}^{N,n}(x)$  (left) and their bootstrap counterparts  $x \mapsto \hat{q}_{\alpha,B}^{N,n}(x)$  for  $B = 50$  (right), based on  $N = 10$  (top),  $N = 25$  (middle), and  $N = 50$  (bottom). In all cases, the sample size is  $n = 500$ , and the quantile orders considered are  $\alpha = 0.05$  (red),  $0.25$  (orange),  $0.5$  (green),  $0.75$  (orange), and  $0.95$  (red); see (2.8) for the data generating model.

not too small. But of course,  $B$  should not be taken very large, as it would increase the computational burden (in the sequel, we use  $B = 50$  throughout).

Figure 1 also shows that the number  $N$  of quantizers has an important impact on the resulting quantile curves. For small  $N$  ( $N = 10$ ), the curves show a large bias and a small variability. For larger  $N$  ( $N = 50$ ), the bias is smaller but the variability is larger (with or without bootstrap). This bias-variance trade-off is standard in the selection of a smoothing parameter. The intermediate value  $N = 25$ , that achieves a nice balance between those two extreme cases, was actually obtained from an original  $N$ -selection method we describe in Section 3 below. This method to choose  $N$ , along with the computation of the (bootstrap or non-bootstrap) quantization-based estimators of conditional quantiles under investigation, are entirely implemented in a package R called `QuantifQuantile`, that is available on the CRAN.

### 3. Data-driven selection of $N$

In this section, we first define (Section 3.1) a natural *integrated square error (ISE)* quantity that is essentially convex in  $N$ , hence allows to identify an optimal value  $N_{\text{opt}}$  for  $N$ . ISEs involve the true unknown conditional quantiles, so that  $N_{\text{opt}}$  cannot be obtained from the data. We therefore propose (Section 3.2) a data-driven selection method for  $N$  obtained by replacing the true ISEs with bootstrap versions. We then investigate how close the resulting  $\hat{N}_{\text{opt}}$  is to  $N_{\text{opt}}$ .

#### 3.1. Infeasible selection of $N$

For any fixed  $N$ , we can consider the ISEs (recall that  $S_X$  denotes the support of  $X$ )

$$\int_{S_X} (\hat{q}_\alpha^{N,n}(x) - q_\alpha(x))^2 dx \quad \text{and} \quad \int_{S_X} (\bar{q}_{\alpha,B}^{N,n}(x) - q_\alpha(x))^2 dx,$$

associated with the non-bootstrap and bootstrap estimators, respectively. These can be approximated by

$$\text{ISE}_{\alpha,J}^{\hat{}}(N) = \frac{1}{J} \sum_{j=1}^J (\hat{q}_\alpha^{N,n}(x_j) - q_\alpha(x_j))^2 \quad \text{and} \quad \text{ISE}_{\alpha,B,J}^{\bar{}}(N) = \frac{1}{J} \sum_{j=1}^J (\bar{q}_{\alpha,B}^{N,n}(x_j) - q_\alpha(x_j))^2,$$

where  $x_1, \dots, x_J$  are equispaced between the minimum and maximum values of  $X_1, \dots, X_n$ .

It is of course natural to consider optimal a value of  $N$  that minimizes these ISEs, which

leads to considering

$$N_{\alpha,J;\text{opt}}^{\wedge} = \arg \min_{N \in \mathbb{N}_0} \text{ISE}_{\alpha,J}^{\wedge}(N) \quad \text{and} \quad N_{\alpha,B,J;\text{opt}}^{-} = \arg \min_{N \in \mathbb{N}_0} \text{ISE}_{\alpha,B,J}^{-}(N).$$

These optimal  $N$ -values may depend on  $\alpha$ , which explains the notation. The dependence on  $J$  of these ISEs and of optimal  $N$ -values is way less crucial than their dependence on  $\alpha$  and  $B$ ; accordingly, we simply write  $\text{ISE}_{\alpha}^{\wedge}(N)$ ,  $\text{ISE}_{\alpha,B}^{-}(N)$ ,  $N_{\alpha;\text{opt}}^{\wedge}$  and  $N_{\alpha,B;\text{opt}}^{-}$  in the sequel.

To illustrate these definitions, we generated mutually independent random samples of size  $n = 300$  according to the models

$$(\mathcal{M1}) \quad Y = \frac{1}{5}X_1^3 + \varepsilon,$$

$$(\mathcal{M2}) \quad Y = f(X_2) + \varepsilon',$$

where  $X_1 = 6Z_1 - 3$  (with  $Z_1 \sim \text{Beta}(0.3, 0.3)$ ),  $X_2 = 3Z_2 - 1.5$  (with  $Z_2 \sim \text{Beta}(2, 2)$ ),  $\varepsilon \sim \mathcal{N}(0, 1)$ , and  $\varepsilon' \sim \chi_2^2$  are mutually independent. Denoting the standard normal density as  $\varphi$ , the link function  $f$  is defined as  $f(x) = \frac{1}{2}\varphi(x) + 10 \sum_{\ell=1}^3 (0.2)^{\ell} \varphi(10(x - \frac{\ell}{2} + 1))$ , a choice that is inspired by [Marron and Wand \(1992\)](#). Obviously,  $q_{\alpha}(x) = \frac{x^3}{5} + \Phi^{-1}(\alpha)$  for  $(\mathcal{M1})$  and  $q_{\alpha}(x) = f(x) + \Psi_2^{-1}(\alpha)$  for  $(\mathcal{M2})$ , where  $\Phi$  and  $\Psi_2$  denote the cumulative distribution functions of the  $\mathcal{N}(0, 1)$  distribution and of the  $\chi_2^2$  distribution, respectively. We evaluated the ISEs above with  $J = 300$  points. [Figure 2](#) focuses on Model  $(\mathcal{M1})$ . Its top panels plot the graphs of  $N \mapsto \text{ISE}_{\alpha}^{\wedge}(N)$  and  $N \mapsto \text{ISE}_{\alpha,B}^{-}(N)$  (with  $B = 50$ ) for a single sample, while the bottom panels plot the corresponding graphs averaged (pointwise in  $x$ ) over  $m = 500$  mutually independent samples. For the sake of brevity, [Figure 3](#), that is related to Model  $(\mathcal{M2})$ , only reports the corresponding plots for bootstrapped estimators (still with  $B = 50$ ).

[Figures 2-3](#) show that ISE curves are indeed essentially convex in  $N$ : while random variability prevents perfect convexity for a single sample, the smoothest case based on 500 replications with  $\bar{q}_{\alpha,B}^{N,n}$  leads to strictly convex I(M)SE curves. Besides, these curves are quite flat in a neighborhood of  $N_{\alpha;\text{opt}}^{\wedge}$  or  $N_{\alpha,B;\text{opt}}^{-}$ , so that picking a value of  $N$  that is close to (but not exactly equal to) the optimal  $N$ -value will not have an important impact on the resulting ISE values. It seems that, as expected,  $N_{1-\alpha;\text{opt}}^{\wedge} = N_{\alpha;\text{opt}}^{\wedge}$  and  $N_{1-\alpha,B;\text{opt}}^{-} = N_{\alpha,B;\text{opt}}^{-}$  when the error  $\varepsilon$  has a symmetric distribution. In contrast, asymmetric errors ([Figure 3](#)) lead to optimal values of  $N$  that are not symmetric in  $\alpha$

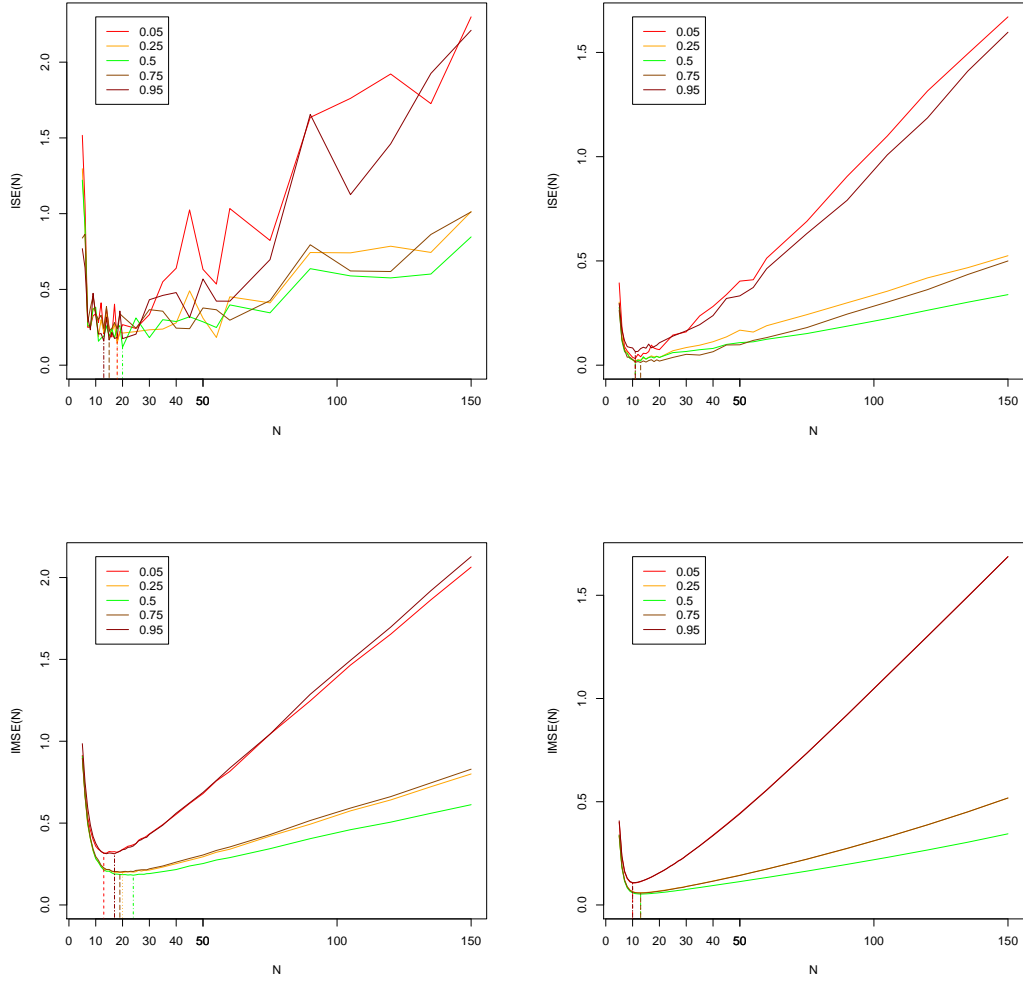


Figure 2: Plots of the mappings  $N \mapsto ISE_{\alpha}(N)$  (top left) and  $N \mapsto ISE_{\alpha,B}^{-}(N)$  with  $B = 50$  (top right) for a random sample of size  $n = 300$  from Model ( $\mathcal{M}1$ ). The bottom panels report the corresponding plots obtained by averaging these mappings over  $m = 500$  mutually independent random samples (all mappings are actually only evaluated at  $N = 5, 6, 7, \dots, 29, 30, 35, 40, \dots, 150$ ).

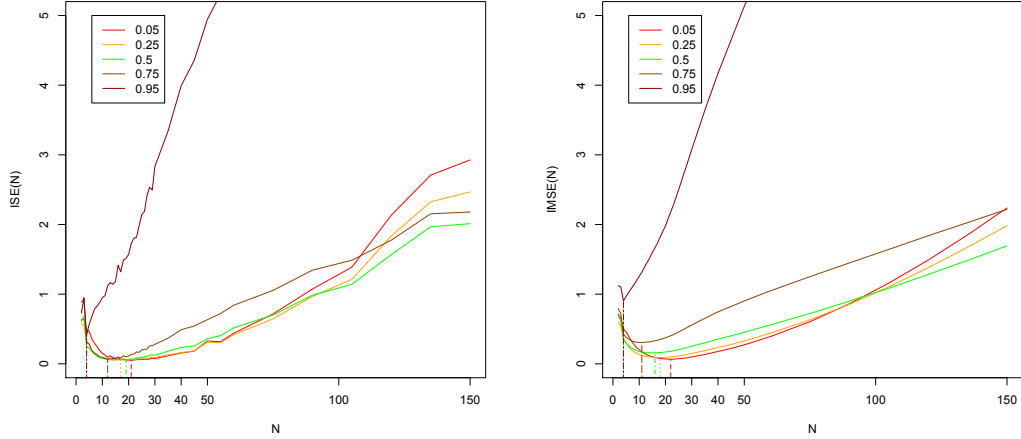


Figure 3: The same plots as in the right panels of Figure 2, but for Model ( $\mathcal{M}2$ ).

and depend more strongly of  $\alpha$ .

To conclude this section, we stress that the existence of such an optimal value of  $N$  does not contradict the theoretical results from Section 2. One might indeed have guessed from (2.3) that the ISE functions above would be monotone decreasing (rather than convex) in  $N$ . The result in (2.3), however, (i) involves the population quantile approximation  $\tilde{q}_\alpha^N(x)$  and not its sample counterpart  $\hat{q}_\alpha^{N,n}(x)$ , and (ii) requires that the projection, for any  $N$ , is performed on an optimal quantization grid. As pointed out earlier, the CLVQ algorithm provides a good approximation of an optimal grid only if the number of iterations, that is equal to the sample size  $n$ , is large compared to  $N$ . Consequently, in the present setup where the sample size  $n$  is fixed, increasingly large values of  $N$  will result into CLVQ grids that are less and less optimal, which explains the increasing ISE values for such values of  $N$ .

### 3.2. Data-driven selection of $N$

In practice, the population conditional quantile function  $q_\alpha(\cdot)$  is unknown, so that it is impossible to obtain optimal  $N$ -values by minimizing  $\text{ISE}_\alpha^{\hat{}}(N)$  and  $\text{ISE}_{\alpha,B}^{\bar{}}(N)$  as above. We then propose the following approach that consists in replacing the unknown

quantile  $q_\alpha(x)$  by a bootstrap estimate (see [Efron \(1982\)](#) for more details on such methods).

Consider a grid  $\{x_1 = X_{(1)}, x_2, \dots, x_{J-1}, x_J = X_{(n)}\}$  of equispaced points between the minimal and maximal observed  $X$ -values. We generate  $\tilde{B}$  bootstrap samples of size  $n$  from  $X_1, \dots, X_n$  and use them as stimuli (that is, as the  $\xi^t$ 's) in the CLVQ algorithm, which provides  $\tilde{B}$  resulting quantization grids. We then consider

$$\widehat{\text{ISE}}_{\alpha, \tilde{B}, J}(N) = \frac{1}{J} \sum_{j=1}^J \left( \frac{1}{\tilde{B}} \sum_{\tilde{b}=1}^{\tilde{B}} (\hat{q}_\alpha^{N,n}(x_j) - \hat{q}_\alpha^{(\tilde{b})}(x_j))^2 \right)$$

and

$$\widehat{\text{ISE}}_{\alpha, B, \tilde{B}, J}^-(N) = \frac{1}{J} \sum_{j=1}^J \left( \frac{1}{\tilde{B}} \sum_{\tilde{b}=1}^{\tilde{B}} (\bar{q}_{\alpha, B}^{N,n}(x_j) - \hat{q}_\alpha^{(\tilde{b})}(x_j))^2 \right), \quad (3.1)$$

where  $\hat{q}_\alpha^{(\tilde{b})}(x_j)$ ,  $\tilde{b} = 1, \dots, \tilde{B}$ , stands for the estimator from (S2) in [Page 5](#) computed from the original sample  $(X_1, Y_1), \dots, (X_n, Y_n)$  and the  $\tilde{b}$ th grid obtained above. Evaluating  $\widehat{\text{ISE}}_{\alpha, B, \tilde{B}, J}^-(N)$  thus requires generating  $B + \tilde{B}$  bootstrap samples of size  $n$ :  $B$  for the construction of  $\bar{q}_{\alpha, B}^{N,n}(x_j)$ , and  $\tilde{B}$  to obtain  $\hat{q}_\alpha^{(\tilde{b})}(x_j)$ ,  $\tilde{b} = 1, \dots, \tilde{B}$ . These sample ISEs are to be minimized in  $N$ . Since not all values of  $N$  can be considered in practice, we rather consider

$$\hat{N}_{\alpha, \tilde{B}, J; \text{opt}} = \arg \min_{N \in \mathcal{N}} \widehat{\text{ISE}}_{\alpha, \tilde{B}, J}(N) \quad \text{and} \quad \hat{N}_{\alpha, B, \tilde{B}, J; \text{opt}}^- = \arg \min_{N \in \mathcal{N}} \widehat{\text{ISE}}_{\alpha, B, \tilde{B}, J}^-(N),$$

where the cardinality of  $\mathcal{N}(\subset \mathbb{N}_0)$  is finite (and may be chosen as a function of  $n$ ). As for their infeasible counterparts, we will not stress in the sequel the dependence on  $J$  in these sample ISEs and optimal  $N$ -values, nor the dependence on  $\tilde{B}$ , that we choose equal to 30 throughout.

[Figure 4](#) plots the mappings  $N \mapsto \widehat{\text{ISE}}_\alpha(N)$  and  $N \mapsto \widehat{\text{ISE}}_{\alpha, B}^-(N)$  and, for the sake of comparison, the (infeasible) mappings  $N \mapsto \widehat{\text{ISE}}_\alpha(N)$  and  $N \mapsto \widehat{\text{ISE}}_{\alpha, B}^-(N)$ , in the setup of Model ( $\mathcal{M}1$ ) with sample size  $n = 300$  (more precisely, the average of the corresponding plots, over 500 mutually independent random samples, are showed there). It is seen that, for large  $N$ , the (averaged) sample ISE functions quite poorly estimate their theoretical versions. Our primary interest, however, rather is in the agreement of the corresponding argmins, which are associated with the feasible or infeasible optimal  $N$ -values. In that respect, the results show that the values of  $N$  minimizing the sample ISE functions tend

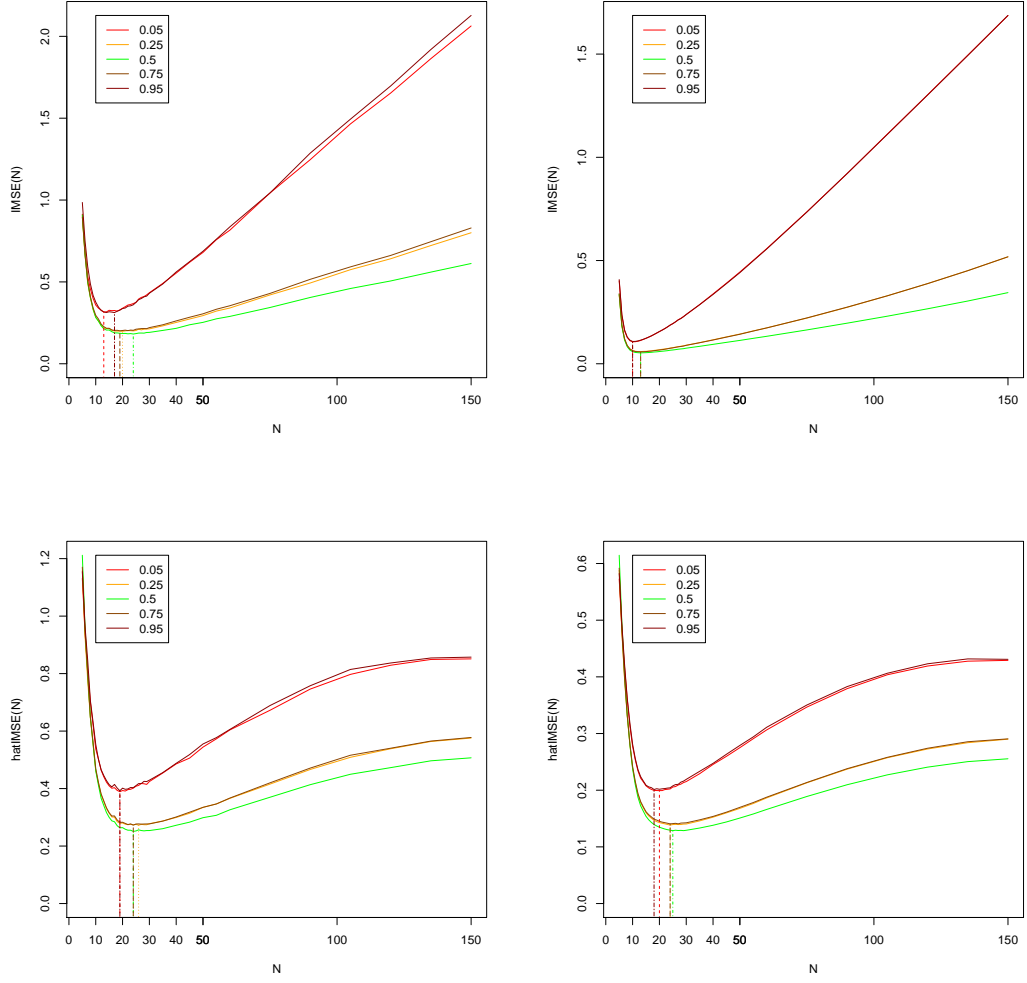


Figure 4: Plots of the mappings  $N \mapsto \text{ISE}_{\alpha}^{\wedge}(N)$  (top left),  $N \mapsto \text{ISE}_{\alpha, B}^{-}(N)$  with  $B = 50$  (top right), and of the sample mappings  $N \mapsto \widehat{\text{ISE}}_{\alpha}^{\wedge}(N)$  with  $\tilde{B} = 30$  (bottom left),  $N \mapsto \widehat{\text{ISE}}_{\alpha, B}^{-}(N)$  with  $\tilde{B} = 30$  and  $B = 50$  (bottom right), averaged over 500 mutually independent random samples from Model ( $\mathcal{M}1$ ) with sample size  $n = 300$ .

to over-estimate the optimal  $N$ -values, with a bias of about 10. Note that this bias is not alarming since Figure 2 shows that virtually the same performances in terms of ISE are obtained with or without bias (particularly so for intermediate values of  $\alpha$ ). Finally, we point out that, when considering a single sample (instead of 500 replications), the estimated optimal  $N$ -value stays in the same interval (the plots are not provided here).

To sum up, the above simulation results suggest that the sample ISE functions allow us to obtain a data-driven value of  $N$  that, for all practical purposes, is as satisfactory as the infeasible optimal  $N$ -value based on the original ISE functions. Note that, in a nonparametric framework, the smoothing parameter is often rather selected by minimizing the mean square error of the estimator at hand. In the context of this paper, however, deriving explicit expressions of the bias and variance (or standard errors) of quantization-based quantile estimators turns out to be extremely challenging. The main reason is that the randomness of observations appears twice in the definition of these estimators : (i) when obtaining the quantization grid (through the CLVQ algorithm, on which few theoretical results are available) and (ii) when calculating the empirical quantile of the selected observations. This explains why we need to rely on the selection procedure described in this section.

#### 4. Comparison with some classical competitors

The aim of this section is to investigate how the proposed estimator of conditional quantile behaves compared to its main competitors, that are introduced in Section 4.1. We both compare the estimated quantile curves themselves (Section 4.2) and the respective ISEs, averaged over  $m = 500$  mutually independent random samples (Section 4.3).

##### 4.1. The competitors considered

The first competing estimator we consider is associated with quantile regression *spline methods*; see [Koenker et al. \(1994\)](#). More precisely, this estimator,  $\hat{q}_\alpha^{\text{sp}}$ , say, is defined as

$$\hat{q}_\alpha^{\text{sp}} = \hat{q}_{\alpha,\lambda}^{\text{sp}} = \arg \min_{g \in \mathcal{G}} \left\{ \sum_{i=1}^n \rho_\alpha(Y_i - g(X_i)) + \lambda P(g) \right\},$$

where  $\lambda$  is a nonnegative real number,  $\mathcal{G}$  is an appropriately chosen space of functions, and the penalty  $P(g) = \int_0^1 |g''(x)| dx$  is the total variation of the first derivative of  $g$ .



It is well-known that  $\widehat{q}_\alpha^{\text{SP}}$  is always piecewise linear. As usual in penalization methods,  $\lambda$  governs the trade-off between fidelity to the sample and smoothness of the resulting estimator. Several methods to select an appropriate value of  $\lambda$  are available in the literature. Below, we will use the AIC criterion, which selects the value  $\lambda = \widehat{\lambda}_{\alpha;\text{opt}}$  minimizing

$$\text{AIC}(\lambda) = \log \left[ \frac{1}{n} \sum_i \rho_\alpha(Y_i - \widehat{g}_\lambda(X_i)) \right] + \frac{p_\lambda}{n},$$

where  $p_\lambda$  is the number of knots defining the spline. The reader can refer to [Koenker et al. \(1994\)](#) and [Koenker and Mizera \(2004\)](#) for more details on how to choose  $\lambda$  (or  $\mathcal{G}$ ). When implementing this method, we used the `rqss` function from the R package `quantreg` to compute  $\widehat{q}_\alpha^{\text{SP}}$  and performed the  $\lambda$ -selection via the R function `AIC` (the package `quantreg` does not propose an automatic  $\lambda$ -selection procedure).

Another competitor is the *kNN estimator*, which is defined as follows; see [Bhattacharya and Gangopadhyay \(1990\)](#). For any integer  $k \in \{1, \dots, n\}$ , this estimator,  $\widehat{q}_{\alpha;k\text{NN}}^k(x)$  say, is defined as the  $\alpha$ -quantile of  $Y_1, \dots, Y_k$  after permuting the observations  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  in such a way that  $|X_i - x|$  is increasing in  $i$ . In other words,  $\widehat{q}_{\alpha;k\text{NN}}^k(x)$  is the  $\alpha$ -quantile of the  $Y_i$ 's associated with the  $k$   $X_i$ 's that are closest to  $x$ . There does not seem to exist an efficient and well-accepted data-driven method to choose  $k$  in the literature. In the sequel, we throughout considered the *kNN estimator*

$$\widehat{q}_{\alpha;k\text{NN}}(x) = \widehat{q}_{\alpha;k\text{NN}}^{k_{\alpha;\text{opt}}}(x),$$

with

$$k_{\alpha;\text{opt}} = \arg \min_{k \in \mathcal{K}} \frac{1}{J} \sum_{j=1}^J (\widehat{q}_{\alpha;k\text{NN}}^k(x_j) - q_\alpha(x_j))^2,$$

where  $\{x_1, \dots, x_J\}$  is chosen as in Section 3.2 and  $\mathcal{K}$  denotes some set of possible values for  $k$ . Clearly,  $k$  cannot be chosen this way in practice since the theoretical quantiles  $q_\alpha(x_j)$  are unknown. To the best of our knowledge, no R package allows to compute  $\widehat{q}_{\alpha;k\text{NN}}(x)$  and we therefore wrote our own implementation to conduct the simulations below.

The last estimators we consider are the *kernel (local linear or local constant) estimators* introduced in [Yu and Jones \(1998\)](#). The local linear estimator is of the form  $\widehat{q}_{\alpha;YJ}(x) =$

$\hat{a}$ , with

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b) \in \mathbb{R} \times \mathbb{R}} \sum_{i=1}^n \rho_\alpha(Y_i - a - b(X_i - x)) K\left(\frac{X_i - x}{h}\right),$$

where  $K$  is a kernel function and  $h$  is the bandwidth. In the sequel,  $K$  will be the standard normal density, and we choose

$$h = \hat{h}_{\alpha;\text{opt}} = \frac{\alpha(1-\alpha)}{(\varphi(\Phi^{-1}(\alpha)))^2} h_{\text{mean}},$$

where  $\varphi$  and  $\Phi$  are respectively the standard normal density and distribution functions, and where  $h_{\text{mean}}$  is the optimal choice of  $h$  for mean regression, selected through cross-validation; see [Yu and Jones \(1998\)](#). The local constant version of this estimator is defined as  $\hat{q}_{\alpha;\text{YJc}}(x) = \hat{a}$ , with

$$\hat{a} = \arg \min_{a \in \mathbb{R}} \sum_{i=1}^n \rho_\alpha(Y_i - a) K\left(\frac{X_i - x}{h}\right),$$

where  $K$  and  $h$  will throughout be chosen as for the local linear estimator. As for  $\hat{q}_{\alpha;k\text{NN}}(x)$ , the simulations below are based on our own R implementation of  $\hat{q}_{\alpha}^{\text{YJ}}(x)$  and  $\hat{q}_{\alpha}^{\text{YJc}}(x)$ .

#### 4.2. Comparison of estimated quantile curves

We now compare our quantization-based estimators with their competitors described above. Since we saw in [Section 2.3](#) that the bootstrap estimators  $\bar{q}_{\alpha,B}^{N,n}$  are to be favored over their original versions  $\hat{q}_{\alpha}^{N,n}$ , we restrict to  $\bar{q}_{\alpha,B}^{N,n}$  below, with the corresponding data-driven value of  $N$ , namely  $\hat{N}_{\alpha,B;\text{opt}}^-$ , that was proposed in [Section 3.2](#). In this section, we also do not consider the local constant estimator  $\hat{q}_{\alpha}^{\text{YJc}}(x)$  since the results in [Section 4.3](#) below show that it is usually outperformed by its local linear version in terms of ISEs. We start the comparison by investigating estimated quantile curves computed from  $n$  independent observations generated according to the models

$$(\mathcal{M}1) \quad Y = \frac{1}{5}X_1^3 + \varepsilon,$$

$$(\mathcal{M}2) \quad Y = f(X_2) + \varepsilon',$$

$$(\mathcal{M}3) \quad Y = \sin(X_3) + \left(0.5 + 1.5 \sin^2\left(\frac{\pi}{2}X_3\right)\right)\varepsilon,$$

where  $X_1 = 6Z_1 - 3$  (with  $Z_1 \sim \text{Beta}(0.3, 0.3)$ ),  $X_2 = 3Z_2 - 1.5$  (with  $Z_2 \sim \text{Beta}(2, 2)$ ),  $X_3 = 6Z_2 - 3$ ,  $\varepsilon \sim \mathcal{N}(0, 1)$ , and  $\varepsilon' \sim \chi_2^2$  are mutually independent. The link function  $f$  is the same as in [Section 3.1](#). For  $n = 300$ , the resulting estimated quantile curves are

plotted in Figures 5-7 for Models ( $\mathcal{M}1$ )-( $\mathcal{M}3$ ), respectively. For the sake of comparison, Figure 8 provides the results for Model ( $\mathcal{M}1$ ) and  $n = 1,000$ .

The quantization-based quantile curves are smooth and adapt well to the polynomial or more complex nature of the link function at hand. In contrast, while the piecewise linear curves obtained from the spline estimator are beneficial for polynomial link functions (see Figure 5), they hurt for more complex link functions (Figure 6, e.g., shows that the spline-based curves miss some of the underlying bumps). The curves resulting from  $k$ NN estimation are relatively close to the theoretical ones in each model, but show some peaks or constant pieces, hence are less pleasant from a visual point of view. Eventually, the curves associated to  $\hat{q}_\alpha^{YJ}(x)$  lack smoothness for medium values of the covariate in Figure 5 (further numerical experiments, not reported here, reveal that this follows from the non-uniform covariate distribution). Also, the local linear estimator  $\hat{q}_\alpha^{YJ}(x)$  does not catch correctly the link function in Figure 7.

The values of the various tuning parameters are given in Table 2. For the sake of brevity, we only comment on the results for Model ( $\mathcal{M}2$ ). We see that the optimal value for  $N$  decreases with  $\alpha$ , which is due to the chi-square error term, that implies that estimation of conditional  $\alpha$ -quantiles should be based on increasingly many data points as  $\alpha$  grows. In accordance with this, the optimal number of neighbours  $k_{\alpha;\text{opt}}$  in  $k$ NN estimation increases with  $\alpha$ . Such a monotonic pattern is not observed for the bandwidth  $\hat{h}_{\alpha;\text{opt}}$  used in kernel estimators since it is chosen such a way that  $\hat{h}_{\alpha;\text{opt}} = \hat{h}_{1-\alpha;\text{opt}}$ . Finally, the parameter  $\hat{\lambda}_{\alpha;\text{opt}}$ , that is a monotone decreasing function of the number of knots of the splines, is related to  $\alpha$  in a quite unclear fashion.

#### 4.3. Comparison of the ISEs

Obtaining well-behaved curves is of course desirable, particularly so in applied statistics, but this should not be achieved at the expense of efficiency. That is why we now compare the various estimators in terms of ISEs. To do so, we generated 500 independent samples from Models ( $\mathcal{M}1$ )-( $\mathcal{M}3$ ) with sample sizes  $n = 300$  and  $n = 1,000$ . In each case, we evaluated the ISEs corresponding to the quantization-based estimators and to their four competitors defined in Section 4.1. For each model, sample size, and quantile order  $\alpha$  considered, this provides a series of 500 observed ISEs for each estimator. Figures 9-11 draw the boxplots of those 500 ISE values for Models ( $\mathcal{M}1$ )-( $\mathcal{M}3$ ), respectively.

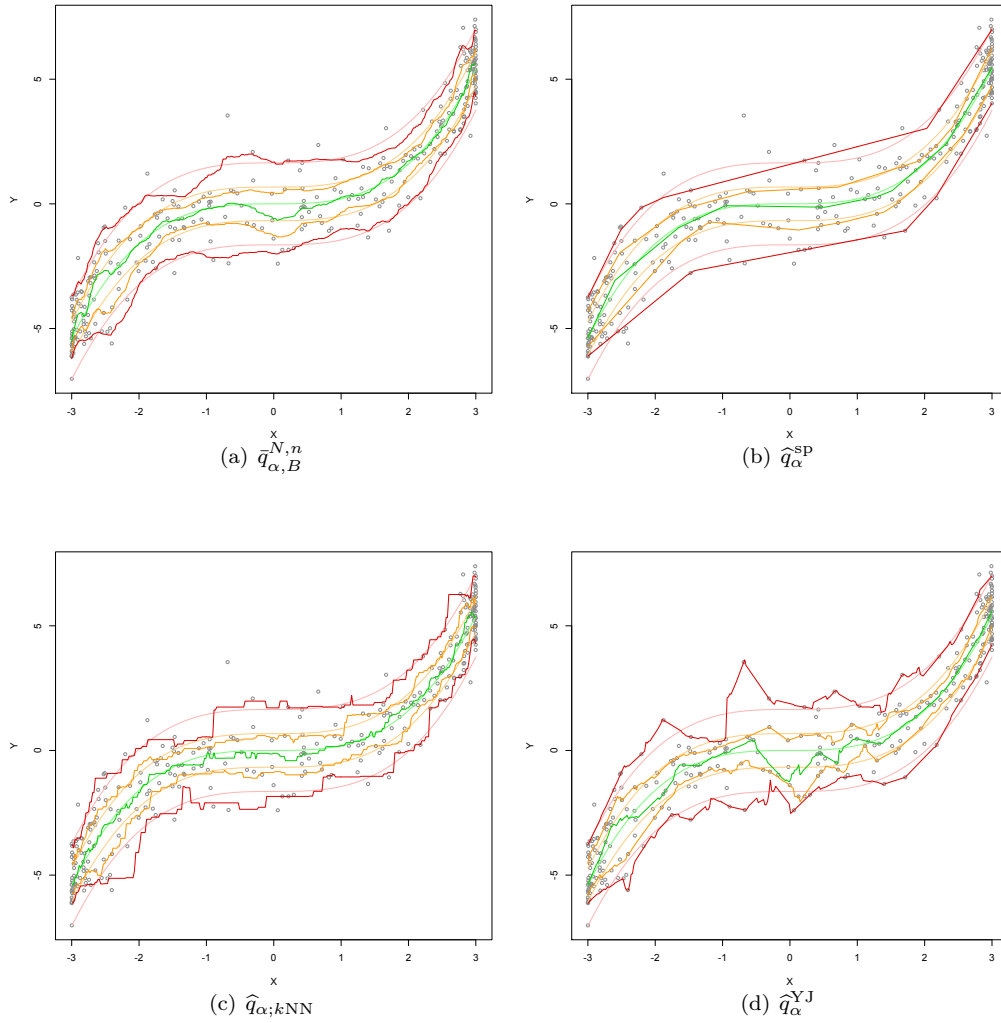


Figure 5: For a random sample of size  $n = 300$  generated according to Model  $(\mathcal{M}1)$ , conditional quantile curves obtained from quantization-based estimation (upper left), spline methods (upper right), nearest-neighbour estimation (lower left), and local linear kernel methods (lower right). In all cases, the quantile levels considered are  $\alpha = 0.05$  (red),  $0.25$  (orange),  $0.5$  (green),  $0.75$  (orange), and  $0.95$  (red). The lighter curves correspond to population conditional quantiles.

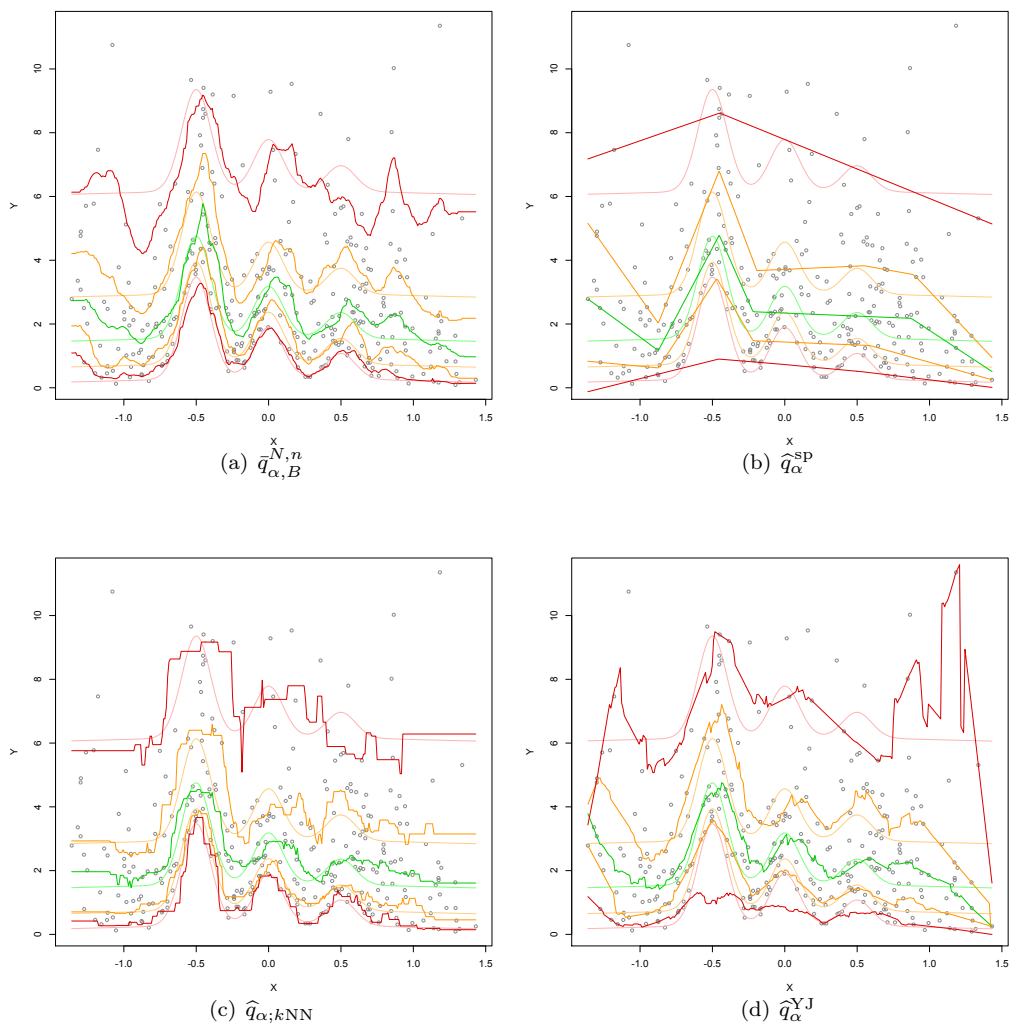


Figure 6: The same plots as in Figure 5, but for Model ( $\mathcal{M}2$ ) and  $n = 300$ .

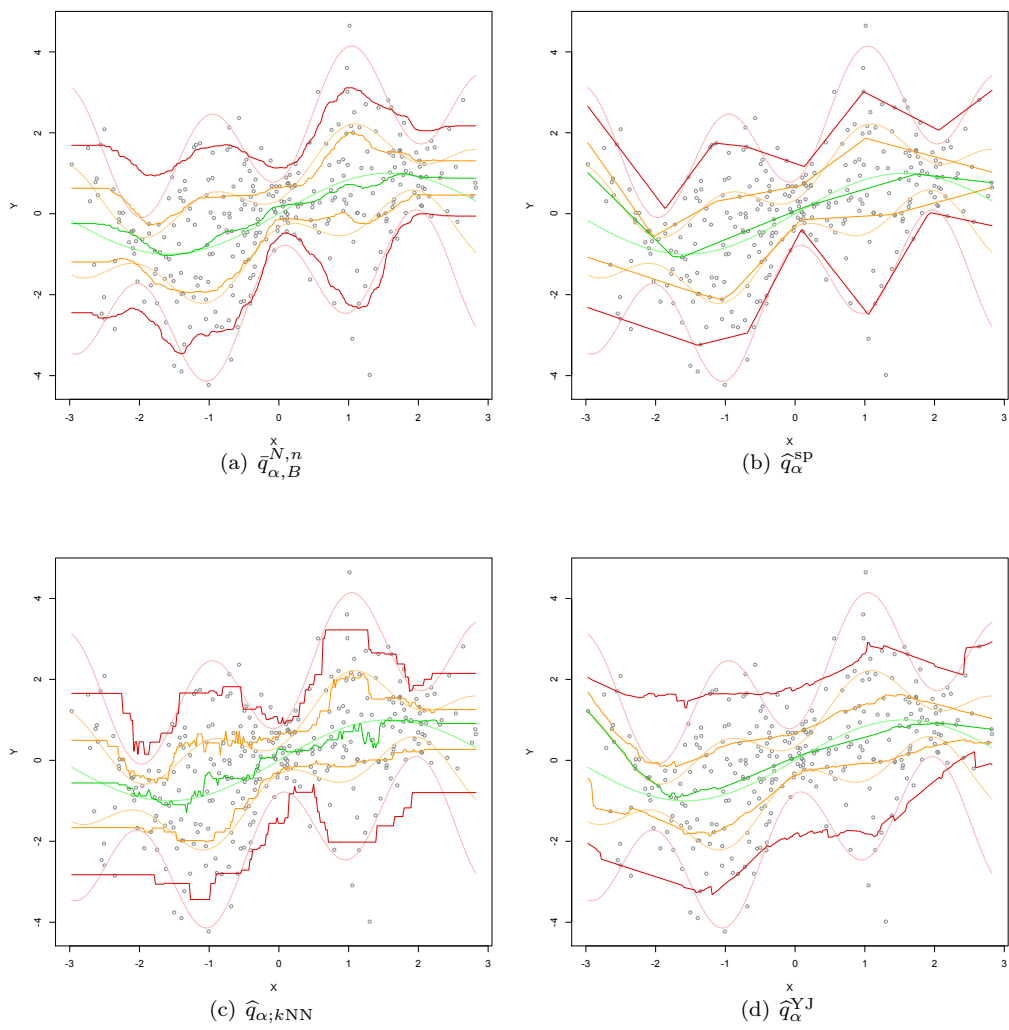


Figure 7: The same plots as in Figures 5-6, but for Model ( $\mathcal{M}3$ ) and  $n = 300$ .

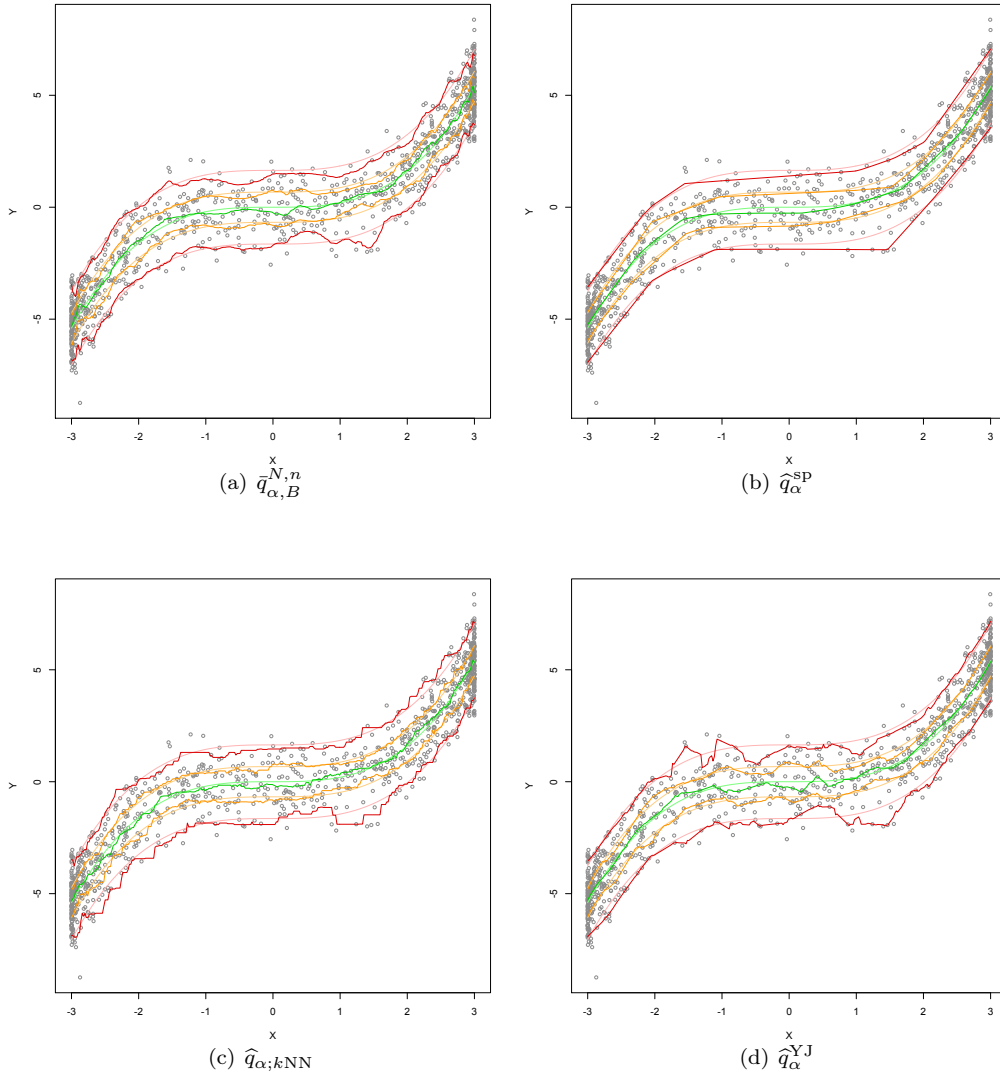


Figure 8: The same plots as in Figures 5-7, but for Model ( $\mathcal{M}1$ ) and  $n = 1,000$ .

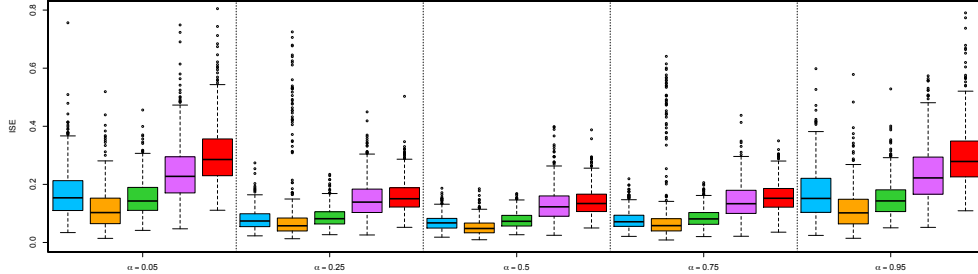
		$\alpha = 0.05$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.95$
$n = 300$	$\hat{N}_{\alpha,B;\text{opt}}$	20	23	23	18	18
	$\hat{\lambda}_{\alpha;\text{opt}}$	0.595	0.971	1.075	0.638	1.033
	$k_{\alpha;\text{opt}}$	25	32	30	28	27
	$\hat{h}_{\alpha;\text{opt}}$	0.202	0.170	0.164	0.170	0.202
$n = 300$	$\hat{N}_{\alpha,B;\text{opt}}$	23	23	19	16	15
	$\hat{\lambda}_{\alpha;\text{opt}}$	0.556	0.672	0.632	0.504	0.716
	$k_{\alpha;\text{opt}}$	16	38	42	43	55
	$\hat{h}_{\alpha;\text{opt}}$	0.135	0.113	0.109	0.113	0.135
$n = 300$	$\hat{N}_{\alpha,B;\text{opt}}$	13	10	9	10	9
	$\hat{\lambda}_{\alpha;\text{opt}}$	0.518	1.127	1.490	0.840	0.516
	$k_{\alpha;\text{opt}}$	70	61	38	28	34
	$\hat{h}_{\alpha;\text{opt}}$	0.594	0.498	0.482	0.498	0.594
$n = 1,000$	$\hat{N}_{\alpha,B;\text{opt}}$	32	43	43	41	33
	$\hat{\lambda}_{\alpha;\text{opt}}$	1.122	1.117	0.927	1.120	1.208
	$k_{\alpha;\text{opt}}$	72	54	84	74	94
	$\hat{h}_{\alpha;\text{opt}}$	0.148	0.124	0.120	0.124	0.148

Table 2: Values of the various tuning parameters involved in the conditional quantile estimators considered in Figures 5-8.

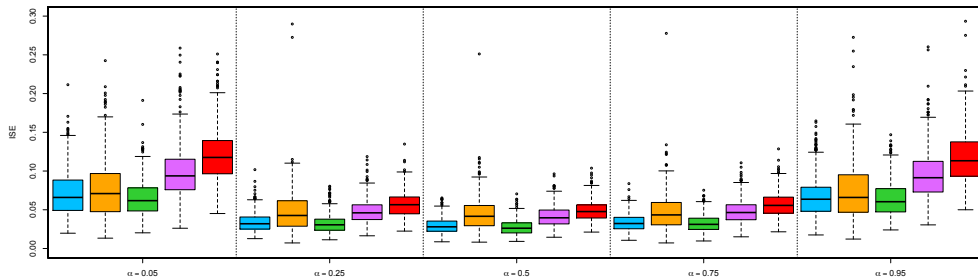
ISEs are, as expected, smaller for  $n = 1,000$  than for  $n = 300$ , but otherwise are quite similar.

For most models and quantile orders considered, the quantization-based estimators, spline estimators, and  $k$ NN estimators provide smaller ISEs than the local constant and local linear estimators. Since the  $k$ NN estimators are based on a selection of  $k$  that is infeasible, the main competitors to our quantization-based estimators are those based on splines. Results reveal that polynomial link functions (see Figure 9) usually are more favourable to splines-based methods (Figure 9(b) provides an exception, though, which shows that the sample size may also play a role). On the contrary, quantization-based estimators are better for more complex link function (see Figures 10-11), which is in line with the comparison of the quantization-based and spline-based estimated curves in





(a)  $n = 300$

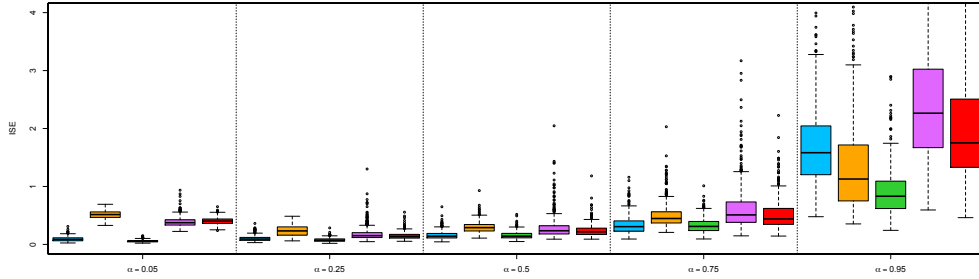


(b)  $n = 1,000$

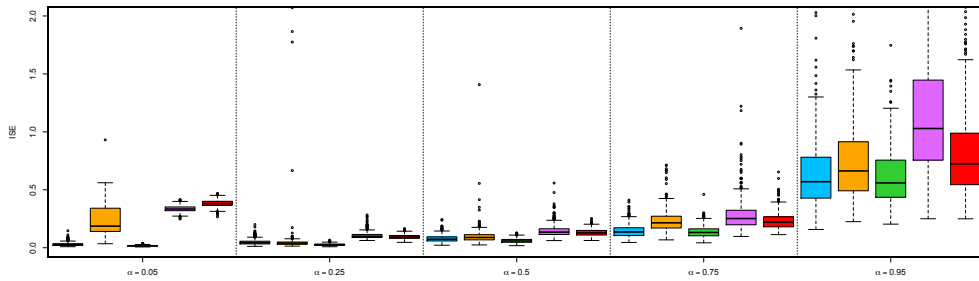
Figure 9: Boxplots, for  $\alpha = 0.05, 0.25, 0.50, 0.75$ , and  $0.95$ , of the ISEs of various conditional  $\alpha$ -quantile estimators obtained from 500 independent random samples from Model  $(\mathcal{M}1)$ , with size  $n = 300$  (top) and  $n = 1,000$  (bottom). The estimators considered are the quantization-based estimator  $\hat{q}_{\alpha,B}^{N,n}$  (in blue), the spline estimator  $\hat{q}_{\alpha}^{\text{SP}}$  (in orange), the  $k$ NN estimator estimator  $\hat{q}_{\alpha;k\text{NN}}$  (in green), the local linear estimator  $\hat{q}_{\alpha}^{\text{YJ}}$  (in purple), and the local constant estimator  $\hat{q}_{\alpha}^{\text{YJc}}$  (in red).

Section 4.2 (note, however, that an exception also appears in Figure 11(b)).

Since the computational burden is also an important issue, we gather in Table 3 the computation times (in seconds) used by each estimator to produce Figure 9(a). To study computation times in the light of efficiency, we further report there, for each estimator, a global measure of efficiency (Eff), defined as the sum, over the five  $\alpha$ -values considered in that figure, of the medians of the 500 fixed- $\alpha$  observed ISEs. Since the choice of the grid  $\mathcal{N}$  plays an important role in the selection of the optimal  $N$ -value (at least from a computational point of view), we considered two grids, namely  $\mathcal{N}_1 = \{5, 6, \dots, 29, 30\}$  and  $\mathcal{N}_5 = \{5, 10, \dots, 25, 30\}$  (the latter one, that may seem too coarse, is actually the one that led to the good ISE performances in Figures 9). Table 3 confirms that there is no free lunch, as it shows that the gain in terms of efficiency (for Model  $(\mathcal{M}1)$  and  $n = 300$ ) has a price in terms of computation time. This price, however, is quite reasonable. Moreover,

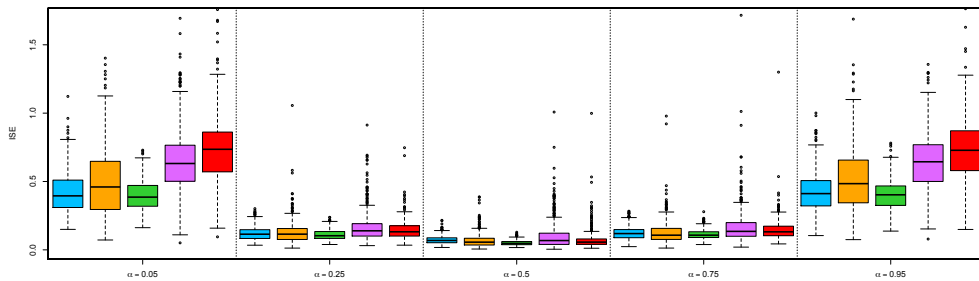


(a)  $n = 300$

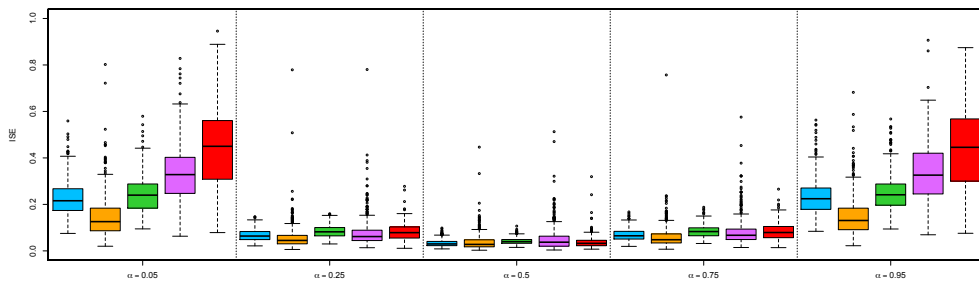


(b)  $n = 1,000$

Figure 10: The same boxplots as in Figure 9, but for Model ( $\mathcal{M}2$ ).



(a)  $n = 300$



(b)  $n = 1,000$

Figure 11: The same boxplots as in Figures 9-10, but for Model ( $\mathcal{M}3$ ).

	$\bar{q}_{\alpha,B}^{N,n}(\mathcal{N}_1)$	$\bar{q}_{\alpha,B}^{N,n}(\mathcal{N}_5)$	$\hat{q}_{\alpha}^{\text{SP}}$	$\hat{q}_{\alpha;k\text{NN}}$	$\hat{q}_{\alpha}^{\text{YJ}}$	$\hat{q}_{\alpha}^{\text{YJc}}$
CPU	11,108.3	2,583.8	2,258.6	541.4	1,319.8	1,178.4
(M1)	0.503	0.518	0.368	0.522	0.845	1.001
(M2)	1.938	2.211	2.614	1.412	3.535	2.950
Eff (M3)	1.108	1.109	1.223	1.046	1.618	1.784

Table 3: (First line:) Computation times (CPU, in seconds) used by each estimator to obtain Figure 9(a). (Subsequent lines:) ISE-based global efficiency measures (Eff) associated with Figures 9(a), 10(a), and 11(a); see Section 4.3 for details.

the Mac and Linux versions of our R package `QuantifQuantile` offer an option for parallel computing, which divides the computing times shown in Table 3 by a factor of 4, which clearly makes our estimators very competitive in this respect. In addition, the computation times for the nearest-neighbour estimator should be considered with care, since  $k\text{NN}$  procedures here are not based on a (typically always quite computationally intensive) data-driven selection of smoothing parameters. This table also gathers the efficiency measures Eff associated with Figures 10(a) and 11(a) (we do not provide the corresponding CPU times, that barely depend on the model considered). The grids  $\mathcal{N}_1$  and  $\mathcal{N}_5$  actually differ across models but not their size.

## 5. A real data example

In this section, we illustrate the behavior of  $\bar{q}_{\alpha,B}^{N,n}$  on a real data set, that was kindly sent to us by Dr. Keming Yu. This data set, of size  $n = 298$ , corresponds to the serum concentration, in grams per liter, of immunoglobulin-G in children aged from 6 months to 6 years. This data set was already investigated in Yu and Jones (1998), where a kernel local linear estimator and a double kernel estimator were considered; we therefore refer to that paper for (smoothed versions of) the corresponding quantile curves. Also, we do not plot the quantile curves resulting from the  $k\text{NN}$  estimator  $\hat{q}_{\alpha;k\text{NN}}$  since the selection of the parameter  $k$  is not data-driven, hence cannot be achieved on a real data set. Consequently, we only compare the proposed quantization-based quantile curves with those obtained from their main competitor, namely the spline estimator  $\hat{q}_{\alpha}^{\text{SP}}$ .

We now describe how we obtained quantization-based quantile curves in the present

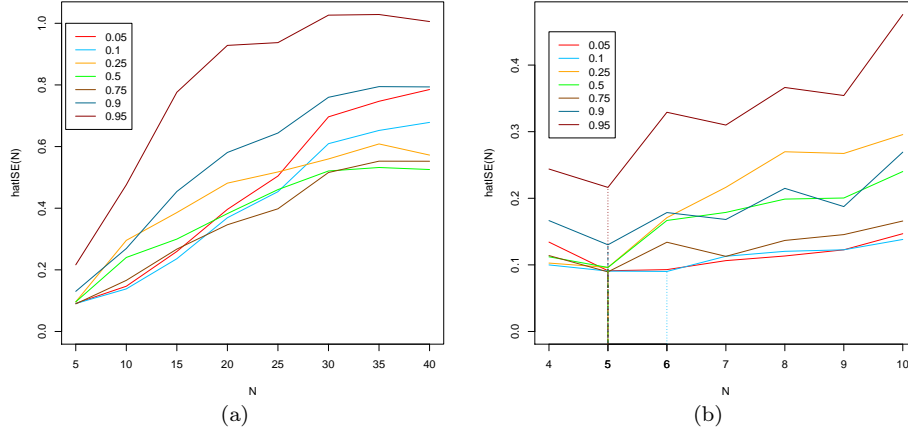


Figure 12: In the context of the real data example of Section 5, this figure plots, for various  $\alpha$ -values,  $N \mapsto \widehat{\text{ISE}}_{\alpha, B}^-(N)$  over  $\mathcal{N} = \{5, 10, \dots, 35, 40\}$  (left) and over  $\mathcal{N} = \{4, 5, \dots, 9, 10\}$  (right).

context, for each  $\alpha = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9$  and  $0.95$ .

- (i) The first step consists in choosing the optimal number  $N$  of quantizers. To do so, we adopted the method proposed in Section 3.2, based on the minimization of  $\widehat{\text{ISE}}_{\alpha, B}^-(N)$  (see (3.1)), for which we considered throughout  $B = 50$ ,  $\tilde{B} = 30$ , and a grid of  $J = 300$  equispaced points between the minimum and maximum values of the  $X$ -part of the sample. We first evaluated  $\widehat{\text{ISE}}_{\alpha, B}^-(N)$  for all  $N \in \mathcal{N} = \{5, 10, \dots, 35, 40\}$ . Figure 12(a) plots the resulting sample ISE curves for each  $\alpha$ . Irrespective of  $\alpha$ , these curves are monotone increasing, which means that the tested values of  $N$  are too large. We therefore did the same exercise for  $\mathcal{N} = \{4, 5, \dots, 9, 10\}$ , which led to Figure 12(b). For each  $\alpha$ , this provides an optimal  $N$  (that is equal to 5 or 6).
- (ii) The second step of course consists in obtaining the estimated quantile curves themselves, based on the selected values of  $N$  above. These are the plots of the mappings  $x \mapsto \tilde{q}_{\alpha, B}^{N, n}(x)$  in (2.7), where we chose  $B = 50$ .

The resulting quantile curves are plotted in Figure 13(a). As announced, we compare these curves with those associated with the spline estimator  $\hat{q}_{\alpha}^{\text{sp}}$ . For each  $\alpha$  considered, the parameter  $\lambda$  was selected according to the AIC procedure described in Section 4.1,

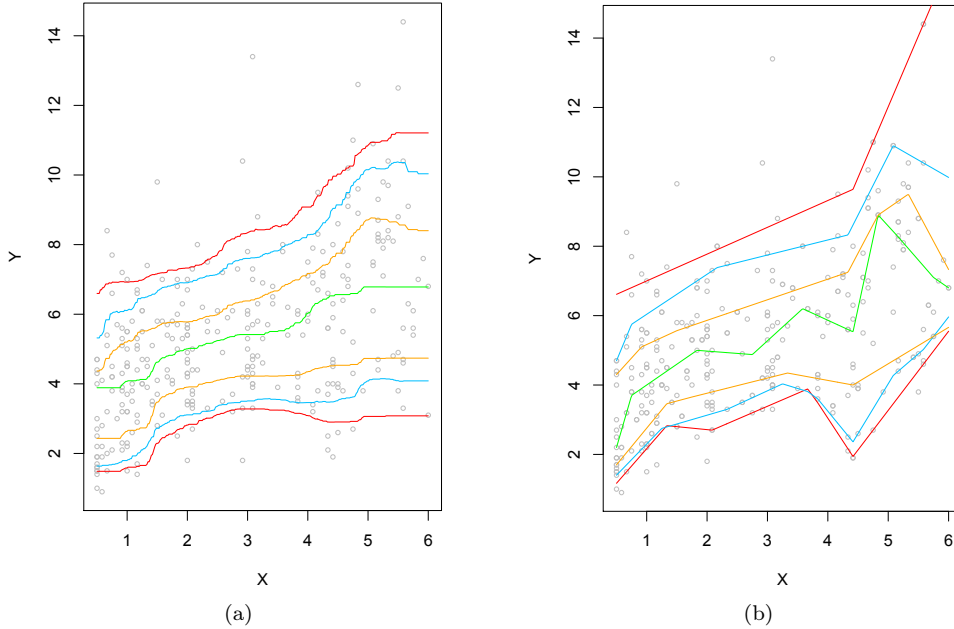


Figure 13: Conditional  $\alpha$ -quantile curves for the real-data example of Section 5 and  $\alpha = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9$  and  $0.95$ . The estimators considered are the quantization-based estimator  $\hat{q}_{\alpha, B}^{N, n}$  with  $B = 50$  (left) and the spline estimator  $\hat{q}_{\alpha}^{\text{SP}}$  (right).

which led to  $\hat{\lambda}_{\alpha; \text{opt}} = 0.40, 0.42, 0.96, 0.63, 0.96, 0.52$ , and  $0.69$ . Figure 13(b) plots the resulting quantile curves. Clearly, these piecewise linear curves are more irregular than the quantization-based ones and show some important peaks and slight crossings.

## 6. Extension to multivariate regressors ( $d > 1$ )

The numerical exercises and simulations in Sections 4 and 5 focused on the case of a single covariate ( $d = 1$ ). All previous definitions and results, however, cover the general case ( $d \geq 1$ ): this not only includes the theoretical results from CPS15 and the definitions of the quantization-based estimators described in Section 2, but also the CLVQ algorithm and the proposed data-driven method to select  $N$ . It is therefore natural to investigate how well quantization-based conditional quantile estimation performs for  $d > 1$ . Clearly, however, in the present context where any sparsity assumption or restriction on the

regression function are avoided,  $d$  should be small enough to avoid the usual curse of dimensionality.

For illustration purposes, we focus on the bivariate case  $d = 2$  to be able to plot the resulting conditional quantile hypersurfaces. We first generated a random sample of size  $n = 3,000$  from the model

$$(\mathcal{M4}) \ Y = \sin(X_1 + X_2) + (0.5 + 1.5 \sin^2(\frac{\pi}{2}(X_1 + X_2)))\varepsilon,$$

with  $X = (X_1, X_2)' = (6Z_1 - 3, 6Z_2 - 3)'$  ( $Z_i \sim \text{Beta}(2, 2)$ ,  $i = 1, 2$ ) and  $\varepsilon \sim \mathcal{N}(0, 9)$ , where  $Z_1, Z_2$  and  $\varepsilon$  are mutually independent. Conditional quantile hypersurfaces of this model for  $\alpha = 0.05, 0.25, 0.5, 0.75$ , and  $0.95$  are plotted in Figure 14(a). Figure 14(b) provides the corresponding sample quantile surfaces obtained from our estimator  $\bar{q}_{\alpha, B}^{N, n}$ , with  $B = 50$  and the optimal  $N$ -values that are obtained through our data-driven  $N$ -selection procedure for the various values of  $\alpha$ . The quantile surfaces were readily obtained from the R package `QuantifQuantile`, that can therefore also be used for  $d > 1$ . Clearly,  $\bar{q}_{\alpha, B}^{N, n}$  provides very smooth surfaces that properly catch the link function, even though the amplitude, for extreme values of  $\alpha$ , is a bit under-estimated.

The results of Section 4 suggest restricting the comparison to spline-based estimators. For this purpose, we consider the triogram-based bivariate splines from [Koenker and Mizera \(2004\)](#), that is implemented in the `rqss` function of the R package `quantreg`. The resulting estimated quantile surfaces, based on the same  $\lambda$ -selection method as for  $d = 1$  in Section 4, are plotted in Figure 14(c). Clearly, these quantile surfaces exhibit more variability than the quantization-based ones. Incidentally, note that, unlike the proposed estimator, this spline-based estimator does not extend easily to  $d > 2$ .

Parallel to our investigation of the case  $d = 1$ , we complement the results above with some efficiency results in terms of ISE. To do so, we generated 50 independent random samples from Model ( $\mathcal{M4}$ ) above with sample size  $n = 3,000$  and computed, for each sample and various values of  $\alpha$ , the quantization-based and spline-based conditional quantile estimators. The corresponding 50 resulting ISEs for each estimator and every  $\alpha$  considered, that are defined in a similar way as in Section 4.3, are represented in Figure 15. In line with the  $d = 1$  results in Figure 11, this shows that, for the complex link function considered,  $\bar{q}_{\alpha, B}^{N, n}$  dominates  $\hat{q}_{\alpha}^{\text{SP}}$  in terms of efficiency. Note that this is particularly true for extreme values of  $\alpha$ .

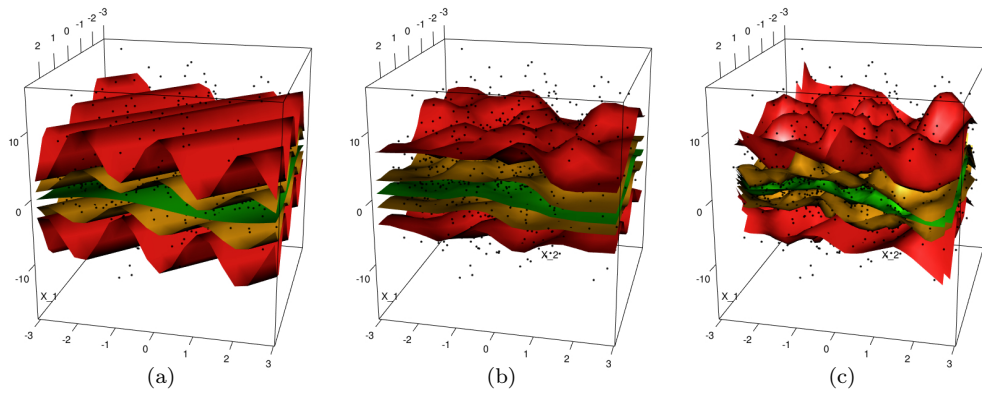


Figure 14: Various conditional  $\alpha$ -quantile surfaces for Model ( $\mathcal{M}4$ ) and  $\alpha = 0.05, 0.25, 0.5, 0.75$  and  $0.95$ . The quantile surfaces are the population ones (left), the ones estimated (from 3,000 independent observations) through quantization (center), and through triogram-based splines (right); see Section 6 for details.

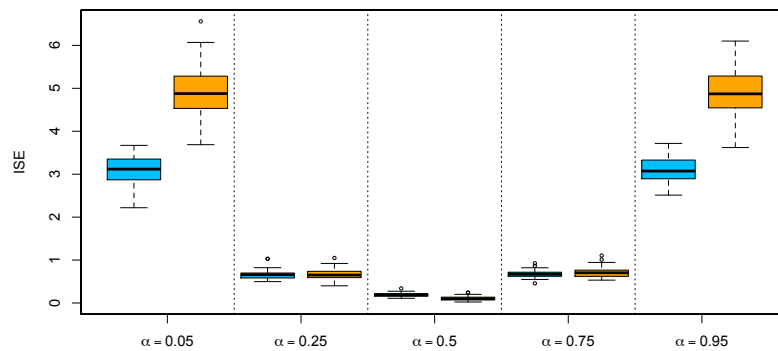


Figure 15: Boxplots, for  $\alpha = 0.05, 0.25, 0.50, 0.75,$  and  $0.95$ , of the ISEs of quantization-based (in blue) and triogram-based spline (in orange) conditional  $\alpha$ -quantile estimators obtained from 500 independent random samples from Model ( $\mathcal{M}4$ ), with size  $n = 3,000$ .

## 7. Summary and final comments

In this paper, we investigated the empirical performances of the quantization-based estimators of conditional quantile that were recently introduced in [CPS15](#). This led us to design an efficient data-driven method to select the number  $N$  of quantizers to be used, that plays the role of a smoothing parameter. From extensive simulations, we concluded that quantization-based estimators compete well with alternative conditional quantile estimators and sometimes dominate its main competitor, spline estimators, particularly when the link function is complex. We treated a real data example, in which we showed that the proposed methodology provides very satisfactory conditional quantile curves, and we saw that the good properties of quantization-based estimators extend to the bivariate covariate case. This should make conditional quantile estimation based on quantization of interest to practitioners; in this spirit, we wrote an R package, named `QuantifQuantile`, that allows to compute in a straightforward way the proposed estimators (the data-driven selection of  $N$  is included) and to plot the resulting quantile curves/surfaces. This package is already available on the CRAN.

We conclude with two final comments.

(i) Since conditional quantiles  $q_\alpha(x)$  are monotone increasing in  $\alpha$  for any fixed  $x$ , population conditional quantile hypersurfaces cannot cross. Yet, like many of their competitors, the proposed quantization-based quantile hypersurfaces might cross. No such crossings occurred in the illustrations of this paper, nor in all other numerical experiments we have conducted. Practitioners, however, usually do not like such crossings, that violate their interpretation of quantiles. An easy way to exclude crossings is to select an optimal  $N$ -value that is independent of  $\alpha$ . A natural way to do so is to minimize

$$N \mapsto \widehat{\text{ISE}}_B(N) = \text{AVE}_\alpha \left[ \widehat{\text{ISE}}_{\alpha,B}(N) \right], \quad (7.1)$$

where the average extends over all  $\alpha$ -values considered. [Figure 16](#) plots the resulting ISE curve for the real data example from [Section 5](#), which leads to the optimal value  $N = 5$ . Another procedure that allows to avoid crossings is the rearrangement procedure from [Chernozhukov et al. \(2010\)](#).

(ii) Both in [CPS15](#) and in this work, quantization was applied to the covariate only. One may wonder whether or not it may be of interest to move entirely to a discrete



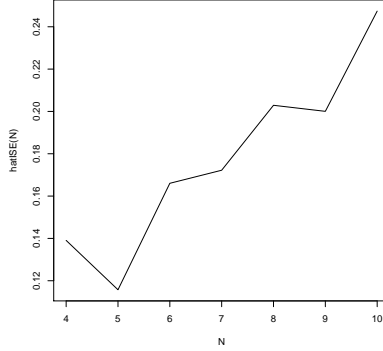


Figure 16: Plot of the mapping  $N \mapsto \widehat{\text{ISE}}_B(N)$  in (7.1) for the real data example of Section 5.

framework by applying quantization to both the covariate and the response. Such a double quantization might involve different numbers of quantizers for  $X$  and  $Y$  ( $N_X$  and  $N_Y$ , say) and would lead to the approximate conditional  $\alpha$ -quantile

$$\widehat{q}_\alpha^{Y, N_X, N_Y}(x) = \arg \min_{a \in \mathbb{R}} \mathbb{E}[\rho_\alpha(\widetilde{Y}^{N_Y} - a) | \widetilde{X}^{N_X} = \tilde{x}],$$

where  $\widetilde{Y}^{N_Y} = \text{Proj}_{\delta^{N_Y}}(Y)$  denotes the projection of  $Y$  onto  $\delta^{N_Y} = (\tilde{y}_1, \dots, \tilde{y}_{N_Y})$ , an optimal  $N$ -grid for  $Y$  ( $\tilde{x}$  still denotes the projection of  $x$  onto the optimal  $N_X$ -grid for  $X$ ). If observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  are available, then the resulting estimator of  $\widehat{q}_\alpha^{Y, N_X, N_Y}(x)$  is the sample  $\alpha$ -quantile,  $\widehat{q}_\alpha^{Y, N_X, N_Y, n}(x)$  say, of the  $\widehat{Y}_i^{N_Y, n} = \text{Proj}_{\delta^{N_Y, n}}(Y_i)$ 's corresponding to the indices  $i$  for which  $\widehat{X}_i^{N_X, n} = \tilde{x}^{N_X, n}$  (the optimal  $y$ -grid  $\widehat{\delta}^{N_Y, n} = (\widehat{y}_1^{N_Y, n}, \dots, \widehat{y}_{N_Y}^{N_Y, n})$  can be obtained from the CLVQ algorithm). Of course, we can also define a bootstrap version  $\widehat{q}_{\alpha, B}^{Y, N_X, N_Y, n}(x)$  of this estimator by proceeding as in Section 2.3.

To investigate the performance of double quantization, we generated a random sample of size  $n = 500$  from the model

$$Y = \frac{1}{5}X^3 + \varepsilon, \quad (7.2)$$

where  $X \sim U(-3, 3)$  and  $\varepsilon \sim \mathcal{N}(0, 1)$  are independent. Figure 17 plots, for five values of  $\alpha$ , the  $\widehat{\text{ISE}}_{\alpha, B}^-(N_X, N_Y)$  quantities achieved by the double-quantization estimator  $\widehat{q}_{\alpha, B}^{Y, N_X, N_Y, n}(x)$  with  $B = 50$ ; we only considered  $N_X \in \{5, 10, \dots, 75, 80\}$  and  $N_Y \in \{5, 10, \dots, 245, 250\}$ . In order to get smoother surfaces, we actually averaged

sample ISEs over 5 independent random samples. The white curve in each panel is the curve associated with  $N_Y \mapsto \widehat{\text{ISE}}_{\alpha,B}^-(\hat{N}_{\alpha,B;\text{opt}}^-(N_Y), N_Y)$ , where  $\hat{N}_{\alpha,B;\text{opt}}^-(N_Y)$  is the optimal value for  $N_X$  at the given  $N_Y$  (i.e., the one that minimizes the ISE function  $N \mapsto \widehat{\text{ISE}}_{\alpha,B}^-(N, N_Y)$ ). It is seen that, irrespective of  $N_Y$  and  $\alpha$ , the optimal number of quantizers for  $X$  stays close to 25. However, there is no optimal value for  $N_Y$ : the larger  $N_Y$ , the smaller the averaged ISE. In conclusion, an optimal choice for  $(N_X, N_Y)$  appears to be  $N_X$  close to 25 and  $N_Y \rightarrow \infty$ . Since letting the size of the quantization grid go to infinity is equivalent to not quantizing at all, this leads to favouring the solution adopted in [CPS15](#) and in the present work over double quantization. Quantizing  $X$  only also makes perfect sense in the problem of estimating nonparametrically a given conditional quantile  $q_\alpha(x)$ . Most estimation methods indeed consist in selecting the observations whose  $X$ -part is the closest to  $x$  (where the meaning of “close” depends on the method at hand) and to take the sample quantile of the corresponding  $Y$ ’s. Quantizing  $Y$  thus does not seem natural, as it adds an unnecessary approximation error.

## Acknowledgments

The authors are grateful to an Associate Editor and two anonymous referees for their careful reading and insightful comments that led to a significant improvement of the original manuscript. The first author’s research is supported by a Bourse FRIA of the Fonds National de la Recherche Scientifique, Communauté française de Belgique. The second author’s research is supported by an A.R.C. contract from the Communauté Française de Belgique and by the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy).

## References

- Belloni, A., Chernozhukov, V., Fernández-Val, I., 2011. Conditional quantile processes based on series or many regressors.
- Bhattacharya, P. K., Gangopadhyay, A. K., 1990. Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* 18 (3), 1400–1415.
- Cattaneo, M. D., Farrell, M. H., 2013. Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators. *J. Econometrics* 174 (2), 127–143.

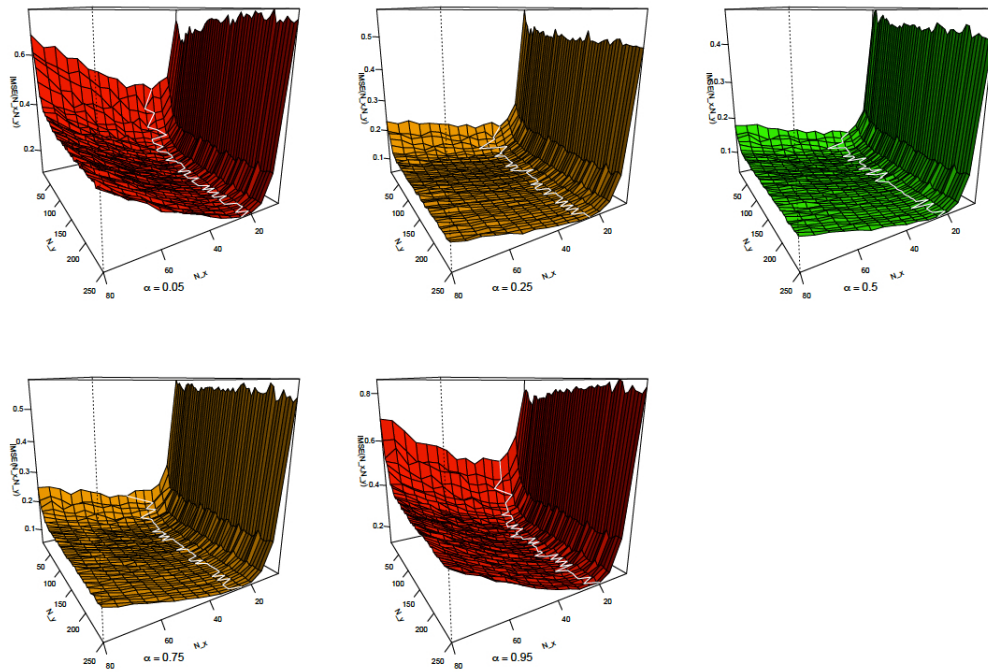


Figure 17: Plots of the mapping  $(N_X, N_Y) \mapsto \widehat{\text{ISE}}_{\alpha, B}^{-}(N_X, N_Y)$  (for the estimator  $\widehat{q}_{\alpha, B}^{Y, N_X, N_Y, n}(x)$  with  $B = 50$ ), for  $\alpha = 0.05, 0.25, 0.50, 0.75$ , and  $0.95$ , averaged over 5 mutually independent random samples from (7.2) with sample size  $n = 500$ .

- Charlier, I., Paindaveine, D., Saracco, J., 2015. Conditional quantile estimators through optimal quantization. *J. Statist. Plann. Inference* 156, 14–30.
- Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* 19 (2), 760–777.
- Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossings. *Econometrica* 78, 1093–1125.
- de Saporta, B., Dufour, F., Gonzalez, K., 2010. Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.* 20 (5), 1607–1637.
- Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. Vol. 38 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa.
- Fan, J., Hu, T.-C., Truong, Y., 1994. Robust nonparametric function estimation. *Scandinavian Journal of Statistics* 21 (4), 433–446.
- Fischer, A., 2010. Quantization and clustering with bregman divergences. *Journal of Multivariate Analysis* 101, 2207–2221.
- Gannoun, A., Girard, S., Guinot, C., Saracco, J., 2002. Reference curves based on non-parametric quantile regression. *Statistics in Medicine* 21 (4), 3119–3135.
- Graf, S., Luschgy, H., 2000. Foundations of quantization for probability distributions. Vol. 1730 of Lecture Notes in Mathematics. Springer-Verlag, Berlin.
- Györfi, L., Kohler, M., Krzyżak, A., Walk, H., 2002. A distribution-free theory of nonparametric regression. *Springer Series in Statistics*. Springer-Verlag, New York.
- Hallin, M., Lu, Z., Yu, K., 2009. Local linear spatial quantile regression. *Bernoulli* 15 (3), 659–686.
- He, X., Ng, P., Portnoy, S., 1998. Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60 (3), 537–550.
- Horowitz, J. L., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* 100 (472), 1238–1249.
- Koenker, R., 2005. Quantile regression. Vol. 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.
- Koenker, R., Bassett, Jr., G., 1978. Regression quantiles. *Econometrica* 46 (1), 33–50.
- Koenker, R., Mizera, I., 2004. Penalized triograms: total variation regularization for bivariate smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (1), 145–163.
- Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. *Biometrika* 81 (4), 673–680.
- Kohler, M., Krzyżak, A., Walk, H., 2006. Rates of convergence for partitioning and nearest neighbor regression estimates with unbounded data. *Journal of Multivariate Analysis* 97, 311–323.
- Lee, S., 2003. Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric theory* 19 (01), 1–31.
- Marron, J. S., Wand, M. P., 1992. Exact mean integrated squared error. *Ann. Statist.* 20 (2), 712–736.
- Pagès, G., 1998. A space quantization method for numerical integration. *J. Comput. Appl. Math.* 89 (1), 1–38.

- Pagès, G., Printems, J., 2003. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl.* 9 (2), 135–165.
- Yu, K., Jones, M. C., 1998. Local linear quantile regression. *J. Amer. Statist. Assoc.* 93 (441), 228–237.
- Yu, K., Lu, Z., Stander, J., 2003. Quantile regression: applications and current research areas. *J. R. Stat. Soc. Ser. D Statistician* 52 (3), 331–350.