Numerical Study of a Conditional Quantile Estimator based on Optimal Quantization

Isabelle Charlier, Davy Paindaveine, Jérôme Saracco

Numerical Study of a Conditional Quantile Estimator based on Optimal Quantization

Isabelle Charlier[(1,2,3)*] Davy Paindaveine[(1,2)†] Jérôme Saracco[(3)]

May 16, 2014

[(1)] *Université Libre de Bruxelles, Département de Mathématique, Boulevard du Triomphe, Campus Plaine, CP210, B-1050, Bruxelles, Belgique.*

ischarli@ulb.ac.be, dpaindav@ulb.ac.be

[(2)] *ECARES, 50 Avenue F.D. Roosevelt, CP114/04, B-1050, Bruxelles, Belgique.*

[(3)] *Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 et INRIA Bordeaux Sud-Ouest, équipe CQFD, 351 Cours de la Libération, 33405 Talence.*

Jerome.Saracco@math.u-bordeaux1.fr

Abstract

Charlier, Paindaveine, and Saracco (2014) recently introduced a promising nonparametric estimator of conditional quantiles based on optimal quantization, but almost exclusively focused on its theoretical properties. In this paper, (i) we discuss its practical implementation (by proposing in particular a method to properly select the corresponding smoothing parameter, namely the number of quantizers) and (ii) we investigate how its finite-sample performances compare with those or classical kernel of nearest-neighbor competitors. Monte Carlo studies show that the quantization-based estimator competes well in all cases (in terms of mean squared errors) and tends to dominate its competitors as soon as the covariate is not uniformly distributed over its support. We also apply our approach to a real data set. While most of the paper focuses on the case of a univariate covariate, we also briefly discuss the multivariate case and provide an illustration for bivariate regressors.

*Research supported by a Bourse F.R.I.A. of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

†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).

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 them, standard (mean) regression, in most cases, is way less informative than quantile regression, that provides a more complete picture of conditional distributions: quantile regression may indeed heteroscedasticity, conditional asymmetry, etc. Conditional quantiles are of major interest since they allow to construct reference hypersurfaces (curves if the covariate is univariate) and conditional prediction intervals. Therefore, they are widely used in many different areas, including medicine, economic or lifetime analysis.

Throughout this paper, we consider quantile regression in the nonparametric regression framework for which a real response variable Y is generated as

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

where X is a d-dimensional covariate and ε is a random error term that is independent of X. In practice, both the distribution of ε and the link function m 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 $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (1.1). Any consistent estimator of conditional quantiles in this context requires some localization technique, the most classical ones being of a kernel or nearest-neighbor nature; see Section **??**.

Recently, Charlier, Paindaveine, and Saracco (2014) introduced an alternative estimator that performs localization through the concept of *optimal quantization*, a numerical tool first used in engineering, and that projects the (continuous) covariate X on a finite grid of N *quantizers*. The emphasis in that paper was mainly on theoretical aspects, quantifying how well the resulting quantization-based conditional approximate the original conditional quantile as N grows (see Theorem ? below) and providing the convergence in probability of a sample quantization-based quantile to its population version (see Theorem ? below). Practical implementation, however, was barely touched in Charlier, Paindaveine, and Saracco (2014), and finite-sample performances were not investigated. The goal of the present work is therefore to carefully cover these aspects.

The outline of the paper is as follows. In Section 2, we discuss briefly the concept of optimal quantization and describe the quantization-based estimators from Charlier et al. (2014). Then we discuss in Section 4 the choice of the number N of quantizers, which, in the nonparametric setup considered, plays the role of a smoothing parameter. In Section 5, we then compare our estimator with other well-known nonparametric estimators of conditional quantiles, namely the k nearest-neighbor (k-NN) estimator (as those of Bhattacharya and Gangopadhyay (1990)) and the local linear and local constant estimators (Yu and Jones (1998)). We first compare

them through some illustrations of the estimated conditional quantile curves in different models. We also base the comparison on empirical mean squared errors, for various models and sample sizes. A real data application is described in Section 6. Finally, some extensions and concluding remarks are given in Section 7.

2 Conditional quantiles through optimal quantization

2.1 Optimal quantization

Initially, quantization consists in discretizing a continuous signal with a finite number of points chosen in an optimal way. Optimality means here that the location of the points has to make the transmission of the signal as efficient as possible. Since then, it was used in other fields as cluster analysis and more recently mathematics. In this context, optimal quantization consists in finding the best approximation of a continuous distribution with a discrete distribution, with support of size N. The quantization was then applied to solve many problems in numerical probabilities, stochastic processes or finance (see Pagès et al. (2004a,b); Bally et al. (2005)), and more rarely in statistics (see Azaïs et al. (2012); Fischer (2010, 2014)). Let us now define more precisely L_p-norm optimal quantization.

Let us consider a random d-vector X defined on a probability space $(\Omega, \mathcal{F}, P)$, with distribution P_X for which $\|X\|_p = (\mathrm{E}[|X|^p])^{1/p}$ is finite (with $p \geq 1$ and $|\cdot|$ denoting the Euclidean norm). The goal of quantization is to replace X with an appropriate random d-vector $\pi(X)$ taking at most N values. Optimal L_p-norm quantization consists then in selecting the vector $\pi(X)$ that minimizes the L_p-norm quantization error

$$\|X - \pi(X)\|_p^p.$$

In other words, quantization finds an N-grid of $\mathbb{R}^d$, denoted γ^N, such that the projection $\widetilde{X}^{\gamma^N} = \mathrm{Proj}_{\gamma^N}(X)$ of X on this grid minimizes the *quantization error* $\|X - \widetilde{X}^{\gamma^N}\|_p^p$.

Under the assumption that P_X does not charge any hyperplane, the existence of such an optimal N-grid has been obtained, see Pagès (1998). However, no unicity result is available. One denotes by $\widetilde{X}^N$ the projection of X on an arbitrary optimal N-grid. The convergence of $\widetilde{X}^N$ to X in L_p is well known. The following result provides the rate of convergence, an essential step to get convergence of our approximation; see, e.g., Graf and Luschgy (2000) for a proof.

Theorem 2.1. *Assume that* $\|X\|_{p+\delta} < \infty$ *for some* $\delta > 0$. *Then, for some* C, $D \in \mathbb{R}$ *and* $N_0 \in \mathbb{N}$, *we have that*

$$\|X - \widetilde{X}^N\|_p^p \leq \frac{1}{N^{p/d}}(C\|X\|_{p+\delta}^{p+\delta} + D),$$

for all $N \geq N_0$.

Generally, there exists no result that describes the geometric structure of optimal N-grids and we use a *stochastic gradient algorithm* to determine an optimal grid computationally. Let $(\xi_t)_{t\in\mathbb{N}_0}$, with $\mathbb{N}_0 = \{1, 2, \ldots\}$, be a sequence of independent and identically P_X-distributed random vectors, and let $(\delta_t)_{t\in\mathbb{N}_0}$ be a deterministic sequence in $(0, 1)$ such that

$$\sum_t \delta_t = +\infty \quad \text{and} \quad \sum_t \delta_t^2 < +\infty.$$

Starting from a deterministic N-tuple $X^0 = x^0$, with N pairwise distinct components, we define recursively the grid X^t, $t \geq 1$, as

$$X^t = X^{t-1} - \frac{\delta_t}{p} \nabla_x d_N^p(X^{t-1}, \xi^t), \tag{2.1}$$

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

$$\big(\nabla_x d_N^p(x, \xi)\big)_i = p\,|x_i - \xi|^{p-1} \frac{x_i - \xi}{|x_i - \xi|} \mathbb{I}_{[x_i = \mathrm{Proj}_x(\xi)]},$$

with the convention $0/0 = 1$ when $x_i = \xi$. Then, X^{t-1} and X^t are the same grids, except one point which has moved: the one corresponding to the non zero component of this gradient.

More details about this algorithm and results of convergence can be found in Pagès (1998) or Pagès and Printems (2003) for instance. Note that it is more commonly used for $p = 2$ and is then referred as the *Competitive Learning Vector Quantization (CLVQ)* algorithm, mainly because the results of convergence are the more satisfactory in this case.

2.2 Conditional quantiles through optimal quantization

We used optimal quantization to approximate conditional quantiles (see Charlier et al. (2014) for more details). We will recall here the corresponding approximation and the main theoretical results obtained. We consider a regression setup where X and Y are continuous variables, and we quantize only the regressor X as explained in that paper.

Fix $p \geq 1$ such that $X \in \mathrm{L}_p$. For any fixed number N of quantizers, we may then construct the optimal discrete approximation $\widetilde{X}^N$ of X, obtained by projecting X onto an optimal N-grid γ^N. Theorem 2.1 implies that this approximation becomes arbitrarily good as N goes to infinity.

Denoting by $\tilde{x}$ the projection of the point $x \in \mathbb{R}^d$ onto the optimal grid γ^N, we may then approximate $q_\alpha(x)$ in (**??**) by

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

Convergence results were derived in Charlier et al. (2014) under the following assumptions.

ASSUMPTION (A) (i) The random vector (X, Y) is generated through $Y = m(X, \varepsilon)$, where the d-dimensional covariate vector X and the error ε are mutually independent; (ii) the link function $(x, z) \mapsto m(x, z)$ is of the form $m_1(x) + m_2(x)z$, where the functions $m_1(\cdot) : \mathbb{R}^d \to \mathbb{R}$ and $m_2(\cdot) : \mathbb{R}^d \to \mathbb{R}_0^+$ are Lipschitz functions; (iii) $\|X\|_p < \infty$ and $\|\varepsilon\|_p < \infty$; (iv) the distribution of X does not charge any hyperplane.

This assumption can be reinforced to obtain a convergence rate.

ASSUMPTION (A′) Same as Assumption (A), but with (iii) replaced by (iii)′ there exists $\delta > 0$ such that $\|X\|_{p+\delta} < \infty$, and $\|\varepsilon\|_p < \infty$.

ASSUMPTION (B) (i) The support S_X of P_X is compact; (ii) ε admits a continuous density $f^\varepsilon : \mathbb{R} \to \mathbb{R}_0^+$ (with respect to the Lebesgue measure on $\mathbb{R}$).

We then have the following results (see Charlier et al. (2014) for discussion and proofs).

Theorem 2.2. *Fix $\alpha \in (0, 1)$. Then (i) under Assumptions (A)-(B),*

$$\|\widetilde{q}_\alpha^N(X) - q_\alpha(X)\|_p \leq 2\sqrt{\max\left(\frac{\alpha}{1-\alpha}, \frac{1-\alpha}{\alpha}\right)}\,[m]_{\mathrm{Lip}}^{1/2}\,\left\|L^N(X)\right\|_p^{1/2}\,\|X - \widetilde{X}^N\|_p^{1/2},$$

for N sufficiently large, where $(L^N(X))$ is a sequence of X-measurable random variables that is bounded in L_p; (ii) under Assumptions (A′)-(B),

$$\|\widetilde{q}_\alpha^N(X) - q_\alpha(X)\|_p = O(N^{-1/2d}),$$

as $N \to \infty$.

Theorem 2.3. *Fix $\alpha \in (0, 1)$. Then, under Assumptions (A)-(B),*

$$\sup_{x \in S_X} \left|\widetilde{q}_\alpha^N(x) - q_\alpha(x)\right| \to 0,$$

as $N \to \infty$.

3 Conditional quantile estimator through optimal quantization

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of (X, Y). We construct in Charlier et al. (2014) an estimator $\widehat{q}_\alpha^{N,n}(x)$ of the conditional quantile $q_\alpha(x)$ from the approximation $\widetilde{q}_\alpha^N(x)$ in (2.2). We proceed in two steps.

(S1) Using the CLVQ algorithm (see Section **??**), X is quantized as follows: (i) the initial grid X^0 is obtained by sampling randomly among the X_i's without replacement under the constraint that the same x-values cannot be picked more than once; (ii) after n iterations (based on $\xi^t = X_t$, $t = 1, \ldots, n$), we get an "optimal" grid denoted by $\hat{\gamma}^{N,n} = (\hat{x}_1^{N,n}, \ldots, \hat{x}_N^{N,n})$, and we write $\widehat{X}^{N,n} (= \mathrm{Proj}_{\hat{\gamma}^{N,n}}(X))$ for the corresponding quantization of X. We will often not stress the dependence on n in these quantities for simplicity of notations.

(S2) Second, $\widetilde{q}_\alpha^N(x) = \arg\min_a \mathrm{E}[\rho_\alpha(Y - a)|\widetilde{X}^N = \tilde{x}]$ is then estimated by

$$\widehat{q}_\alpha^{N,n}(x) = \arg\min_a \sum_{i=1}^n \rho_\alpha(Y_i - a)\, \mathbb{I}_{[\widehat{X}_i^N = \hat{x}^N]},$$

where $\widehat{X}_i^N = \widehat{X}_i^{N,n} = \mathrm{Proj}_{\hat{\gamma}^{N,n}}(X_i)$ and $\hat{x}^N = \hat{x}^{N,n} = \mathrm{Proj}_{\hat{\gamma}^{N,n}}(x)$.

It is important to note that this grid is not optimal in the sense that it minimizes the quantization error. However, it converges towards an optimal one under some assumptions (see Section **??**), as the number of iterations (i.e. the sample size) increases. Then, it is expected that only moderate-to-large n will provide reasonable approximations of optimal N-grids.

We obtained and proved in Charlier et al. (2014) the consistency of this estimator $\widehat{q}_\alpha^{N,n}(x)$ (for fixed N) to our approximation $\widetilde{q}_\alpha^N(x)$ with the following additional assumption.

Assumption (C) P_X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

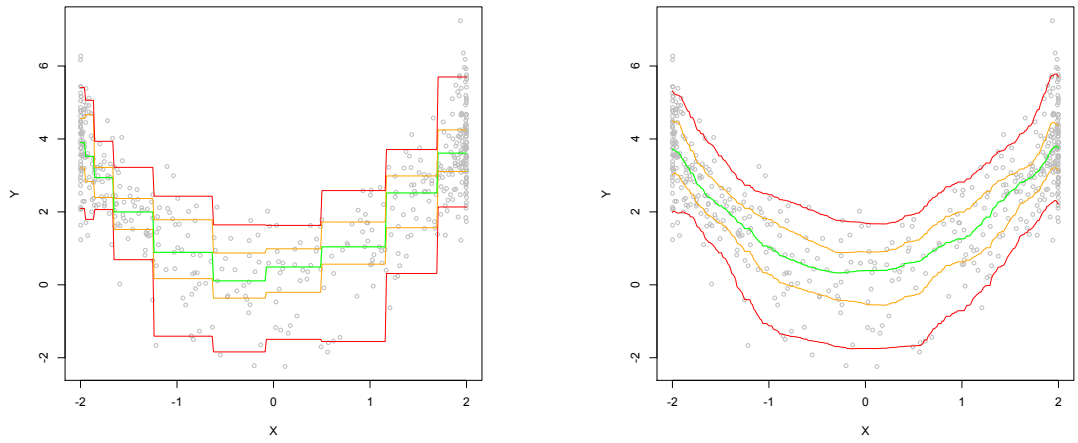
Theorem 3.1. *Fix $\alpha \in (0,1)$, $x \in S_X$ and $N \in \mathbb{N}_0$. Then, under Assumptions (A), (B)(i), and (C), we have that, as $n \to \infty$,*

$$|\widehat{q}_\alpha^{N,n}(x) - \widetilde{q}_\alpha^N(x)| \to 0,$$

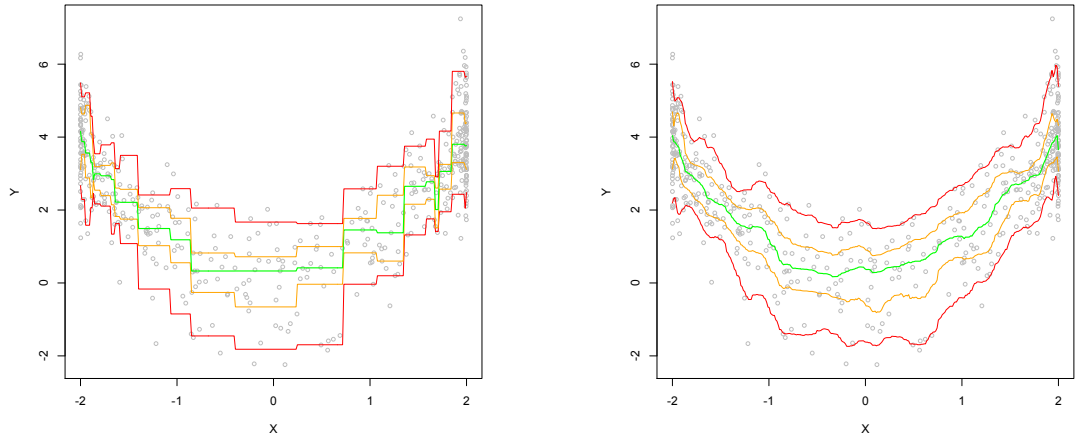
in probability, provided that quantization is based on $p = 2$.

We discussed in Charlier et al. (2014) that it is more appropriate in practice to use a bootstrap version of this estimator, particularly when the sample size is small or moderate. Taking $B \in \mathbb{N}_0$, we generate B samples of size n from the original sample with replacement. For each of these samples, we apply (S1) to obtain B grids, and (S2) with these grids and the original sample $\{(X_i, Y_i),\ i = 1, \ldots, n\}$, which provides B estimations $\hat{q}_\alpha^{(b)}(x) = \hat{q}_\alpha^{(b),N,n}(x)$, $b = 1, \ldots, B$. We define the bootstrap estimator $\bar{q}_\alpha^{N,n}(x)$ by taking the mean :

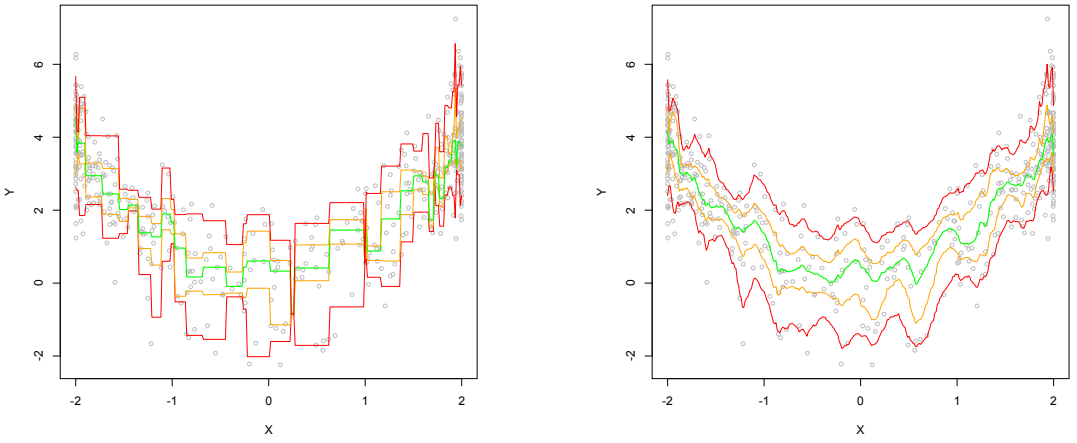
$$\bar{q}_\alpha^{N,n}(x) = \frac{1}{B} \sum_{b=1}^B \hat{q}_\alpha^{(b)}(x).$$



(a) $N = 10, B = 1$ (b) $N = 10, B = 50$



(c) $N = 25, B = 1$ (d) $N = 25, B = 50$



(e) $N = 50, B = 1$ (f) $N = 50, B = 50$

Figure 1: Estimated conditional quantile curves $x \mapsto \widehat{q}_\alpha^{N,n}(x)$ (left) and $x \mapsto \bar{q}_\alpha^{N,n}(x)$ for $B = 50$ (right), based on various N. In all cases, the sample size is $n = 500$, the quantile levels considered are $\alpha = 0.05$, 0.25, 0.5, 0.75 and 0.95; see (3.1) for the data generating model.

In the sequel, we will adopt the notation $B = 1$ when we do not perform bootstrap.

Figure 1 provides a first illustration of the behavior of the bootstrap and non bootstrap versions of the estimator. We computed $\widehat{q}_\alpha^{N,n}(x)$ and $\bar{q}_\alpha^{N,n}(x)$, for $N = 10$, $N = 25$ and $N = 50$ from $n = 500$ data points obtained with the model

$$Y = X^2 + \varepsilon, \tag{3.1}$$

where $X = 4Z - 2$, with $Z \sim \text{Beta}(0.3, 0.3)$, and $\varepsilon \sim \mathcal{N}(0, 1)$ are independent. One can observe that the number of quantizers has an important impact on the estimated curves. The middle value $N = 25$ is considered as optimal since it was chosen with a method proposed in Section 4. For small N ($N = 10$), the curves are not smooth enough without bootstrap and show a large bias. When bootstrap is performed, the curves are really smoother but it remains a bound effect. For larger N ($N = 50$), the bias is smaller but the variability is large (in both non-bootstrap and bootstrap versions). We conclude from this figure that bootstrapping clearly and significantly smooths all curves if B is not too small. However, B should not be taken very large, otherwise the computational time will increase. We propose, and use in the sequel, $B = 50$.

4 Determining the number of quantizers N

This section is devoted to the development of a method for determining the tuning parameter N. First, we propose to investigate in Section 4.1 the true MSE according to the size N of the grid and we observe its convexity and the existence of a minimum. Second, we propose in Section 4.2 a data-driven selection method for N obtained by substituting the true MSE for an estimated bootstrap version. We then compare the value of N selected with our data-based method and the one with the true MSE.

4.1 Choice of N based on the true Mean Squared Error

We simulate data set of size n according to the models

($\mathcal{M}1$) $Y = \frac{1}{5}X^3 + \varepsilon$,
($\mathcal{M}2$) $Y = \frac{1}{5}(X')^3 + \varepsilon$,

where $X = 6Z - 3$, with $Z \sim \text{Beta}(0.3, 0.3)$, $X' \sim U(-3, 3)$ and with $\varepsilon \sim \mathcal{N}(0, 1)$ independent of X and X'. We consider a grid of $\mathcal{N}_x$ points equispaced in the interval $(-3, 3)$. For each point x_j of this grid, we calculate the theoretical conditional quantile $q_\alpha(x_j)$ and we estimate this conditional quantile for various values of N. In this model ($\mathcal{M}1$), we have $q_\alpha(x) = x^3/5 + (F^\varepsilon)^{-1}(\alpha)$, where F^ε denotes the cumulative distribution function of ε, since $P(Y \leq q_\alpha(x)|X = x) = P(\varepsilon \leq$

$q_\alpha(x) - x^3/5) = \alpha$ (similarly for X'). The fact that we know the exact value of the conditional quantile (since the model is simulated) is used to evaluate the quality of the estimators by comparing the estimated conditional quantiles to the true ones using the mean squared error criterion.

For a fixed N, we evaluate the MSE(N) on this grid:

$$\text{MSE}(N) = \frac{1}{\mathcal{N}_x} \sum_{j=1}^{\mathcal{N}_x} \left(q_\alpha(x_j) - q_\alpha'^N(x_j)\right)^2,$$

with $q_\alpha'^N(x_j) = \widehat{q}_\alpha^{N,n}(x_j)$ or $\bar{q}_\alpha^{N,n}(x_j)$ (depending on the estimator we choose). One way to determine an optimal value of N is to choose the value of N that minimizes this MSE. We define

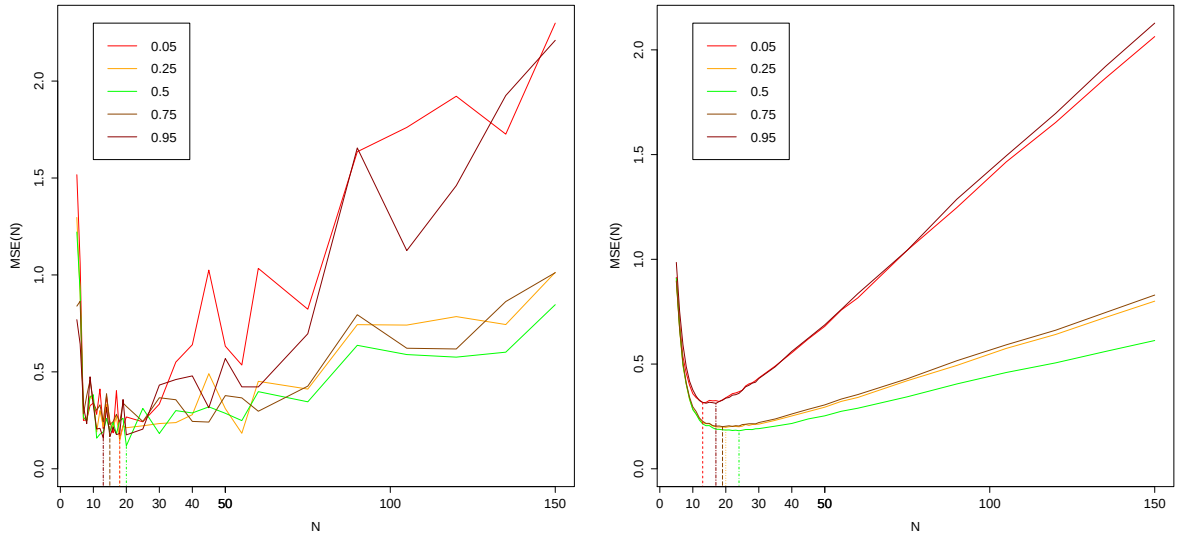
$$N^* = \arg\min_{N \in \mathfrak{N}} \text{MSE}(N),$$

where $\mathfrak{N}$ denotes a grid of values for N chosen according to the sample size of the considered data set. We will also determine whether this optimal value N^* depends on α.

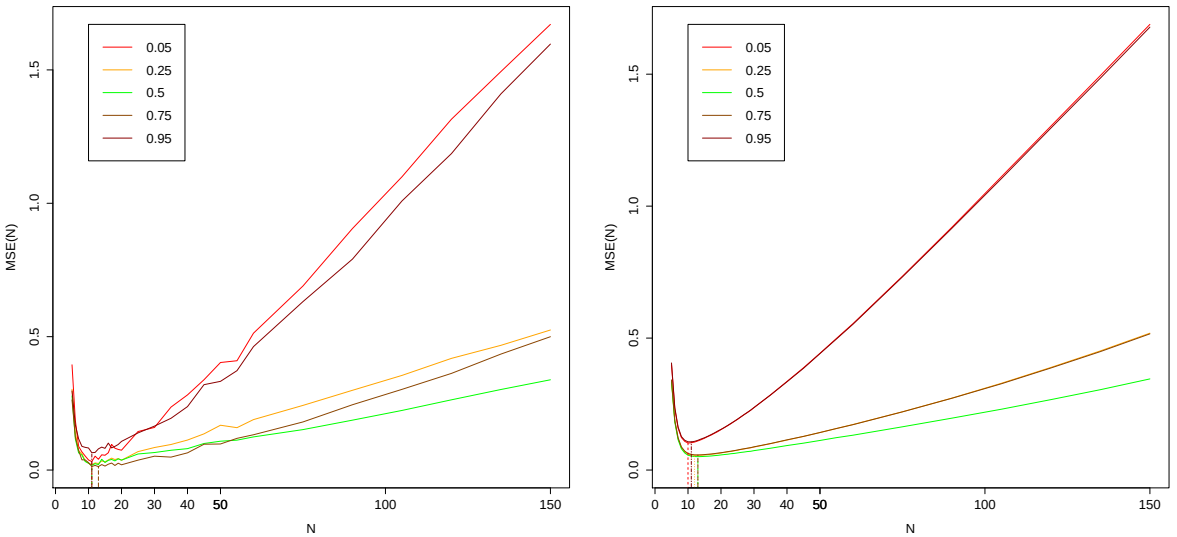
First, we realized some graphs representing the function of the MSE given N for different values of α (α=0.05, 0.25, 0.5, 0.75 and 0.95), with $n = 300$. Figures 2 and 3 correspond respectively to data generated from models ($\mathcal{M}1$)and ($\mathcal{M}2$). On the left side of Figures 2 and 3, we consider only one sample and the upper and lower graphs were obtained by using $\widehat{q}_\alpha^{N,n}(x)$ and $\bar{q}_\alpha^{N,n}(x)$ respectively. The right graphs are analogous, but we calculated the mean of the MSE over 500 sample replications. The aim is to determine whether the conclusion made after 500 replications remain valid for one sample taken separately, and whether there is a difference for the choice of N when using $\widehat{q}_\alpha^{N,n}(x)$ or $\bar{q}_\alpha^{N,n}(x)$.

We observe in Figures 2 and 3 that the function MSE(N) is at least globally convex. Indeed, for the smoothest case based on 500 replications with $\bar{q}_\alpha^{N,n}(x)$, this function has a unique minimum for each α. Besides, the MSE's increase quite slowly after the minimum and just before, which implies that there is an interval of values for N that could be taken without important impact on the MSE's value. In other words, it does not matter if we choose a value for N not optimal but close to the optimal one, denoted N^* (i.e. $|N - N^*| \leq 5$). We notice that the left graphs, although less smooth (especially without bootstrap), are globally convex and the minimum stays in the same interval.

Another important remark is that N^* appears to be chosen in a symmetric way according to α in these models where the error ε is symmetric (see Figures 2(d) and 3(d)).

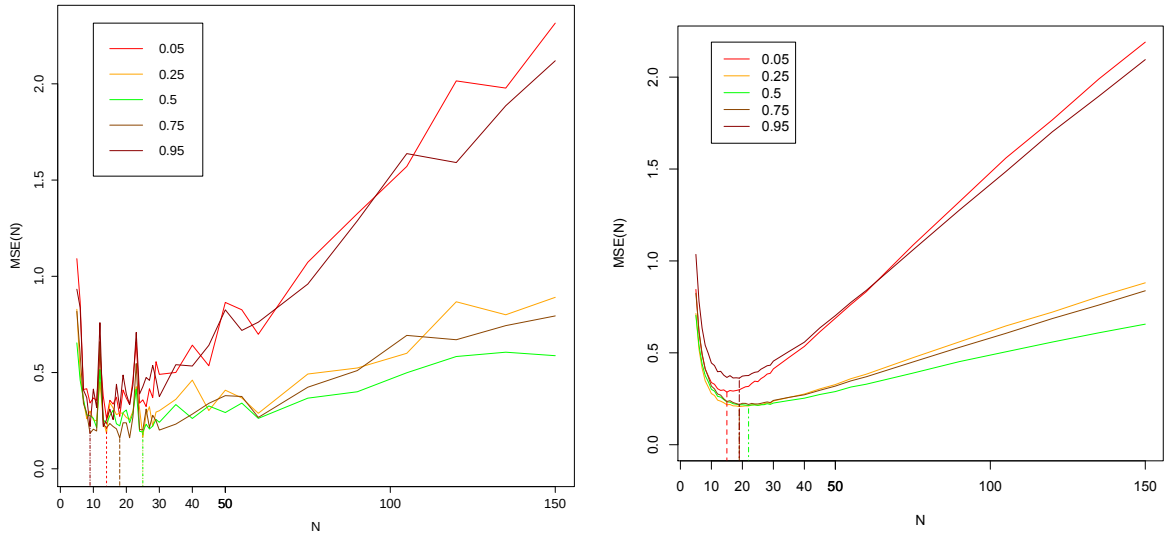


(a) $B = 1$, $m = 1$ (b) $B = 1$, $m = 500$

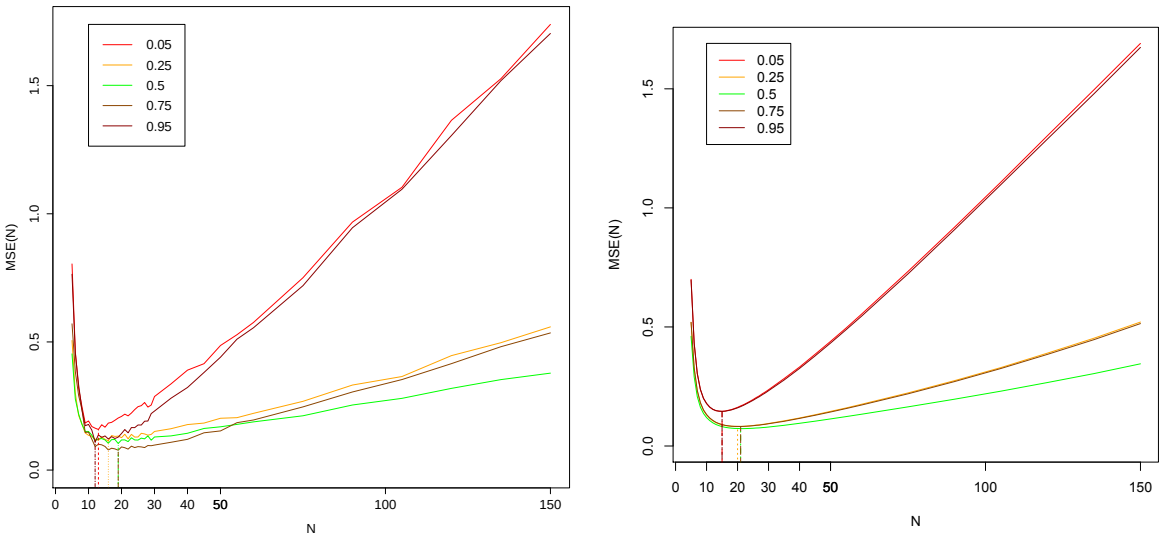


(c) $B = 50$, $m = 1$ (d) $B = 50$, $m = 500$

Figure 2: For $n = 300$, model $(\mathcal{M}1)$ and N from 5 to 30 by step of 1 and from 35 to 150 by step of 5, the four graphs represent the MSE according to N for (a) one sample with $\widehat{q}_\alpha^{N,n}(x)$, (b) 500 replications with $\widehat{q}_\alpha^{N,n}(x)$, (c) one sample with $\bar{q}_\alpha^{N,n}(x)$, (d) 500 replications with $\bar{q}_\alpha^{N,n}(x)$.



(a) $B = 1$, $m = 1$ (b) $B = 1$, $m = 500$



(c) $B = 50$, $m = 1$ (d) $B = 50$, $m = 500$

Figure 3: For $n = 300$, model $(\mathcal{M}2)$ and N from 5 to 30 by step of 1 and from 35 to 150 by step of 5, the four graphs represent the MSE according to N for (a) one sample with $\widehat{q}_\alpha^{N,n}(x)$, (b) 500 replications with $\widehat{q}_\alpha^{N,n}(x)$, (c) one sample with $\bar{q}_\alpha^{N,n}(x)$, (d) 500 replications with $\bar{q}_\alpha^{N,n}(x)$.

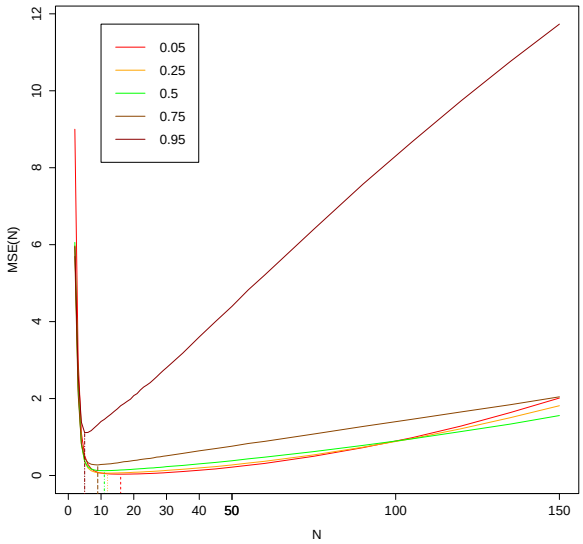


Figure 4: For $n = 300$ and N from 2 to 30 by step of 1 and from 35 to 150 by step of 5, the MSE according to N for 100 replications of model $(\mathcal{M}3)$ with $\bar{q}_\alpha^{N,n}(x)$.

We conclude this section by briefly considering the same model but with an asymmetric error:

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

where $X = 6Z - 3$, with $Z \sim \text{Beta}(0.3, 0.3)$ and $\varepsilon' \sim \chi^2(2)$ are independent. The main effect of this change is the dispersion of the data according to Y. Indeed, the greater Y is, the larger is the dispersion. For this reason, the observations used to estimate the αth conditional quantile curve become more and more spaced when α increases. The optimal value N^* is then expected to depend on α. This feature is checked in Figure 4 for $n = 300$ and 100 replications with the bootstrap-based estimator for the choice $B = 50$. Indeed, the curves of the MSE according to N are still convex but N^* decreases as α increases.

4.2 Practical choice of N based on estimation of the Mean Squared Error

In practical situations, it is not possible to use the previous criterion, based on MSE(N), to determine an optimal N since the user does not know the true underlying model. Thus he cannot calculate the theoretical conditional quantiles $q_\alpha(x_j)$. Therefore, we develop a computational method for this usual practical situation that will consist in estimating the MSE. Let us detail this approach in a general case (i.e. for any underlying model).

We fix a grid $\{x_1, \ldots, x_{\mathcal{N}_x}\}$ of $\mathcal{N}_x$ points between the minimum and the maximum values taken by the X-part of the sample. We "estimate" MSE(N) by a bootstrap-based approach.

Estimate is here in quotes since the aim is not to estimate the value of the MSE as a function of N, but the minimal argument of this function.

For each point of the grid, we estimate the true conditional quantile $q_\alpha(x_j)$ by $\widehat{q}_\alpha^{N,n}(x_j)$ or $\bar{q}_\alpha^{N,n}(x_j)$ using the available sample, as explained before. We also generate $\widetilde{B}$ bootstrap samples of size n from the initial sample (with replacement) allowing us to obtain $\widetilde{B}$ quantization grids with the CLVQ algorithm. We then calculate the estimations $\widehat{q}_\alpha^{N,n}(x_j)$ using each of these grids and the original sample, that we denote $\hat{q}_\alpha^{(\tilde{b})}(x_j)$, for $\tilde{b} = 1, \ldots, \widetilde{B}$. We propose to use the following bootstrap estimate of MSE(N):

$$\widehat{\mathrm{MSE}}(N) = \frac{1}{\mathcal{N}_x} \sum_{j=1}^{\mathcal{N}_x} \left(\frac{1}{\widetilde{B}} \sum_{\tilde{b}=1}^{\widetilde{B}} \left(\hat{q}_\alpha^{(\tilde{b})}(x_j) - q_\alpha'^N(x_j) \right)^2 \right),$$

where $q_\alpha'^N(x_j)$ stands for $\widehat{q}_\alpha^{N,n}(x_j)$ or $\bar{q}_\alpha^{N,n}(x_j)$. Notice that, when taking $q_\alpha'^N(x_j) = \bar{q}_\alpha^{N,n}(x)$, we generate $B + \widetilde{B}$ bootstrap samples of size n from the initial sample: B for the construction of $\bar{q}_\alpha^{N,n}(x_j)$, and $\widetilde{B}$ as explained before. We search for the value of N that minimizes this estimated MSE, and we define

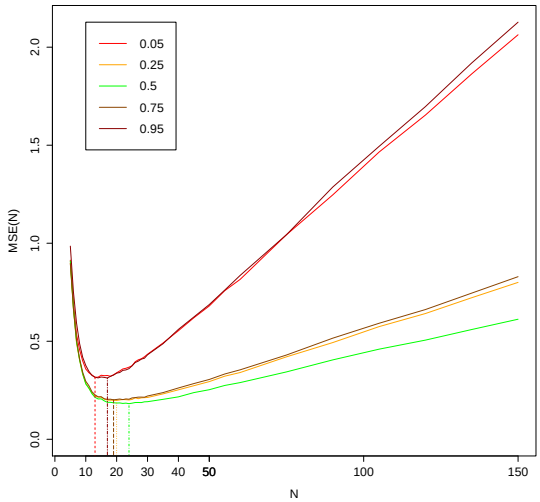
$$\widehat{N}^* = \arg\min_{N \in \mathfrak{N}} \widehat{\mathrm{MSE}}(N).$$

Therefore, we can use this estimated MSE to construct some graphs analogous to Figures 2 and 3. Our aim is to check if the curves of $\widehat{\mathrm{MSE}}(N)$ are convex as their theoretical versions and if they are minimized for the same value of N. Indeed, since we work with simulated models, we know the true MSE's values and which N appears to be the best one. With a real data set, we could not compare our estimated MSE to the true one; this is then important to be sure that $\widehat{\mathrm{MSE}}(N)$, and particularly $\widehat{N}^*$, correspond to the theoretical ones. Figure 5 corresponds to a sample size $n = 300$ generated from model ($\mathcal{M}1$) and is composed of four subfigures :

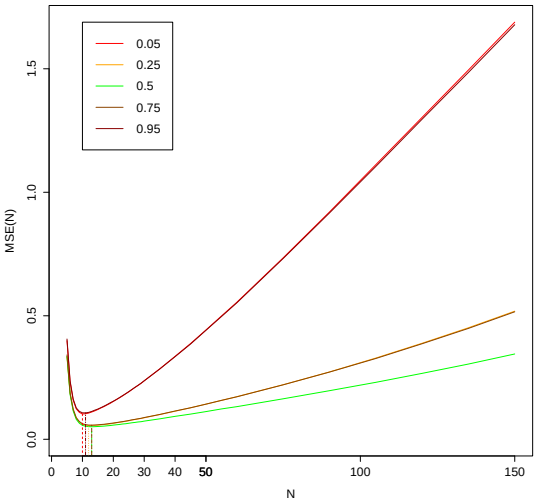
- Figure 5(a) represents the theoretical MSE according to N with $\widehat{q}_\alpha^{N,n}(x)$, with 500 replications,
- Figure 5(b) is analogous with $\bar{q}_\alpha^{N,n}(x)$,
- Figure 5(c) gives the estimated MSE, $\widehat{\mathrm{MSE}}(N)$, for the choices $\widetilde{B} = 30$, $q_\alpha'^N(x_j) = \widehat{q}_\alpha^{N,n}(x_j)$ and with 500 replications,
- Figure 5(d) is similar with $\widehat{\mathrm{MSE}}(N)$ based on the bootstrap version of the estimator, $\bar{q}_\alpha^{N,n}(x_j)$.

Figure 6 is analogous for data generated from model ($\mathcal{M}2$).

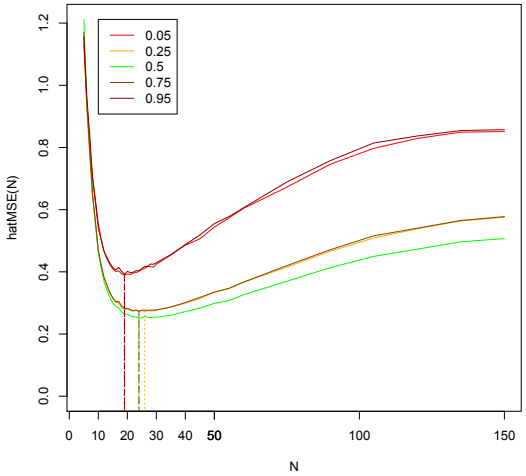
We observe that we obtain very similar optimal values of N in Figures 5(c) and 5(d). Since we noticed previously that $\bar{q}_\alpha^{N,n}(x)$ provides better estimated quantile curves, and since the



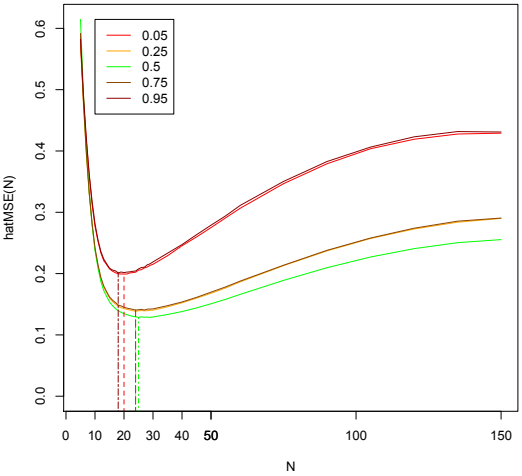
(a) MSE(N), $B = 1$, $m = 500$



(b) MSE(N), $B = 50$, $m = 500$

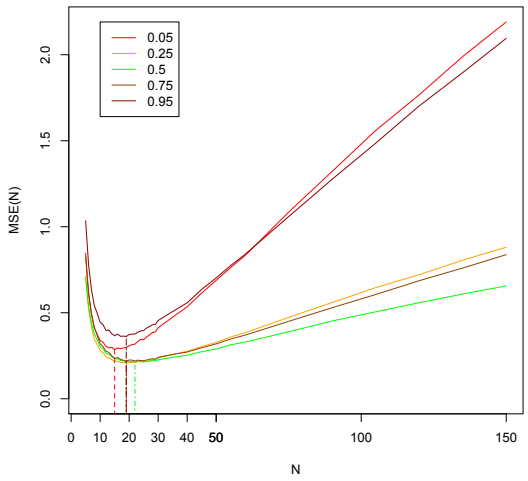


(c) $\widehat{\text{MSE}}(N)$, $B = 1$, $m = 500$

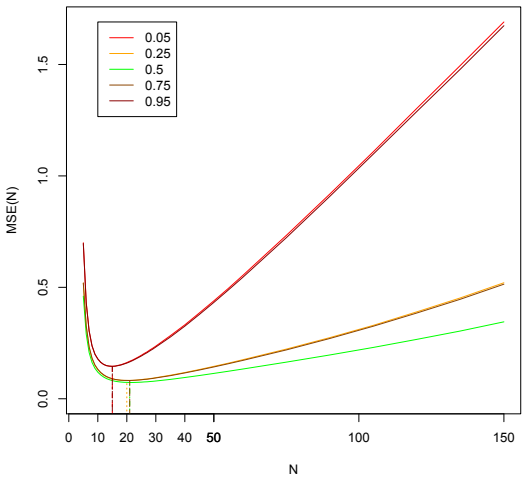


(d) $\widehat{\text{MSE}}(N)$, $B = 50$, $m = 500$

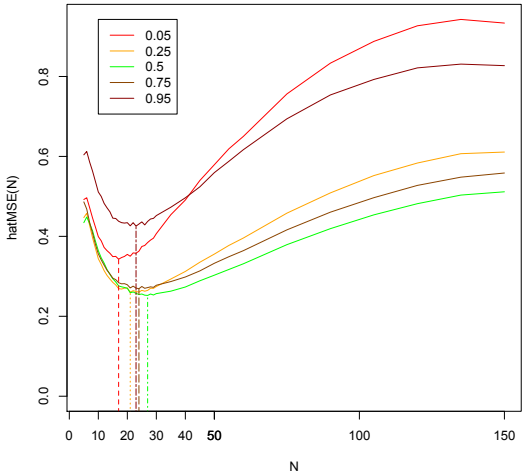
Figure 5: For $n = 300$ and data generated from model ($\mathcal{M}1$): (a) MSE(N) with $\widehat{q}_\alpha^{N,n}(x)$, 500 replications,(b) MSE(N) with $\bar{q}_\alpha^{N,n}(x)$, $B = 50$, 500 replications, (c) $\widehat{\text{MSE}}(N)$ with $\widetilde{B} = 30$ and $q_\alpha'^N(x_j) = \widehat{q}_\alpha^{N,n}(x_j)$, 100 replications, (d) $\widehat{\text{MSE}}(N)$ with $\widetilde{B} = 30$ and $q_\alpha'^N(x_j) = \bar{q}_\alpha^{N,n}(x_j)$, $B = 50$, 100 replications.



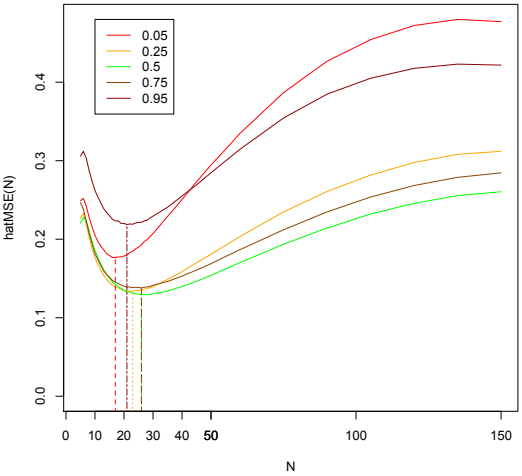
(a) MSE(N), $B = 1$, $m = 500$



(b) MSE(N), $B = 50$, $m = 500$



(c) $\widehat{\text{MSE}}(N)$, $B = 1$, $m = 500$



(d) $\widehat{\text{MSE}}(N)$, $B = 50$, $m = 500$

Figure 6: For $n = 300$ and data generated from model ($\mathcal{M}2$): (a) MSE(N) with $\widehat{q}_\alpha^{N,n}(x)$, 500 replications,(b) MSE(N) with $\bar{q}_\alpha^{N,n}(x)$, $B = 50$, 500 replications, (c) $\widehat{\text{MSE}}(N)$ with $\widetilde{B} = 30$ and $q_\alpha'^N(x_j) = \widehat{q}_\alpha^{N,n}(x_j)$, 500 replications, (d) $\widehat{\text{MSE}}(N)$ with $\widetilde{B} = 30$ and $q_\alpha'^N(x_j) = \bar{q}_\alpha^{N,n}(x_j)$, $B = 50$, 500 replications.

curves in Figure 5(d) are smoother than in Figure 5(c), we would rather use Figure 5(d) for determining $\widehat{N}^*$. This remark is also valid for model ($\mathcal{M}2$) and Figure 6. Let us now compare the values of N^* and $\widehat{N}^*$ for each model. For model ($\mathcal{M}1$), Figure 5 shows us that our criterion tends to over-estimate the value of N^*,with a gap of 10. However, this is not alarming: if we consider the true MSE curves in the top of Figure 5, we observe that the value of the MSE is only slightly larger for $\widehat{N}^*$ than for N^*, especially for middle values of α. Hence, the error do not significantly increase with the choice $\widehat{N}^*$. Concerning model ($\mathcal{M}2$), we see in Figure 6 that our criterion provides really satisfying values for N, since the gap between N^* and $\widehat{N}^*$ is around 5 for each α. This is also interesting to note that the same conclusions are obtained without replications (not graphed here): the only difference concerns the smoothness and the perfect convexity of the curves, but the optimal value of N stays in the same interval.

To conclude, the estimator of the MSE, $\widehat{\mathrm{MSE}}(N)$, allows us to find a value of N, denoted $\widehat{N}^*$, which is really close to the optimal one in our simulated model. This encourages us to recommend this approach for real data sets when we do not know the theoretical quantiles and then we cannot calculate the true MSE to determine N^*.

5 Comparison with existing conditional quantile estimators

The aim of this section is to compare the numerical performance of our estimators with the performance of some other well-known conditional quantile estimators. We introduce in Section 5.1 the alternative estimators considered. This comparison is realized in two ways. First, we analyze in Section 5.3 the boxplots of the mean squared error for each estimator and for different models and sample sizes. We then graphed in Section 5.2 some estimated curves corresponding to each model considered.

5.1 Alternative conditional quantile estimators

We consider two well-known competitors in our simulation study. The first one is the k-NN estimator (see Bhattacharya and Gangopadhyay (1990)). First of all, let us define it more precisely.

According to Bhattacharya and Gangopadhyay (1990), this estimator of $q_\alpha(x)$ is defined as follows. Let $X_i^* = |X_i - x|$ for $i = 1, \ldots, n$ and let $X_{n1}^* < \cdots < X_{nn}^*$ denote the order statistics of $X_1^*, \ldots, X_n^*$ and $Y_{n1}, \ldots, Y_{nn}$ the induced order statistics of $(X_1^*, Y_1), \ldots, (X_n^*, Y_n)$, i.e. $Y_{ni} = Y_j$ if $X_{ni}^* = X_j^*$. For any positive integer $k \leq n$, one will define the k-NN estimator $\widehat{q}_\alpha^k(x) = \widehat{q}_\alpha^{k,n}(x)$ as the $[k\alpha]$th order statistics of $Y_{n1}, \ldots, Y_{nn}$. In a more visual way, the idea is to select the k points of the data such that their X's are the nearest of x, hence the name "k nearest-neighbor",

and to calculate the quantile of order α of their Y's. Indeed, with a null probability, there is no point of the data whose X is equal to x, but it can be very near: one thus selects the k nearest ones. However, the problem of the choice of k follows directly. Since we did not find in the literature an efficient method to select k only based on the data, we will choose k in the sequel by minimizing the mean squared error among a set of values for k, that we will denote k^*. More precisely, taking an equispaced grid $\{x_1, \ldots, x_{\mathcal{N}_x}\}$ on the support of X, we consider

$$\mathrm{MSE}(k) = \frac{1}{\mathcal{N}_x} \sum_{i=1}^{\mathcal{N}_x} \big(q_\alpha(x_i) - \widehat{q}_\alpha^k(x_i)\big)^2,$$

and we choose

$$k^* = \arg\min_{k \in \mathfrak{K}} \mathrm{MSE}(k),$$

for some set $\mathfrak{K}$ of values for k. This method then requires theoretical conditional quantiles and cannot be used in practical situation.

The second competitor is the kernel weighted local linear estimator introduced by Yu and Jones (1998). This estimator is defined as $\widehat{q}_\alpha^{\mathrm{YJ}}(x) = \hat{a}$, with

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

where K is a kernel function, that we will choose as the standard normal density, and where h is the bandwidth. In the sequel, we will choose h according to α as

$$h_\alpha = h_{\mathrm{mean}} \left(\frac{\alpha(1-\alpha)}{\varphi(\Phi^{-1}(\alpha))^2}\right),$$

where φ and Φ are respectively the standard normal density and distribution functions, and where h_{mean} is the optimal choice of h for regression mean estimation, selected thanks to a cross-validation criterion (see Yu and Jones (1998)). We will also consider the local constant version of this estimator. More precisely, it is defined as $\widehat{q}_\alpha^{\mathrm{YJc}}(x) = \hat{a}$, with

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

and the kernel function and the bandwidth are chosen as in the local linear case.

5.2 Comparison of estimated reference curves

For this comparison, we consider the following models:

$(\mathcal{M}1)$ $Y = \frac{1}{5}X^3 + \varepsilon$,
$(\mathcal{M}2)$ $Y = \frac{1}{5}(X')^3 + \varepsilon$,
$(\mathcal{M}3)$ $Y = \frac{1}{5}X^3 + \varepsilon'$,

$(\mathcal{M}4)$ $Y = 1 + 2\tilde{X} + \tilde{\varepsilon}\tilde{X}$,

where the covariates are defined as $X = 6Z - 3$, $\tilde{X} = 10Z$, with $Z \sim \text{Beta}(0.3, 0.3)$, and $X' \sim U(-3, 3)$, and the error terms $\varepsilon \sim \mathcal{N}(0, 1)$, $\varepsilon' \sim \chi^2(2)$ and $\tilde{\varepsilon} \sim \mathcal{N}(0, 0.36)$ are independent of X, X' and $\tilde{X}$.

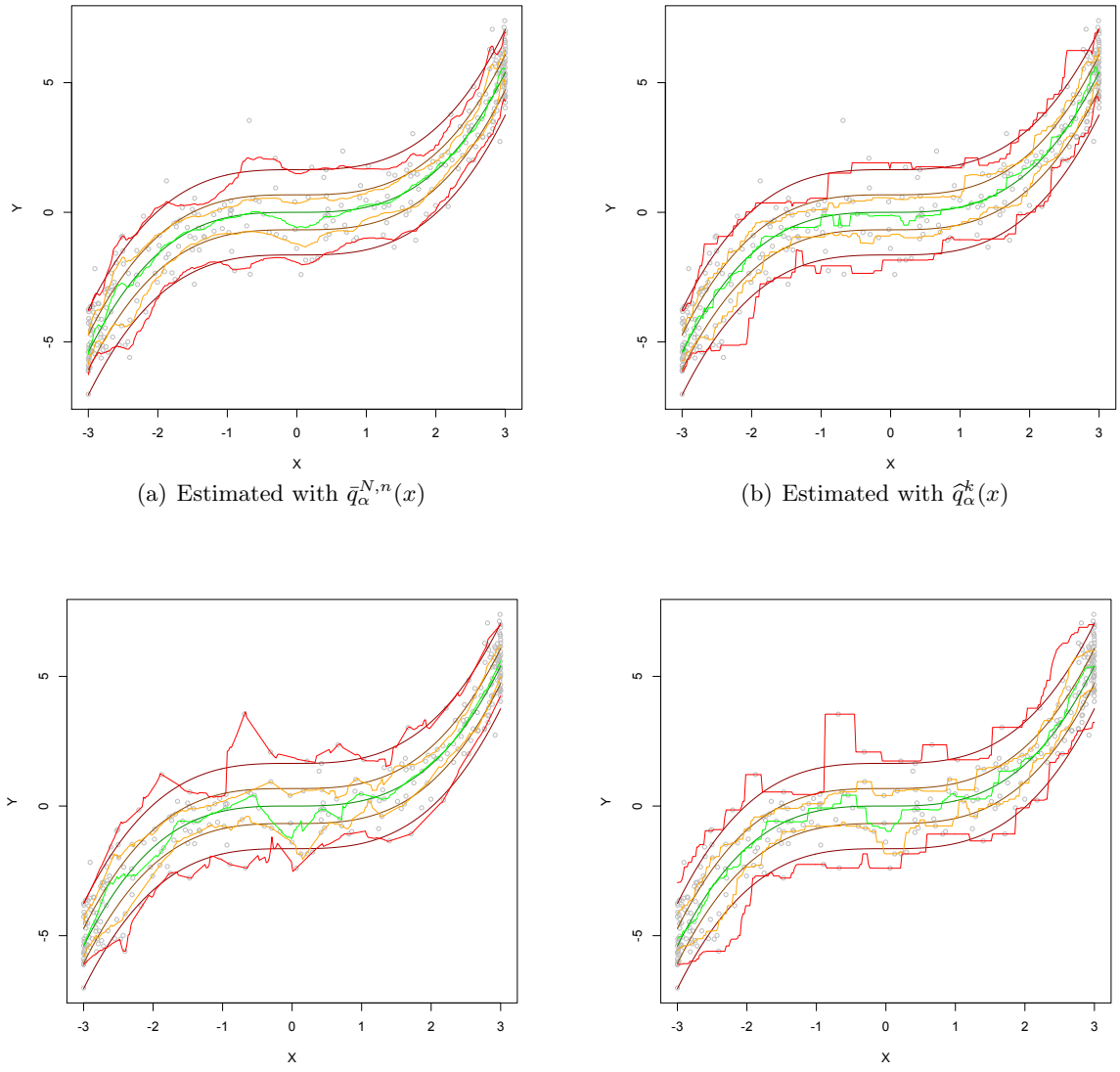
Assume then that we have a sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ generated by one of the previous models. Since we know the underlying model, we can then compare the theoretical conditional quantiles with some estimates. We consider the following estimators:

- $\bar{q}_\alpha^{N,n}(x)$ with the optimal number $\widehat{N}^*$ of quantizers only determined from the data, see Section 4.2,
- $\widehat{q}_\alpha^{k}(x)$, the k nearest-neighbor estimator of Bhattacharya and Gangopadhyay (1990), with the choice of k given in Section 5.1,
- $\widehat{q}_\alpha^{\text{YJ}}(x)$, the local linear estimator of Yu and Jones (1998), with the choice of h detailed in Section 5.1,
- $\widehat{q}_\alpha^{\text{YJc}}(x)$, the local constant estimator of Yu and Jones (1998), with the choice of h detailed in Section 5.1.

We saw in Figure 1 the dominance of $\bar{q}_\alpha^{N,n}(x)$ on $\widehat{q}_\alpha^{N,n}(x)$ thanks to some illustrations. We now proceed similarly by comparing $\bar{q}_\alpha^{N,n}(x)$ with the above-mentioned competitors. We generate samples of size $n = 300$ from each model $(\mathcal{M}1)$ - $(\mathcal{M}4)$ the corresponding estimated conditional quantiles curves are plotted in Figures 7, 8, 9 and 10 respectively. They are all composed of four graphs (one for each kind of estimators) and the tuning parameters used are given for each figure in Tables 2, 3, 4 and 5 respectively.

It is important to notice that the choice of $\widehat{N}^*$ depends on α. This has as an advantage that this choice provides the smallest (bootstrap-based) MSE. However, the major drawback is that the estimated quantiles curves may cross each other, which highly undesirable. Nevertheless, this do not happen in practice since we consider values of α not too close to each other.

We observe in Figure 7 that the local linear and constant estimators $\widehat{q}_\alpha^{\text{YJ}}(x)$ and $\widehat{q}_\alpha^{\text{YJc}}(x)$ suffer from the non uniformity of the distribution of X: less observations are available in the middle of the interval, inducing more irregularities that at the boundaries. Even if the k nearest-neighbor estimator $\widehat{q}_\alpha^{k}(x)$ does not show this difference between the middle and the boundaries of the interval, we observe that our estimator $\bar{q}_\alpha^{N,n}(x)$ provides the smoothest curves in this case.



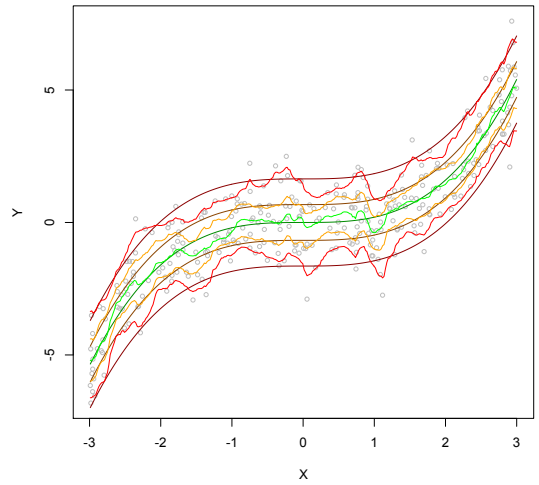
(a) Estimated with $\bar{q}_\alpha^{N,n}(x)$

(b) Estimated with $\widehat{q}_\alpha^{k}(x)$

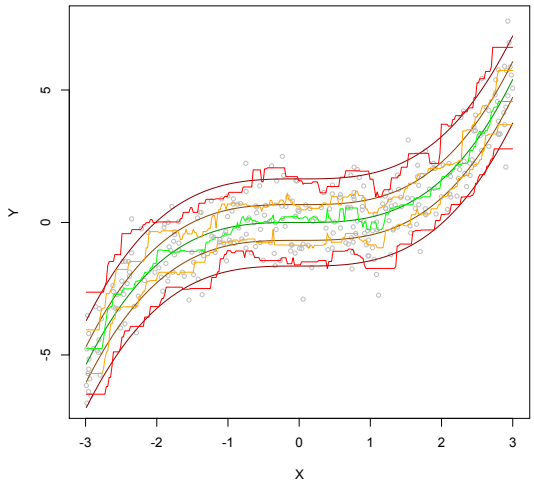
(c) Estimated with $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$

(d) Estimated with $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$

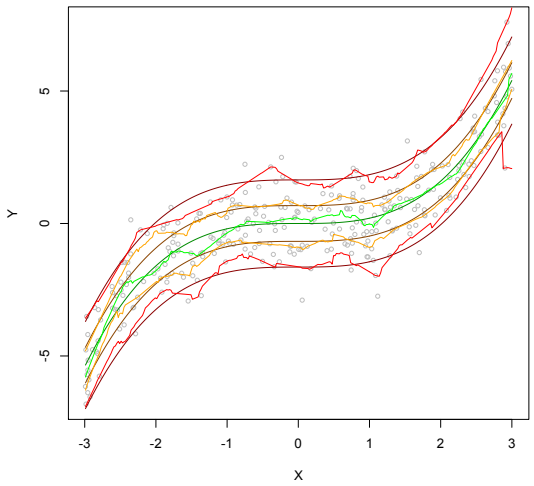
Figure 7: A sample of size $n = 300$ generated according to model ($\mathcal{M}1$). The darker curves represent the theoretical conditional quantile curves and the lighter ones the estimated curves.



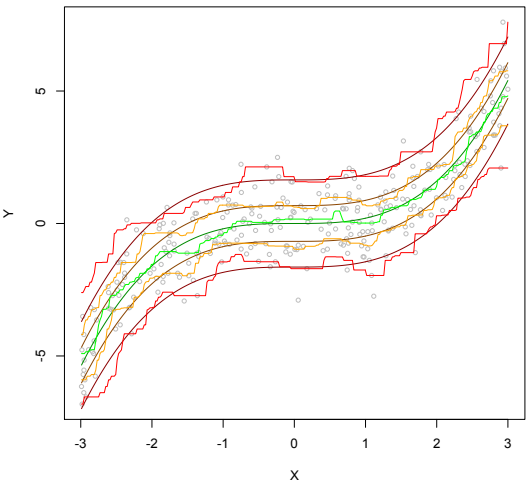
(a) Estimated with $\bar{q}_\alpha^{N,n}(x)$



(b) Estimated with $\widehat{q}_\alpha^{k}(x)$,

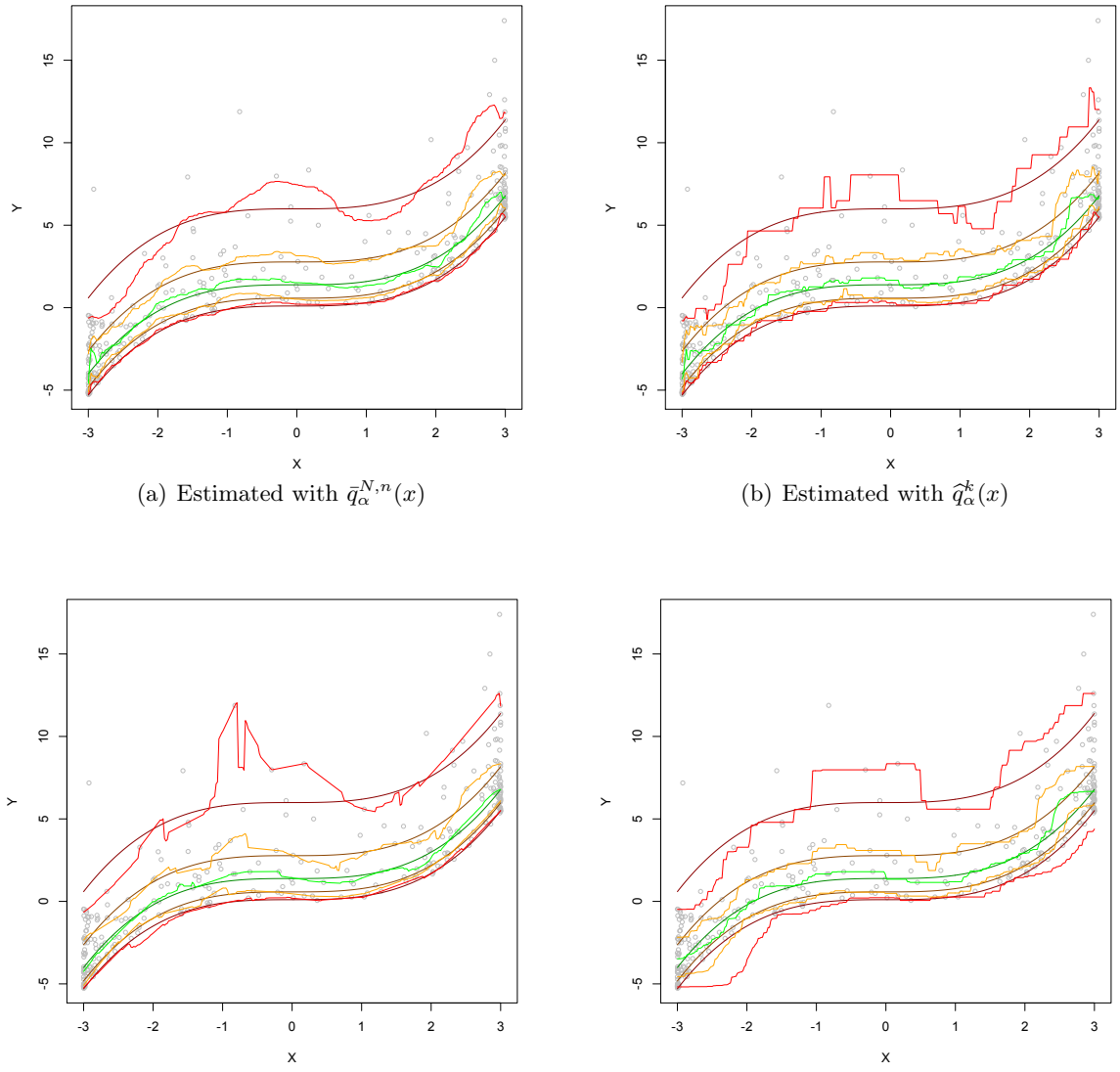


(c) Estimated with $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$



(d) Estimated with $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$

Figure 8: A sample of size $n = 300$ generated according to model ($\mathcal{M}2$). The darker curves represent the theoretical conditional quantile curves and the lighter ones the estimated curves.



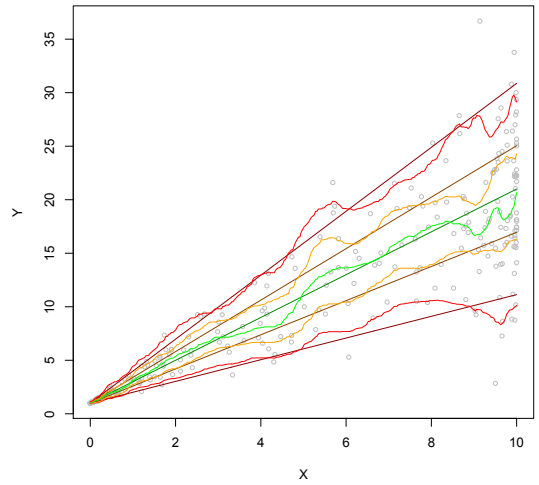
(a) Estimated with $\bar{q}_\alpha^{N,n}(x)$

(b) Estimated with $\widehat{q}_\alpha^{k}(x)$

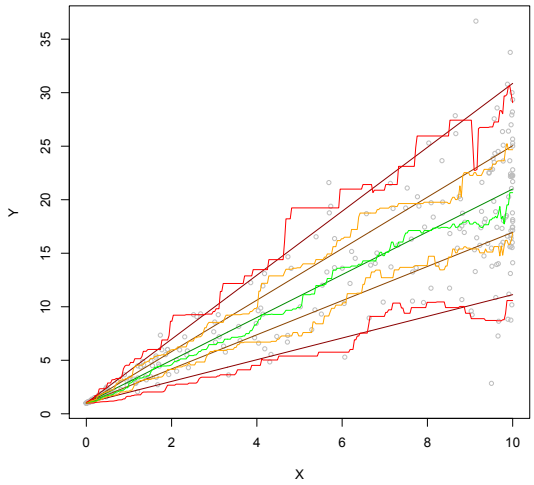
(c) Estimated with $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$

(d) Estimated with $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$

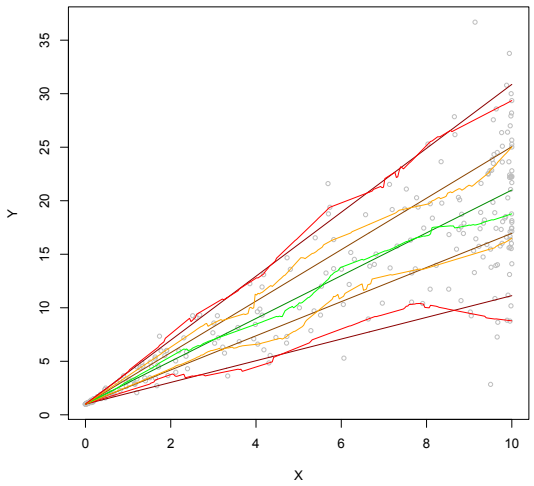
Figure 9: A sample of size $n = 300$ generated according to model ($\mathcal{M}3$). The darker curves represent the theoretical conditional quantile curves and the lighter ones the estimated curves.



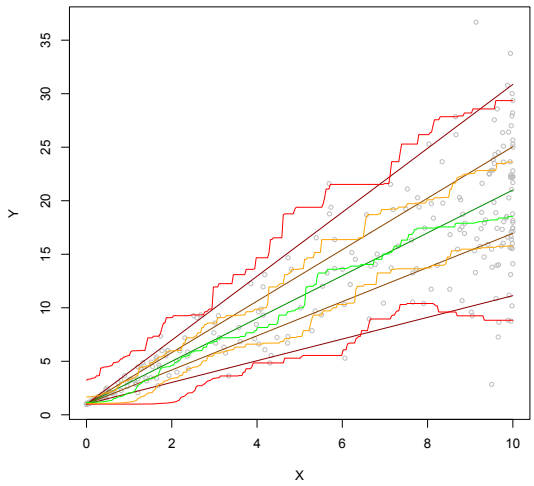
(a) Estimated with $\bar{q}_{\alpha}^{N,n}(x)$, $\widehat{N}^* = 15$



(b) Estimated with $\widehat{q}_{\alpha}^{k}(x)$, $k^* = 40$



(c) Estimated with $\widehat{q}_{\alpha}^{\mathrm{YJ}}(x)$



(d) Estimated with $\widehat{q}_{\alpha}^{\mathrm{YJc}}(x)$

Figure 10: A sample of size $n = 300$ generated according to model ($\mathcal{M}4$). The darker curves represent the theoretical conditional quantile curves and the lighter ones the estimated curves.

This graphical dominance will be checked in the boxplot comparison of the following section (see Figure 11).

In Figure 8, the covariate is uniformly distributed, which induces more smoothness for $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$ and $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ than in previous cases. Here, $\bar{q}_\alpha^{N,n}(x)$ and $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ seem to provide the most desirable curves.

Figure 9 shows us that the estimated curves generally fit very well the theoretical ones for small α, but some irregularities tend to stay for large α (more significant for $\alpha = 0.95$). The difference in the quality of estimation according to α is observed for each estimators, however the curves obtained with $\bar{q}_\alpha^{N,n}(x)$ appear smoother.

For Figure 10, the conclusions are analogous to the ones of Figure 9: the curves obtained with $\bar{q}_\alpha^{N,n}(x)$ and $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ seem to be smoother.

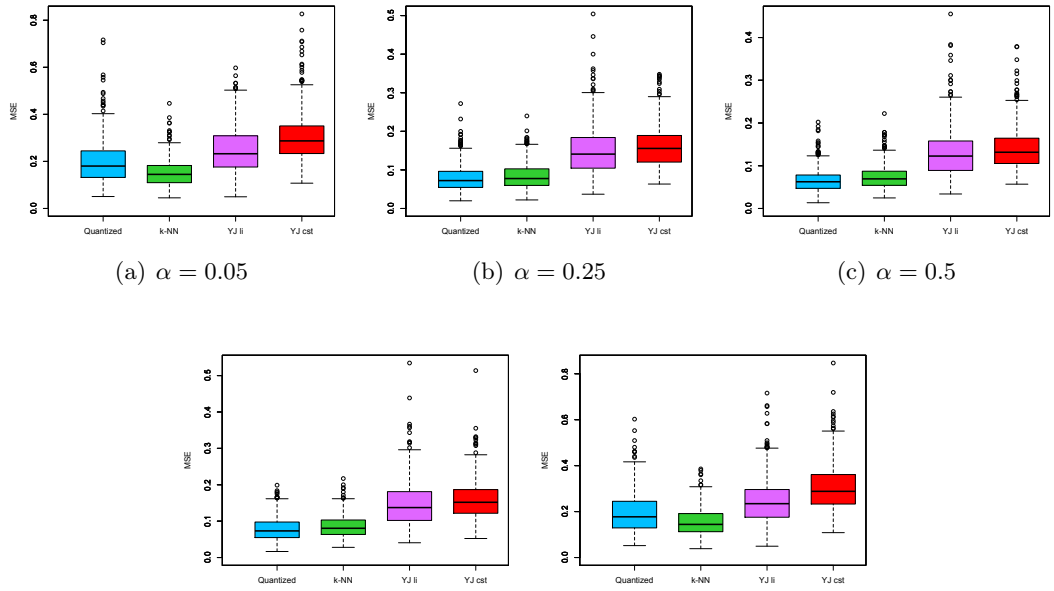
5.3 Comparison of the Mean Squared Error thanks to boxplots

We generate 500 samples for each model ($\mathcal{M}1$)-($\mathcal{M}4$) and for various sample sizes $n = 300$, $n = 500$ and $n = 1,000$. For each simulated sample, we calculate the estimates $\bar{q}_\alpha^{N,n}(x)$, $\widehat{q}_\alpha^{k}(x)$, $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ and $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$ for some x on a fixed grid and the corresponding MSE.

We represent in Figures 11, 12, 13 and 14 the boxplots of the MSE according to these four estimators for each of these models and $n = 300$. We observe in each figure that $\bar{q}_\alpha^{N,n}(x)$ provides a better quality of estimation that $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$. The same conclusions are generally valid for $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ in Figures 11 and 12, when the covariate is not uniformly distributed (except for $\alpha = 0.05$). In contrast in Figure 14, all methods appear equivalent.

The main advantage of quantization is that the radius of a Voronoi cell (i.e. the larger distance between a point of the grid and any point projected on it) is adaptive: when there are many points in a region, the number of quantizers is more important in this region than in other less dense regions. The radius of the cells can be seen as a data-adaptive bandwidth parameter. It is then not surprising that $\bar{q}_\alpha^{N,n}(x)$ outperforms $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ in a situation where this feature appears, since the bandwidth h is the same for any x in $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$. In this case, the MSE's for $\bar{q}_\alpha^{N,n}(x)$ are significantly smaller than for $\widehat{q}_\alpha^{\mathrm{YJ}}(x)$ and $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$ (see Figure 11).

Concerning $\widehat{q}_\alpha^{k}(x)$ in these models, the dominance of $\bar{q}_\alpha^{N,n}(x)$ is also often observed but tends to depend on α. Even if the MSE's for $\widehat{q}_\alpha^{k}(x)$ are sometimes smaller than to the ones of $\bar{q}_{\alpha,n}^{X}(x)$ (for extreme values of α), we have to keep in mind that the selection method for k is not data-based and then unfeasible in practice.



(a) $\alpha = 0.05$ (b) $\alpha = 0.25$ (c) $\alpha = 0.5$

(d) $\alpha = 0.75$ (e) $\alpha = 0.95$

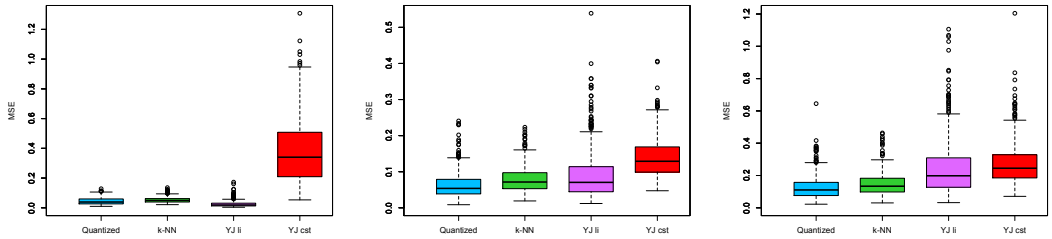
Figure 11: For 500 replications of sample of size $n = 300$ from model ($\mathcal{M}1$), the boxplots of the MSE in the estimation of the conditional quantile curves for different values of α: in blue, with $\bar{q}_\alpha^{N,n}(x)$, in green, with $\widehat{q}_\alpha^{k}(x)$, in purple, with $\widehat{q}_\alpha^{YJ}(x)$ and in red, with $\widehat{q}_\alpha^{\mathrm{YJc}}(x)$.

	$\bar{q}_\alpha^{N,n}(x)$ (step=1)	$\bar{q}_\alpha^{N,n}(x)$ (step=5)	$\widehat{q}_\alpha^{k}(x)$	$\widehat{q}_\alpha^{\mathrm{YJ}}(x)$	$\widehat{q}_\alpha^{\mathrm{YJc}}(x)$
CPU	11108.278	2509.359	546.093	1312.166	1152.017
$\sum_\alpha$ median	0.552	0.566	0.516	0.868	1.014

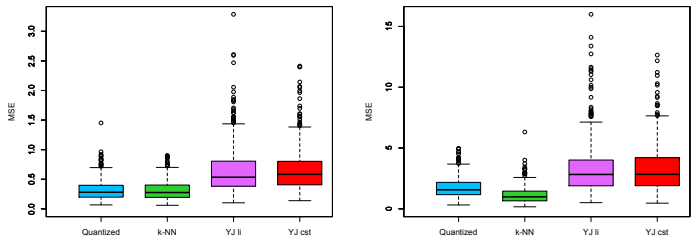
Table 1: This table provides the computation time of the 500 replications of Figure 11 for each estimator and the corresponding sum on α of the median values.

Figures 15 and 16 correspond to Figure 11 for sample sizes $n = 500$ and $n = 1,000$ respectively. Again, $\bar{q}_\alpha^{N,n}(x)$ provides smaller MSE and outperforms other methods for middle values of α, and only the constant and local linear estimators for extreme values of α.

It is also important to focus on computation times. We gather in Table 1 the computation time needed to realize the 500 replications of Figure 11 for each estimator. We also calculate the corresponding sum of the median for each α to compare computation time and quality of estimation. Since the choice of the grid $\mathfrak{N}$ plays a important role in the selection of $\widehat{N}^*$, we give the corresponding values for two different grids $\mathfrak{N}$: a more accurate grid of values for N where we test the values between 5 and 30 by step of one, and the one used to obtain the given boxplots where we test them by step of five. This accuracy increases hugely the computation time for

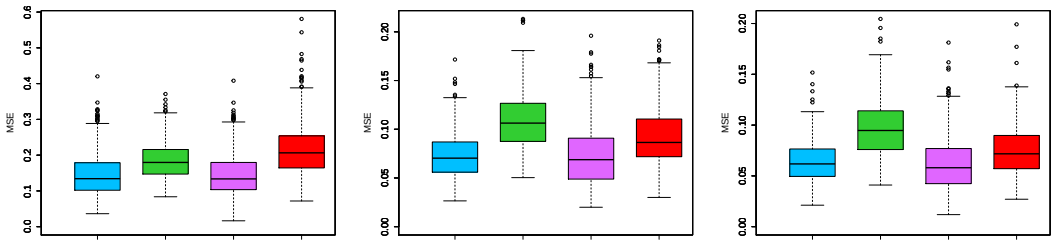


(a) $\alpha = 0.05$ (b) $\alpha = 0.25$ (c) $\alpha = 0.5$

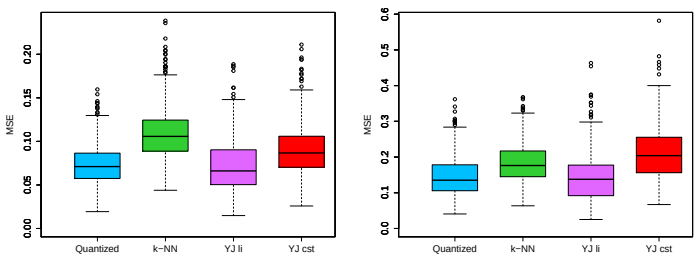


(d) $\alpha = 0.75$ (e) $\alpha = 0.95$

Figure 12: Idem as Figure 11, but for model ($\mathcal{M}3$).

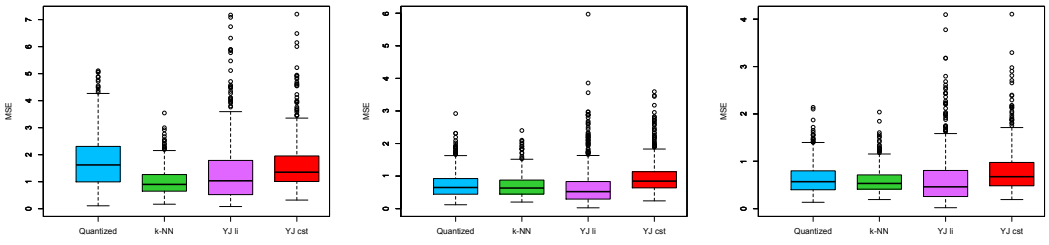


(a) $\alpha = 0.05$ (b) $\alpha = 0.25$ (c) $\alpha = 0.5$

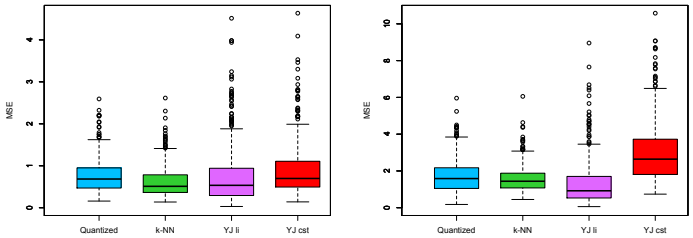


(d) $\alpha = 0.75$ (e) $\alpha = 0.95$

Figure 13: Idem as Figure 11, but for model ($\mathcal{M}2$).

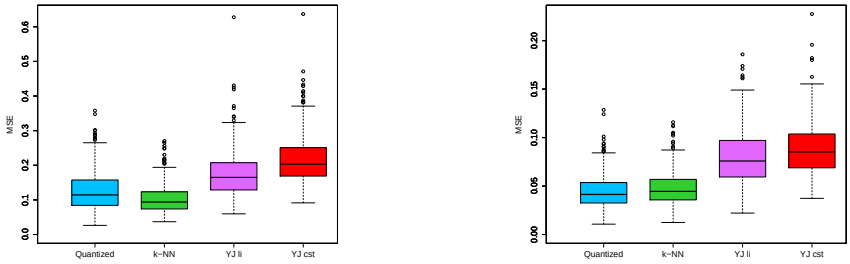


(a) $\alpha = 0.05$ (b) $\alpha = 0.25$ (c) $\alpha = 0.5$



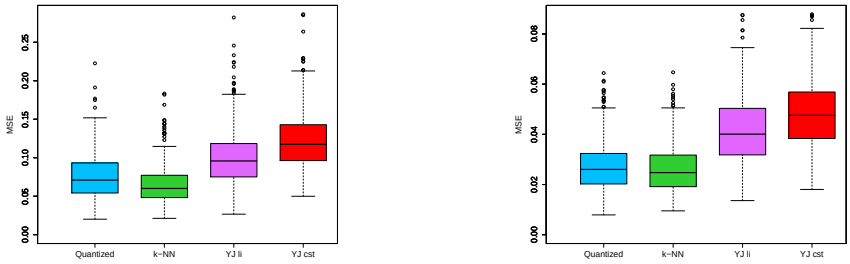
(d) $\alpha = 0.75$ (e) $\alpha = 0.95$

Figure 14: Idem as Figure 11, but for model ($\mathcal{M}4$).



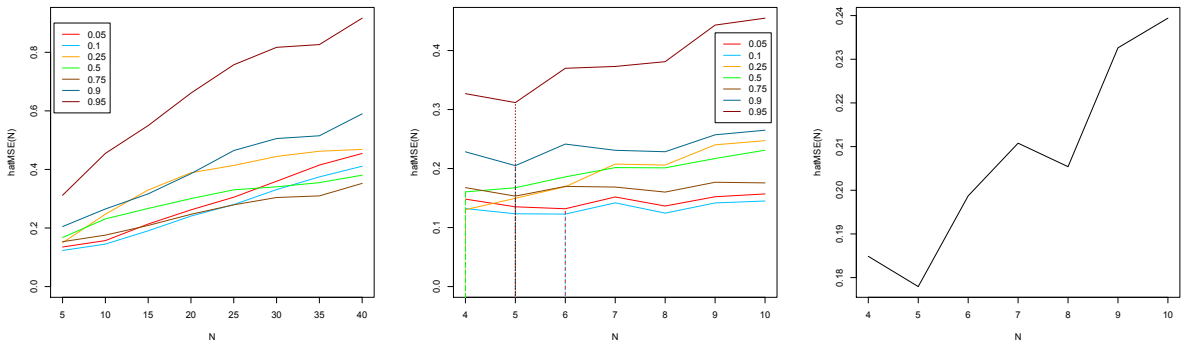
(a) $\alpha = 0.05$ (b) $\alpha = 0.5$

Figure 15: Idem as Figure 11, but for $n = 500$.



(a) $\alpha = 0.05$ (b) $\alpha = 0.5$

Figure 16: Idem as Figure 11, but for $n = 1000$.



(a) N from 5 to 40 by step of 5 (b) N from 4 to 10 by step of 1 (c) Mean of α

Figure 17: Selecting method for N. (a) $\widehat{\mathrm{MSE}}(N)$ for a first grid of values of N for different α, (b) $\widehat{\mathrm{MSE}}(N)$ for a best-adapted grid of values of N for different α, (c) the mean for α of $\widehat{\mathrm{MSE}}(N)$.

only a slight improvement (in terms of reduction of the sum of the median of the errors).

6 Application on real data set

In this section, we illustrate the behavior of $\bar{q}_\alpha^{N,n}(x)$ on a real data set of size $n = 298$. This data set corresponds to the serum concentration, in grams per liter, of immunoglobulin-G in children aged from 6 months to 6 years which was kindly sent to us by Dr. K. Yu. It was already investigated in Yu and Jones (1998). They considered two estimators, a kernel weighted local linear estimator and a double kernel estimator. The reader can refer to this paper where the corresponding (smoothed) estimated quantiles curves are graphed.

We construct estimated conditional quantile curves as follows. The first step consists in choosing the optimal number $\widehat{N}^*$ of quantizers (optimal is here in the sense that it minimizes the error $\widehat{\mathrm{MSE}}(N)$). As above-mentioned, we cannot compare the estimated conditional quantiles to the true ones since we do not know them. We then use the estimated MSE that we introduced in Section 4.2, with $q_\alpha'^N(x) = \bar{q}_\alpha^{N,n}(x)$.

We calculate $\widehat{\mathrm{MSE}}(N)$ for various values of N and minimize it for different values of α taken separately. We choose α= 0.05, 0.1, 0.25, 0.5, 0.75, 0.9 and 0.95, and we evaluate the conditional quantiles at 300 equispaced points between the minimum and maximum values of the X-part of the sample. This selection of N is realized in two steps:

(S1) We first calculate the estimated MSE for N between 5 and 40 by step of 5. In order to get smooth $\widehat{\mathrm{MSE}}(N)$ curves, we realize 10 replications of this procedure (the sample is of course the same but the quantization grids vary at each replication). Figure 17(a) provides

the resulting curves. The values 5 then 10 are optimal. We conclude that the testing values for N were not well chosen.

(S2) We then calculate $\widehat{\mathrm{MSE}}(N)$ for N between 4 and 10. We represent it in Figure 17(b). It appears that $\widehat{N}^*$ depends on α but $\widehat{N}^* = 5$ or 6 show up for the majority of α's.

As already mentioned, choosing N according to α could be problematic here since the aim of this estimation is to plot the obtained curves and different choices of N could provide crossing curves. Two types of N-selection can then be considered. The first consists in taking the risk of having crossing curves (which rarely happens) and choosing different N for each α's. The second one consists in summing up the estimated error for each value of α (with N fixed), and observe for which value of N this sum is minimal. This procedure is represented in Figure 17(c). The value $N = 5$ is then optimal.

It is important to specify the number B of bootstrap replications we used. A large value of B does not change the optimal value of N, but slightly smoothes the curves. For this reason, we use $B = 50$ for determining N to save time, but we will use larger B when plotting the curves.

Figure 18 gathers the two possible selection for N, for this two values of B. The top figures are obtained with $\widehat{N}^* = 5$ while $\widehat{N}^*$ depends on α in the bottom ones. These graphs are quite similar. Since the left panel still contains some irregularities, we increase the number of bootstrap replications to get the right panel.

7 Extensions

We have realized a simulation study to investigate the numerical performances of the estimator of conditional quantile based on optimal quantization, introduced in Charlier et al. (2014). This study allowed us to determine an efficient data-based selecting method for the number N of quantizers. We then compared our estimators with two alternative non parametric estimators and we observed that the proposed estimator $\bar{q}_\alpha^{N,n}(x)$ generally outperforms its competitors where the covariate is not uniformly distributed. In the considered real data set, we also obtained satisfying estimated conditional quantile curves using the above-mentioned estimator. A R package was developed, named `QuantifQuantile` and is available on the CRAN. We will conclude this paper with some possible extensions of this work.

7.1 Case of $d > 1$

The simulation study in this paper was realized for the particular case $d = 1$. However, our theoretical results were proven for any d in Charlier et al. (2014). We can then extend the

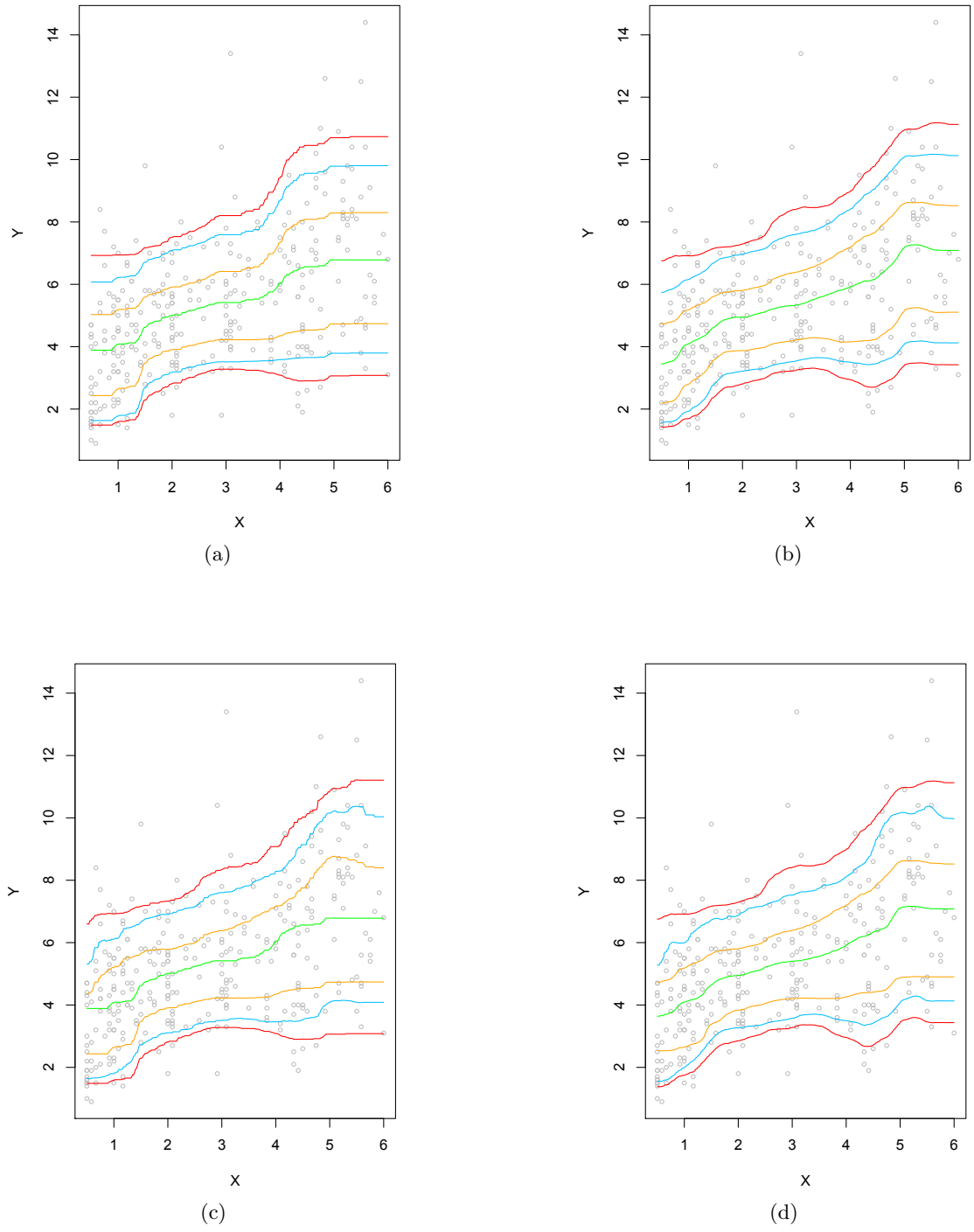


Figure 18: For the Immunoglobulin-G data, (a) Quantile curves obtained with $\bar{q}_\alpha^{N,n}(x)$, $\widehat{N}^* = 5$, $B = 50$, (b) Idem with $B = 500$, (c) Quantile curves obtained with $\bar{q}_\alpha^{N,n}(x)$, $B = 50$ and N depending on α, (d) Idem with $B = 500$.

results of this paper for $d > 1$. Our method still works well, but the computation time increases. Figure 19 shows us what kind of quantile surfaces we can obtain for a sample size $n = 1,000$. Generated from the model

$$(\mathcal{M}5)\ \ Y = \frac{1}{4}\sum_{i=1}^{d} X_i^3 + \varepsilon,$$

where $X \sim U\big([-3,3]^d\big)$, X_i denotes the ith component of X and $\varepsilon \sim \mathcal{N}(0,9)$ is independent of X. The left panel of Figure 19 represents the theoretical conditional quantile surfaces with different views while the right panel of Figure 19 corresponds to the estimated conditional quantile surface with the choice $\widehat{N}^* = 140$ and $B = 50$. The value of $\widehat{N}^*$ was determined by minimizing $\widehat{\mathrm{MSE}}(N)$.

These figures correspond to the case $d = 2$. We can of course estimate conditional quantiles of order α when $d > 2$. However, graphical representation is no longer available. For this reason, we do not illustrate our method for $d > 2$.

7.2 Quantization in Y

Optimal quantization allows us to construct discrete versions of any continuous law. In our work, we used this tool to quantize the covariate X. It is then natural to wonder why not quantizing also the response variable Y and moving entirely to a distcrete framework. This can be done by defining the following approximation and estimator of conditional quantiles. Let us write $\delta^N = (\tilde{y}_1, \ldots, \tilde{y}_N)$ for an optimal N-grid for Y and $\hat{\delta}^{N,n} = (\hat{y}_1^{N,n}, \ldots, \hat{y}_N^{N,n})$ for the grid provided by the CLVQ algorithm. We define

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

$$\widehat{q}_\alpha^{Y,N,n}(x) = \arg\min_{a\in\mathbb{R}} \sum_{i=1}^{N} \rho_\alpha(\widehat{Y}_i^N - a)\#\{j : (\widehat{X}_j^N, \widehat{Y}_j^N) = (\hat{x}^N, \hat{y}_i^{N,n})\},$$

where $\widetilde{Y}^N = \mathrm{Proj}_{\delta^N} Y$ and $\widehat{Y}_i^N = \widehat{Y}_i^{N,n} = \mathrm{Proj}_{\hat{\delta}^{N,n}}(Y_i)$. We can also define a bootstrap version of this estimator following the exact same procedure as when we do not quantize Y.

We did not study them in Charlier et al. (2014) and in this paper since we observe from simulations that $\bar{q}_\alpha^{N,n}(x)$ outperforms significantly its Y-quantized corresponding version. Indeed, Figure 20 represents the boxplots of the mean squared error for these two estimators for 500 replications of samples of size $n = 300$ generated from model $(\mathcal{M}2)$.

Note that the number N of quantizers is the same for X and Y in the construction of $\widehat{q}_\alpha^{Y,N,n}(x)$. It is then interesting to wonder whether we could obtain better results with the former if the number of quantizers is different for X and Y, that we will denote N_X and N_Y.

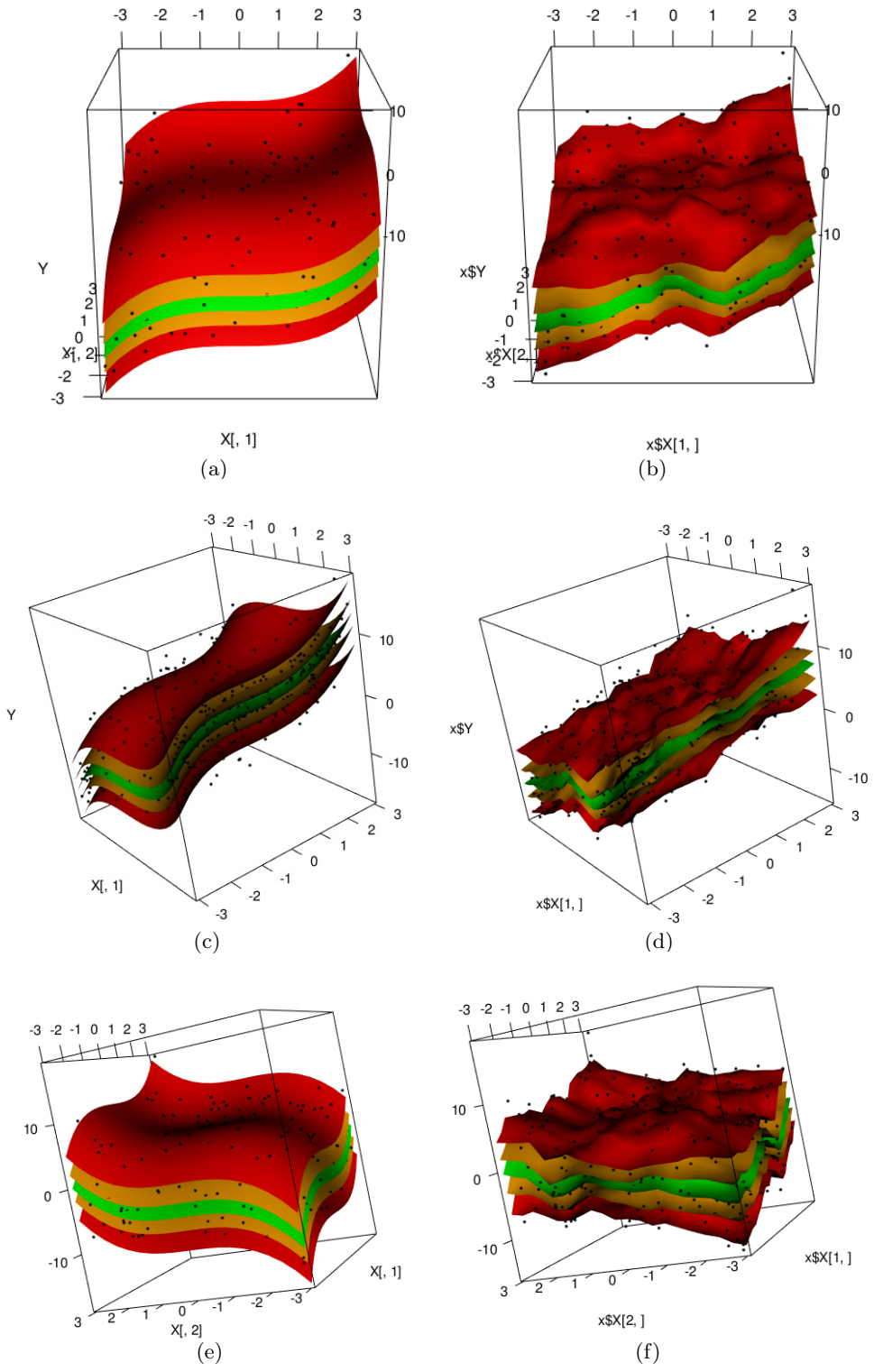


Figure 19: For a sample of size $n = 1,000$ from model ($\mathcal{M}5$), the left panel corresponds to the theoretical conditional quantile surfaces with different views, and the right panel the corresponding estimated conditional quantile surfaces $x \mapsto \bar{q}_\alpha^{N,n}(x)$ with $B = 50$ and $\widehat{N}^* = 140$.

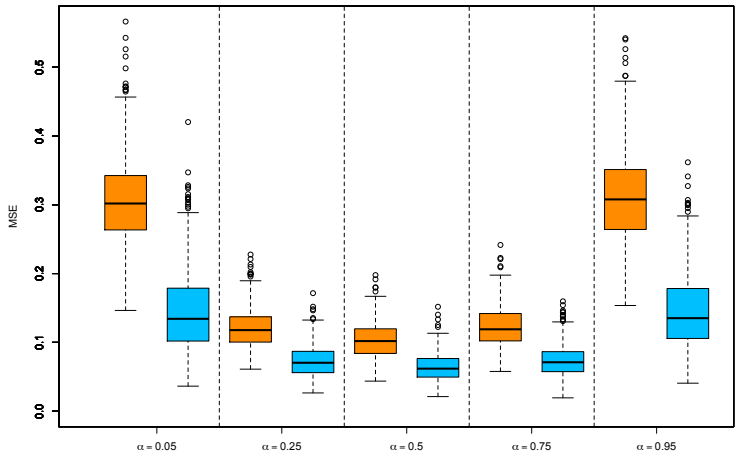


Figure 20: For 500 replications of sample of size $n = 300$ from model ($\mathcal{M}2$), the boxplots of the MSE in the estimation of conditional quantiles curves: in orange, the bootstrap-based version of $\widehat{q}_\alpha^{Y,N,n}(x)$ and in blue, $\bar{q}_\alpha^{N,n}(x)$.

For this reason, we calculate the MSE for different values of N_X and N_Y and we minimize the two-dimensional function $\mathrm{MSE}(N_X, N_Y)$. We take N_X between 5 and 80 and N_Y between 5 and 250, by step of five. In order to get smoother surface, we realize 5 replications. The corresponding three dimensional graph is represented in Figure 21 for five values of α. For each fixed N_Y, we calculate the optimal value for N_X and we plot the curve of $N_X^*(N_Y)$: it is the white line in each graph. This figure shows us that the optimal number of quantizers for X stay in the same interval for any values of N_Y and α (between 25 and 35). However, we cannot find an optimal value for N_Y (in the sense that the MSE is minimum): the greater we take N_Y, the smaller the MSE is. This fact encourages us not to quantize in Y, in other words to prefer $\bar{q}_\alpha^{N,n}(x)$, as mentioned previously.

A Appendix

References

Azaïs, R., A. Gégout-Petit, and J. Saracco (2012). Optimal quantization applied to sliced inverse regression. *J. Statist. Plann. Inference 142*(2), 481–492.

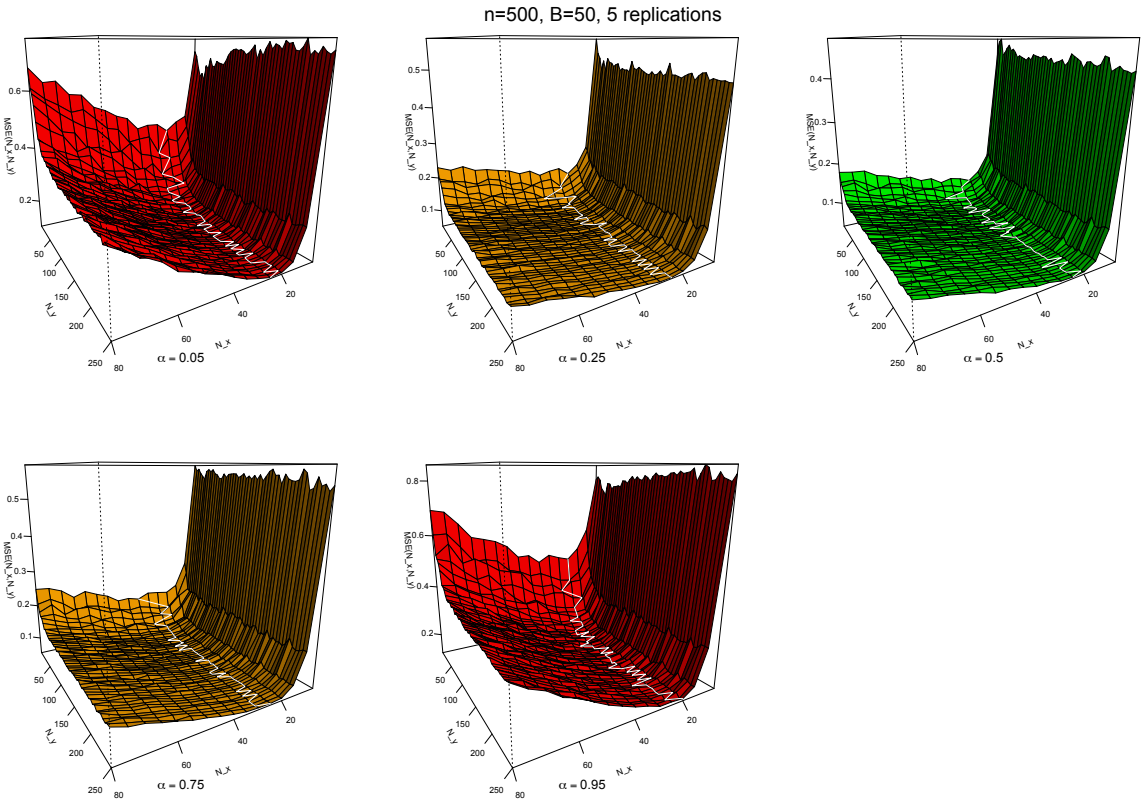


Figure 21: For a sample of size $n = 500$ from model ($\mathcal{M}2$), the surface of the MSE according to N_X and N_Y for five values of α, for the estimator $\bar{q}_\alpha^{N,n}(x)$, B=50.

	$\alpha = 0.05$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.95$
$\widehat{N}^*$	20	21	20	22	22
k^*	25	32	30	28	27
h_{YJ}	0.202	0.170	0.164	0.170	0.202
h_{YJcst}	0.202	0.170	0.164	0.170	0.202

Table 2: For each value of α, the tuning parameters selected by each method and used to construct graphs of Figure 7.

	$\alpha = 0.05$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.95$
$\widehat{N}^*$	26	26	26	26	26
k^*	29	25	25	20	25
h_{YJ}	0.229	0.192	0.186	0.192	0.229
h_{YJcst}	0.229	0.192	0.186	0.192	0.229

Table 3: For each value of α, the tuning parameters selected by each method and used to construct graphs of Figure 8.

	$\alpha = 0.05$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.95$
$\widehat{N}^*$	20	20	20	13	9
k^*	14	19	37	35	37
h_{YJ}	0.364	0.306	0.296	0.306	0.364
h_{YJcst}	0.364	0.306	0.296	0.306	0.364

Table 4: For each value of α, the tuning parameters selected by each method and used to construct graphs of Figure 9.

	$\alpha = 0.05$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.95$
$\widehat{N}^*$	11	11	17	15	17
k^*	49	31	42	42	26
h_{YJ}	0.809	0.679	0.657	0.679	0.809
h_{YJcst}	0.809	0.679	0.657	0.679	0.809

Table 5: For each value of α, the tuning parameters selected by each method and used to construct graphs of Figure 10.

Bally, V., G. Pagès, and J. Printems (2005). A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance 15*(1), 119–168.

Bhattacharya, P. K. and A. K. Gangopadhyay (1990). Kernel and nearest-neighbor estimation

of a conditional quantile. *Ann. Statist. 18*(3), 1400–1415.

Charlier, I., D. Paindaveine, and J. Saracco (2014). Conditional quantile estimators through optimal quantization. *Submitted*.

de Saporta, B., F. Dufour, and K. Gonzalez (2010). Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab. 20*(5), 1607–1637.

Fischer, A. (2010). Quantization and clustering with bregman divergences. *Journal of Multivariate Analysis 101*, 2207–2221.

Fischer, A. (2014). Deux méthodes d'apprentissage non supervisé : synthèse sur la méthode des centres mobiles et présentation des courbes principales. *J. Soc. Fr. Stat. 155*(2), 2–35.

Graf, S. and H. Luschgy (2000). *Foundations of quantization for probability distributions*, Volume 1730 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag.

Koenker, R. and G. Bassett, Jr. (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Pagès, G. (1998). A space quantization method for numerical integration. *J. Comput. Appl. Math. 89*(1), 1–38.

Pagès, G., H. Pham, and J. Printems (2004a). An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems. *Stoch. Dyn. 4*(4), 501–545.

Pagès, G., H. Pham, and J. Printems (2004b). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pp. 253–297. Boston, MA: Birkhäuser Boston.

Pagès, G. and J. Printems (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl. 9*(2), 135–165.

Yu, K. and M. C. Jones (1998). Local linear quantile regression. *J. Amer. Statist. Assoc. 93*(441), 228–237.