Multiple-Output Regression through Optimal Quantization

Isabelle Charlier
SBS-EM, ECARES and Département de Mathématique Université libre de Bruxelles
and Université de Bordeaux

Davy Paindaveine
SBS-EM, ECARES and Département de Mathématique, Université libre de Bruxelles

Jérôme Saracco
Université de Bordeaux

April 2016

ECARES working paper 2016-18
Multiple-Output Quantile Regression through Optimal Quantization

Isabelle Charlier\(^{(1,2)}\)∗ Davy Paindaveine\(^{(1)}\)† Jérôme Saracco\(^{(2)}\)

April 3, 2016

\(^{(1)}\) Université Libre de Bruxelles, ECARES and Département de Mathématique, 50, Avenue F.D. Roosevelt, CP114/04, B-1050, Brussels, Belgium. ischarli@ulb.ac.be, dpaindav@ulb.ac.be

\(^{(2)}\) Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251 and INRIA Bordeaux Sud-Ouest, CQFD team, 351 Cours de la Libération, 33405 Talence, France. Jerome.Saracco@math.u-bordeaux1.fr

Abstract

Charlier et al. (2015a,b) developed a new nonparametric quantile regression method based on the concept of optimal quantization and showed that the resulting estimators often dominate their classical, kernel-type, competitors. The construction, however, remains limited to single-output quantile regression. In the present work, we therefore extend the quantization-based quantile regression method to the multiple-output context. We show how quantization allows to approximate the population multiple-output regression quantiles introduced in Hallin et al. (2015), which are conditional versions of the location multivariate quantiles from Hallin et al. (2010). We prove that this approximation becomes arbitrarily accurate as the size of the quantization grid goes to infinity. We also consider a sample version of the proposed quantization-based quantiles and establish their weak consistency for their population version. Through simulations, we compare the performances of the proposed quantization-based estimators with their local constant and local bilinear kernel competitors from Hallin et al. (2015). We also compare the corresponding sample quantile regions. The results reveal that the proposed quantization-based estimators, which are local constant in nature, outperform their kernel counterparts and even often dominate their local bilinear kernel competitors.

∗Research is supported by a Bourse F.R.I.A. of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.
†Research is supported by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), the Crédit de Recherche J.0113.16 of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique, and a grant from the National Bank of Belgium.
1 Introduction

Since its introduction in the seminal paper Koenker and Bassett (1978), quantile regression has met a tremendous success. Unlike standard least squares regression that focuses on the mean of a scalar response $Y$ conditional on a $d$-dimensional covariate $X$, quantile regression is after the corresponding conditional quantile of any order $\alpha \in (0, 1)$, hence provides a thorough description of the conditional distribution of the response. It is well-known that quantile regression, as an $L_1$-type method, dominates least squares regression methods in terms of robustness, while remaining very light on the computational side since it relies on linear programming methods. Quantile regression was first defined for linear regression but was of course later extended to nonlinear/nonparametric regression. For a modern account of quantile regression, we refer to the monograph Koenker (2005).

Now, quantile regression for long has been restricted to the single-output context for which the response is univariate. The reason is of course that the lack of a canonical ordering in $\mathbb{R}^m$, $m > 1$, makes unclear how to define a suitable concept of quantile in the multivariate setting; we refer to Serfling (2002) for a nice review on multivariate quantiles. More recently, Hallin et al. (2010) introduced a multivariate quantile that enjoys all properties usually expected from a quantile and that provides, as in the univariate case, a strong connection with the concept of halfspace depth. The same paper also based on this quantile a concept of multiple-output linear regression quantiles, extending to the multiple-output setup the linear regression quantiles from Koenker and Bassett (1978). Hallin et al. (2015) — hereafter, HLPS15 — then extended this construction to nonparametric regression by introducing local constant and local bilinear versions of these multiple-output regression quantiles. The resulting nonparametric multiple-output regression quantiles make use of kernel smoothing, hence actually extend to the multiple-output setup the single-output local constant and local linear kernel quantile regression methods from Yu and Jones (1998).

Kernel methods, however, are not the only smoothing techniques that can be used to perform nonparametric quantile regression. In the single-output context, Charlier et al. (2015a) indeed showed recently that nonparametric quantile regression can alternatively be based on optimal quantization, which is a method that provides a discretized approximation $\tilde{X}^N$ of size $N$ of the $d$-dimensional (typically continuous) covariate $X$ (this discretization is obtained by projecting $X$ onto an optimal quantization grid, that is onto a collection of $N$ points minimizing the $L_p$-norm of $\tilde{X}^N - X$). Charlier et al. (2015a) introduced a nonparametric regression quantile method based on optimal quantization and derived some of its theoretical properties. Charlier et al. (2015b) then showed through simulations that the resulting sample regression quantiles often dominate their kernel competitors in terms of integrated square errors.
This dominance of quantization-based regression quantiles over their kernel competitors provides a natural motivation to try and define quantization-based analogs of the kernel multiple-output regression quantiles from HLPS15. This is the objective of the present paper, that is organized as follows. Section 2 describes the population multiple-output regression quantiles considered in this work, namely the conditional version of the multivariate quantiles from Hallin et al. (2010), and explains how these can be approximated through optimal quantization. The approximation is shown to be arbitrarily accurate as the number \(N\) of grid points used in the quantization goes to infinity. Section 3 defines the corresponding sample quantization-based regression quantiles and establishes their consistency (for the fixed-\(N\) approximation of multiple-output regression quantiles). Section 4 is devoted to numerical results: first, a data-driven method to select the smoothing parameter \(N\) is described (Section 4.1). Then, a comparison with the kernel-based competitors from HLPS15 is performed, based on empirical integrated square errors and on visual inspection of the resulting conditional quantile regions (Section 4.2). Finally, Section 5 concludes and an appendix collects the proofs.

2 Quantization-based multiple-output regression quantiles

As mentioned above, the main objective of this paper is to estimate through optimal quantization the population multiple-output regression quantiles from Hallin et al. (2015). We start by describing these quantiles.

2.1 The population multiple-output regression quantiles considered

Let \(X\) be a vector of covariates of dimension \(d\) and \(Y\) be a vector of response variables of dimension \(m\). The multivariate quantiles from HLPS15 are indexed by a vector \(\alpha\) ranging over \(B^m := \{y \in \mathbb{R}^m : 0 < |y| < 1\}\), the open unit ball of \(\mathbb{R}^m\) deprived of the origin (throughout, \(|\cdot|\) denotes the Euclidean norm). This index \(\alpha\) factorizes into \(\alpha = \alpha u\), with \(\alpha = |\alpha| \in (0, 1)\) and \(u \in S^{m-1} := \{y \in \mathbb{R}^m : |y| = 1\}\). Letting \(\Gamma_u\) be an arbitrary \(m \times (m - 1)\) matrix whose columns form, jointly with \(u\), an orthonormal basis of \(\mathbb{R}^m\), the population regression \(\alpha\)-quantiles we consider are the following.

**Definition 2.1** (HLPS15). The \(\alpha\)-quantile of \(Y\) given \(X = x\) is any element of the collection of hyperplanes \(\pi_{\alpha,x} := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^m : u'y = c'_{\alpha,x} \Gamma_u y + a_{\alpha,x}\}\) such that

\[
q_{\alpha,x} := \left(\frac{a_{\alpha,x}}{c_{\alpha,x}}\right) = \arg\min_{(\alpha, c') \in \mathbb{R}^m} \mathbb{E}[\rho_\alpha(Y_u - c' Y_u^\perp - a)|X = x],
\]

where \(\alpha =: \alpha u \in B^m\), \(Y_u := u'Y\), \(Y_u^\perp := \Gamma_u' Y\) and \(z \mapsto \rho_\alpha(z) := z(\alpha - \mathbb{1}_{z<0})\) is the usual check function (throughout, \(\mathbb{1}_A\) is the indicator function of \(A\)).
Like their original location version introduced in Hallin et al. (2010), the $\alpha$-quantiles from Definition 2.1 are directional quantiles. The vector $u$ is the direction of the quantile, whereas the scalar quantity $\alpha$ refers to the order of the quantile. The $y$-part of the $\alpha$-quantile from Definition 2.1, that is,

$$\{ y \in \mathbb{R}^m : u'y = c_{\alpha,x}' \Gamma_u'y + a_{\alpha,x} \},$$

(2.2)

actually provides the Hallin et al. (2010) (location) multivariate $\alpha$-quantile associated with the conditional distribution $P^{Y|X=x}$ of $Y$ given $X = x$. We refer to Hallin et al. (2010) for the exact interpretation of this quantile. It then follows directly from Hallin et al. (2010) that, under mild assumptions, the regression $\alpha$-quantile region

$$R_{\alpha,x} := \bigcap_{u \in S^{m-1}} \{ y \in \mathbb{R}^m : u'y \geq c_{\alpha,x}' \Gamma_u'y + a_{\alpha,u,x} \}, \quad \alpha \in (0, \frac{1}{2}),$$

(2.3)

coincides with the level-$\alpha$ halfspace depth region of $P^{Y|X=x}$ (that is, with the set of $y$'s whose halfspace depth with respect to $P^{Y|X=x}$ is larger than or equal to $\alpha$). In the single-output case ($m = 1$), the quantile in (2.2), for $u = 1$ (resp., for $u = -1$), reduces to the usual $\alpha$-quantile (resp., $(1 - \alpha)$-quantile) of $P^{Y|X=x}$, and $R_{\alpha,x}$ coincides with the interval whose end points are these two quantiles.

If $P^{Y|X=x}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with a density that has a connected support and admits finite first-order moments, the minimization problem in (2.1) admits a unique solution; see Hallin et al. (2010). Moreover, $(a, c') \mapsto G_{a,c}(x) = E[\rho_a(Y_u - c'Y_u^\top - a)|X = x]$ is then convex and continuously differentiable on $\mathbb{R}^m$. Therefore, under these assumptions, $q_{\alpha,x} = (a_{\alpha,x}, c_{\alpha,x}')'$ is alternatively characterized as the unique solution of the system of equations

$$\partial_a G_{a,c}(x) = P[u'Y < a + c'\Gamma_u'y|X = x] - \alpha = 0$$

(2.4)

$$\nabla_c G_{a,c}(x) = E[\Gamma'_uY(\alpha - \|u'y < a + c'\Gamma_u'y\|)|X = x] = 0;$$

(2.5)

see Lemma A.4 in Appendix A. As we will see in the sequel, twice differentiability of $(a, c') \mapsto G_{a,c}(x)$ actually requires slightly stronger assumptions.

### 2.2 Quantization-based multiple-output regression quantiles

We now construct an approximation of the above regression quantiles based on optimal quantization. For any fixed $N \in \mathbb{N}_0:= \{1, 2, \ldots \}$, quantization replaces the random $d$-vector $X$ by a discrete version $\tilde{X}^{\gamma^N} := \text{Proj}_{\gamma^N}(X)$ obtained by projecting $X$ onto the $N$-quantization grid $\gamma^N(\in (\mathbb{R}^d)^N)$. The quantization grid is optimal if it minimizes the quantization error $\|\tilde{X}^{\gamma^N} - X\|_p$, where $\|Z\|_p := (E[|Z|^p])^{1/p}$ denotes the $L_p$-norm of $Z$. Existence (but not
unicity) of such an optimal grid is guaranteed if the distribution of $X$ does not charge any hyperplane; see, e.g., Pagès (1998). In the sequel, $\tilde{X}^N$ will denote the projection of $X$ onto an arbitrary optimal $N$-grid. This approximation becomes more and more precise as $N$ increases since $\|\tilde{X}^N - X\|_p = O(N^{-1/d})$ as $N \to \infty$; see, e.g., Graf and Luschgy (2000). More details on optimal quantization can be found in Pagès (1998), Pagès and Printems (2003) or Graf and Luschgy (2000).

Let $p \geq 1$ such that $\|X\|_p < \infty$ and let $\gamma^N$ be an optimal quantization grid. Replacing $X$ in (2.1) by its projection $\tilde{X}^N$ onto $\gamma^N$ leads to considering

$$\tilde{q}^N_{\alpha,x} = \left( \tilde{q}^N_{\alpha,x}, \tilde{c}^N_{\alpha,x} \right) = \arg\min_{(a,c') \in \mathbb{R}^m} \mathbb{E}[\rho_\alpha(Y_u - c' Y_u^\perp - a)|\tilde{X}^N = \tilde{x}], \quad (2.6)$$

where $\tilde{x}$ denotes the projection of $x$ onto $\gamma^N$. A quantization-based approximation of the multiple-output regression quantile from Definition 2.1 above is thus any hyperplane of the form

$$\tilde{\pi}^N_{\alpha,x} := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^m : u'y = (\tilde{c}^N_{\alpha,x})' \Gamma_u y + \tilde{a}^N_{\alpha,x} \}.$$ 

This quantization-based quantile being entirely characterized by $\tilde{q}^N_{\alpha,x}$, we will investigate the quality of this approximation through $\tilde{q}^N_{\alpha,x}$. Since $\tilde{X}^N - X$ goes to zero in $L_\infty$-norm as $N$ goes to infinity, we may expect that $\tilde{q}^N_{\alpha,X} - q_{\alpha,X}$ also converges to zero in an appropriate sense. To formalize this, the following assumptions are needed.

**Assumption (A)** (i) The random vector $(X, Y)$ is generated through $Y = M(X, \epsilon)$, where the $d$-dimensional covariate vector $X$ and the $m$-dimensional error vector $\epsilon$ are mutually independent; (ii) the support $S_X$ of the distribution $P_X$ of $X$ is compact; (iii) denoting by $GL_m(\mathbb{R})$ the set of $m \times m$ invertible real matrices, the link function $M : S_X \times \mathbb{R}^m \to \mathbb{R}^m$ is of the form $(x, z) \mapsto M(x, z) = M_{a,x} + M_{b,x} z$, where the functions $M_{a,} : S_X \to \mathbb{R}^m$ and $M_{b,} : S_X \to GL_m(\mathbb{R})$ are Lipschitz with respect to the Euclidean norm and operator norm, respectively (see below); (iv) the distribution $P_X$ of $X$ does not charge any hyperplane; (v) $\|X\|_p < \infty$ and $\|\epsilon\|_p < \infty$.

For the sake of clarity, we make precise that the Lipschitz properties of $M_{a,}$ and $M_{b,}$ in Assumption (A)(iii) mean that there exist constants $C_1, C_2 > 0$ such that

$$\forall x_1, x_2 \in \mathbb{R}^d, \quad |M_{a,x_1} - M_{a,x_2}| \leq C_1|x_1 - x_2|, \quad (2.7)$$

$$\forall x_1, x_2 \in \mathbb{R}^d, \quad \|M_{b,x_1} - M_{b,x_2}\| \leq C_2|x_1 - x_2|, \quad (2.8)$$

where $\|A\| = \sup_{u \in S^{m-1}} |Au|$ denotes the operator norm of $A$. The smallest constant $C_1$ (resp., $C_2$) that satisfies (2.7) (resp., (2.8)) will be denoted as $[M_{a,}]_{\text{Lip}}$ (resp., $[M_{b,}]_{\text{Lip}}$).
Assumption (B) The distribution of $\varepsilon$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with a density $f^\varepsilon : \mathbb{R}^m \to \mathbb{R}_+^+$ that is bounded, has a connected support, admits finite second-order moments, and satisfies, for some constants $C > 0$, $r > m - 1$ and $s > 0$,

$$|f^\varepsilon(z_1) - f^\varepsilon(z_2)| \leq C |z_1 - z_2|^s \left(1 + \frac{1}{2} |z_1 + z_2|^2 \right)^{-\left(3 + r + s\right)/2},$$

(2.9)

for all $z_1, z_2 \in \mathbb{R}^m$.

We refer to Hallin et al. (2010) for more details on Assumption (B) and more particularly on (2.9). For technical reasons, the condition $r > m - 2$ from Hallin et al. (2010) had to be slightly reinforced into $r > m - 1$ above. As shown in Lemma A.4, Assumption (B) ensures twice continuous differentiability of $(a, c)' \mapsto G_{a,c}(x)$.

Under these assumptions, we have the following result (see Appendix A for the proof).

**Theorem 2.1.** Let Assumptions (A) and (B) hold. Then, for any $\alpha \in \mathcal{B}^m$,

$$\sup_{x \in \mathcal{S}_X} |\tilde{q}_N^\alpha, x - q_{\alpha,x}| \to 0,$$

as $N \to \infty$.

This result confirms that, as the size $N$ of the optimal quantization grid goes to infinity, the quantization-based approximation $\tilde{q}_N^\alpha, x$ of $q_{\alpha,x}$ becomes arbitrarily precise. Clearly, the approximation is actually uniform in $x$. This makes it natural to try and define, whenever observations are available, a sample version of $\tilde{q}_N^\alpha, x$ that will then be an estimator of $q_{\alpha,x}$ from which one will be able to obtain in particular sample versions of the regression quantile regions $R_{\alpha,x}$ in (2.3).

### 3 Sample quantization-based multiple-output quantiles

We now consider the problem of defining, from independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of $(X, Y)$, a sample version $\hat{q}_N^{\alpha,n}$ of the quantization-based regression quantile coefficients $\tilde{q}_N^\alpha, x$ in (2.6).

#### 3.1 Definition of the estimator

No closed form is available for an optimal quantization grid, except in some very particular cases. The definition of $\hat{q}_N^{\alpha,n}$ thus first requires constructing an (approximate) optimal grid. This may be done through a stochastic gradient algorithm, which proceeds as follows to quantize a $d$-dimensional random vector $X$.
Let \((\xi^t)_{t\in\mathbb{N}_0}\) be a sequence of independent copies of \(X\), and let \((\delta_t)_{t\in\mathbb{N}_0}\) be a deterministic sequence in \((0,1)\) satisfying \(\sum_{t=1}^\infty \delta_t = +\infty\) and \(\sum_{t=1}^\infty \delta_t^2 < +\infty\). Starting from an initial \(N\)-grid \(\hat{\gamma}^{N,0}\) with pairwise distinct clusters, the algorithm recursively defines the grid \(\hat{\gamma}^{N,t}\), \(t \in \mathbb{N}_0\), as

\[
\hat{\gamma}^{N,t} = \hat{\gamma}^{N,t-1} - \frac{\delta_t}{p} \nabla_x d_N^p(\hat{\gamma}^{N,t-1}, \xi^t),
\]

where \(\nabla_x d_N^p(x, \xi)\) is the gradient with respect to the \(x\)-component of the so-called local quantization error \(d_N^p(x, \xi) = \min_{i=1, \ldots, N} |x_i - \xi|^p\), with \(x = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N\) and \(\xi \in \mathbb{R}^d\). Since \((\nabla_x d_N^p(x, \xi))_i = p|x_i - \xi|^{p-2}(x_i - \xi)\I_{[x_i = \Proj_x(\xi)]}, i = 1, \ldots, N\), two consecutive grids \(\hat{\gamma}^{N,t-1}\) and \(\hat{\gamma}^{N,t}\) differ by one point only, namely the point corresponding to the non-zero component of this gradient. The reader can refer to Pagès (1998), Pagès and Printems (2003) or Graf and Luschgy (2000) for more details on this algorithm, which, for \(p = 2\), is known as the Competitive Learning Vector Quantization (CLVQ) algorithm.

The construction of \(\hat{q}^{N,n}_{\alpha,x}\) then proceeds in two steps.

(S1) An “optimal” quantization grid is obtained from the algorithm above. First, an initial grid \(\hat{\gamma}^{N,0}\) is selected by sampling randomly without replacement among the \(X_i\)'s, under the constraint that the same value cannot be picked more than once (a constraint that is relevant only if there are ties in the \(X_i\)'s). Second, \(n\) iterations of the algorithm are performed, based on \(\xi^t = X_i\), for \(t = 1, \ldots, n\). The resulting optimal grid is denoted as \(\hat{\gamma}^{N,n} = (\hat{x}^{N,n}_1, \ldots, \hat{x}^{N,n}_n)\).

(S2) The approximation \(\hat{q}^{N}_{\alpha,x} = \arg\min_{a,c'} \mathbb{E}[\rho_{\alpha}(u'Y - c'\Gamma_u Y - a) | \hat{X}^N = \hat{x}]\) in (2.6) is then estimated by

\[
\hat{q}^{N,n}_{\alpha,x} = \left(\frac{\hat{q}^{N,n}_{\alpha,x}}{\hat{c}^{N,n}_{\alpha,x}}\right) = \arg\min_{a,c' \in \mathbb{R}^m} \sum_{i=1}^n \rho_{\alpha}(u'y_i - c'\Gamma_u Y - a) \I_{[\hat{X}^N_i = \hat{x}]} \tag{3.1}
\]

where \(\hat{X}^N = (\hat{X}^N_1, \ldots, \hat{X}^N_n) = \Proj_{\mathbb{R}^N}(X)\) and \(\hat{x} = (\hat{x}^{N,n}_1, \ldots, \hat{x}^{N,n}_n)\).

An estimator of the multiple-output regression quantiles from Definition 2.1 is then any hyperplane of the form

\[
\hat{\pi}^{N,n}_{\alpha,x} := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^m : u'y = (\hat{c}^{N,n}_{\alpha,x})'\Gamma_u y + \hat{q}^{N,n}_{\alpha,x}\}.
\]

Since this estimator is entirely characterized by \(\hat{q}^{N,n}_{\alpha,x}\), we may focus on \(\hat{q}^{N,n}_{\alpha,x}\) when investigating the properties of these sample quantiles. We will show that, for fixed \(N(\in \mathbb{N}_0)\) and \(x(\in S_X)\), \(\hat{q}^{N,n}_{\alpha,x}\) is a weakly consistent estimator for \(q^{N}_{\alpha,x}\). The result requires restricting to \(p = 2\) and reinforcing Assumption (A)(iv) into the following assumption.
Theorem 3.1. Let Assumption (A)' hold. Then, for any $\alpha \in B^m$, $x \in S_X$ and $N \in \mathbb{N}_0$,

$$\left| \hat{q}_{\alpha,x}^{N,n} - \tilde{q}_{\alpha,x}^{N} \right| \to 0 \quad \text{as } n \to \infty$$

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

In principle, Theorem 2.1 and Theorem 3.1 could be combined to provide an asymptotic result stating that

$$\left| \hat{q}_{\alpha,x}^{N,n} - q_{\alpha,x} \right| \to 0 \quad \text{as } n \to \infty$$

in probability, with $N = N_n$ going to infinity at an appropriate rate. However, obtaining such a result is extremely delicate, since all convergence results for the CLVQ algorithm are as $n \to \infty$ with $N$ fixed.

3.2 A bootstrap modification

For small sample sizes, the stochastic gradient algorithm above is likely to provide a grid that is far from being optimal, which may have a negative impact on the proposed sample quantiles. To improve on this, we propose the same bootstrap approach as the one adopted in the single-output context by Charlier et al. (2015a,b):

(S1) For some integer $B$, we first generate $B$ samples of size $n$ with replacement from the initial sample $X_1, \ldots, X_n$, that we denote as $\{\xi^i_t, t = 1, \ldots, n\}, b = 1, \ldots, B$. We also generate initial grids $\hat{\gamma}^{N,0}_b$ as above, by sampling randomly among the corresponding $\{\xi^i_t, t = 1, \ldots, n\}$ under the constraints that the $N$ values are pairwise distinct. We then perform $B$ times the CLVQ algorithm with iterations based on $\{\xi^i_t, t = 1, \ldots, n\}$ and with initial grid $\hat{\gamma}^{N,0}_b$. This provides $B$ optimal grids $\hat{\gamma}^{N,n}_b, b = 1, \ldots, B$ (each of size $N$).

(S2) Each of these grids is then used to estimate multiple-output regression quantiles. Working again with the original sample $(X_i, Y_i), i = 1, \ldots, n$, we project the $X$-part onto the grids $\gamma^{N,n}_b, b = 1, \ldots, B$. Therefore, (3.1) provides $B$ estimates of $q_{\alpha,x}$, denoted as $\hat{q}_{\alpha,x}^{(b),N,n}$, $b = 1, \ldots, B$. This leads to the bootstrap estimator

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

obtained by averaging these $B$ estimates.

Denoting by $\hat{R}_{\alpha,x}$ the resulting sample quantile regions (see Section 4.2.3 for more details), the parameter $B$ should be chosen large enough to smooth the mappings $x \mapsto \hat{R}_{\alpha,x}$, but not too
large to keep the computational burden under control. We use $B = 50$ or $B = 100$ in the sequel. The choice of $N$, that plays the role of the smoothing parameter in the nonparametric regression method considered, has an important impact on the proposed estimators and is discussed in the next section.

4 Numerical results

In this section, we explore the numerical performances of the proposed estimators. We first introduce in Section 4.1 a data-driven method for selecting the size $N$ of the quantization grid. In Section 4.2, we then compare the proposed (bootstrap) quantization-based estimators with their kernel-type competitors from HLPS15.

4.1 Data-driven selection of $N$

In this section, we extend the $N$-selection criterion developed in Charlier et al. (2015b) to the present multiple-output context. This criterion is based on the minimization of an empirical integrated square error (ISE) quantity that is essentially convex in $N$, which allows to identify an optimal value $N_{\text{opt}}$ of $N$.

Let $x_1, \ldots, x_J$ be values of interest in $S_X$ and $u_1, \ldots, u_K$ be directions of interest in $S^{m-1}$, with $J, K$ finite. The procedure to select $N$ works as follows. For any combination of $x_j$ and $u_k$, we first compute $q_{\alpha u_k, x_j}^{N,n} = \frac{1}{B} \sum_{b=1}^{B} q_{\alpha u_k, x_j}^{(b),N,n}$ from $B$ bootstrap samples as above. We then generate $\tilde{B}$ further samples of size $n$ with replacement from the initial sample $X_1, \ldots, X_n$, and we perform $\tilde{B}$ times the CLVQ algorithm with iterations based on these samples. This provides $\tilde{B}$ optimal quantization grids. Working again with the original sample $(X_i, Y_i)$, $i = 1, \ldots, n$ and using the $\tilde{b}$th grid, (3.1) provides $\tilde{B}$ new estimations, denoted $q_{\alpha u_k, x_j}^{(B+b),N,n}$, $\tilde{b} = 1, \ldots, \tilde{B}$. We then consider

$$\text{ISE}_{\alpha,B,\tilde{B},J,K}(N) = \frac{1}{J} \sum_{j=1}^{J} \left( \frac{1}{K} \sum_{k=1}^{K} \left( \frac{1}{\tilde{B}} \sum_{\tilde{b}=1}^{\tilde{B}} \left| q_{\alpha u_k, x_j}^{N,n} - \hat{q}_{\alpha u_k, x_j}^{(B+b),N,n} \right|^2 \right) \right).$$

To make the notation lighter, we simply denote these integrated square errors as $\text{ISE}_{\alpha}(N)$ (throughout, our numerical results will be based on $m = 2$, $\tilde{B} = 30$ and $K$ equispaced directions in $S^1$; the values of $B$, $K$ and $x_1, \ldots, x_J$ will be made precise in each case).

These sample ISEs are to be minimized in $N$. Since not all values of $N$ can be considered in practice, we rather consider

$$\hat{N}_{\alpha,\text{opt}} = \arg \min_{N \in \mathcal{N}} \text{ISE}_{\alpha}(N),$$

where the cardinality of $\mathcal{N} \subset \mathbb{N}_0$ is finite and may be chosen as a function of $n$.  

9
For illustration purposes, we simulated random samples of size $n = 999$ according to the model

$$(M1) \quad (Y_1, Y_2) = (X, X^2) + (1 + X^2)\varepsilon,$$

where $X \sim U([-2, 2])$, $\varepsilon$ has independent $N(0, 1/4)$ marginals, and $X$ and $\varepsilon$ are independent. Figure 1 plots, for $\alpha = 0.2$ and $\alpha = 0.4$, the graphs of $N \mapsto \widehat{\text{ISE}}_\alpha(N)$, where the ISEs are based on $B = 50, K = 40$ and $x_1 = -1.89, x_2 = -1.83, x_3 = -1.77, \ldots, x_J = 1.89$ (more precisely, the figure shows the average of the corresponding graphs, computed from 100 mutually independent replications). It is seen that ISE curves are indeed essentially convex in $N$ and allow to select $N$ equal to 10 for both values of $\alpha$.

### 4.2 Comparison with competitors

In this section, we investigate the numerical performances of our estimator $\bar{q}_{N,n}^{\alpha, x}$. In Section 4.2.1, we first define the competitors that will be considered. Then we compare the respective ISEs through simulations (Section 4.2.2) and show how the estimated quantile regions compare on a given sample (Section 4.2.3).

#### 4.2.1 The competitors considered

The main competitors are the local constant and local bilinear estimators from HLPS15, that extend to the multiple-output setting the local constant and local linear estimators of Yu and Jones (1998), respectively. To describe these estimators, fix a kernel function $K : \mathbb{R}^d \to \mathbb{R}^+$ and
a bandwidth \( h \). Writing \( Y_{iu} := u' \xi_i \) and \( Y_{iu}^\perp := \Gamma_u \xi_i \), the local constant estimator is then the minimizer \( \hat{q}_{\alpha,x}^c = (\hat{\beta}_{\alpha,x}^c, (\hat{\gamma}_{\alpha,x}^c)' )' \) of

\[
q \mapsto \sum_{i=1}^{n} K_{(X_i - x)} \rho_\alpha (Y_{iu} - q' \lambda_{iu}^c), \quad \text{with } \lambda_{iu}^c := \left( \frac{1}{Y_{iu}} \right).
\]

As for the local (bi)linear estimator \( \hat{q}_{\alpha,x}^\ell = (\hat{\beta}_{\alpha,x}^\ell, (\hat{\gamma}_{\alpha,x}^\ell)' )' \), its transpose vector \( (\hat{q}_{\alpha,x}^\ell)' \) is given by the first row of the \( (d + 1) \times m \) matrix \( Q \) that minimizes

\[
Q \mapsto \sum_{i=1}^{n} K_{(X_i - x)} \rho_\alpha (Y_{iu} - (Q' \lambda_{iu}^\ell)), \quad \text{with } \lambda_{iu}^\ell := \left( \frac{1}{Y_{iu}} \right) \otimes \left( \frac{1}{X_i - x} \right).
\]

As explained in HLPS15, the local bilinear approach is more informative than the local constant one and should be more reliable close to the boundary of the covariate support. However, the price to pay is an increase of the covariate space dimension (\( \lambda_{iu}^c \) is of dimension \( m \), whereas \( \lambda_{iu}^\ell \) is of dimension \( m(d + 1) \)). We refer to HLPS15 for more details on these approaches.

In the sequel, we consider \( d = 1 \) and \( m = 2 \) in order to provide graphical representations of the corresponding quantile regions. The kernel \( K \) will be the density of the bivariate standard Gaussian distribution and we choose, as in most applications in HLPS15,

\[
h = \frac{3s_x}{n^{1/5}},
\]

where \( s_x \) stands for the empirical standard deviation of \( X_1, \ldots, X_n \).

### 4.2.2 Comparison of ISEs

We now compare our bootstrap estimators with the competitors above in terms of ISEs. To do so, we generated 500 independent samples of size \( n = 999 \) from

(M1) \( (Y_1, Y_2) = (X, X^2) + (1 + X^2) \varepsilon_1 \),

(M2) \( (Y_1, Y_2) = (X, X^2) + \varepsilon_1 \),

(M3) \( (Y_1, Y_2) = (X, X^2) + (1 + \frac{3}{2} \left( \sin \left( \frac{x}{2} X \right) \right)^2) \varepsilon_2 \),

where \( X \sim U([-2, 2]) \), \( \varepsilon_1 \) has independent \( \mathcal{N}(0, 1/4) \) marginals, \( \varepsilon_2 \) has independent \( \mathcal{N}(0, 1) \) marginals, and \( X \) is independent of \( \varepsilon_1 \) and \( \varepsilon_2 \) (these models were already considered in HLPS15).

Both the proposed quantization-based quantiles and their competitors are indexed by a scalar order \( \alpha \in (0, 1) \) and a direction \( u \in S^1 \). In this section, we compare efficiencies when estimating a given conditional quantile \( q_{\alpha u,x} \). In the sequel, we still work with \( \alpha = 0.2, 0.4 \) and we fix \( u = (0, 1)' \).

For each replication in each model, the various quantile estimators were computed, based on the bandwidth \( h \) in (4.4) for the HLPS15 estimators and based on \( B = 100 \) and the \( N \)-selection procedure described in Section 4.1 (with \( x_1 = -1.89, x_2 = -1.83, \ldots, x_J = 1.89, \mathcal{N} = \)
\{10, 15, 20\} and \( K = 1 \) direction, namely the direction \( \mathbf{u} = (0, 1)' \) above) for the quantization-based estimators. For each estimator, we then evaluated

\[
\text{ISE}^a_{\alpha} = \sum_{j=1}^{J} (\hat{a}_{\alpha u,x_j} - a_{\alpha u,x_j})^2 \quad \text{and} \quad \text{ISE}^c_{\alpha} = \sum_{j=1}^{J} (\hat{c}_{\alpha u,x_j} - c_{\alpha u,x_j})^2,
\]

still for \( x_1 = -1.89, x_2 = -1.83, \ldots, x_J = 1.89 \); here, \( \hat{a}_{\alpha u,x_j} \) stands for \( \hat{\alpha}_{\alpha u,x_j}^N, \hat{\alpha}_{\alpha u,x_j}^c \) or \( \hat{\alpha}_{\alpha u,x_j}^\ell \) and \( \hat{c}_{\alpha u,x_j} \) for \( \hat{\ell}_{\alpha u,x_j}^N, \hat{\ell}_{\alpha u,x_j}^c \) or \( \hat{\ell}_{\alpha u,x_j}^\ell \). Figure 2 reports, for each model and each estimator, the boxplots of ISE compared to its local bilinear kernel competitor.

Results reveal that the proposed estimator \( \hat{q}_{\alpha,x}^N \) and the local bilinear estimator \( \hat{q}_{\alpha,x}^\ell \) perform significantly better than the local constant estimator \( \hat{q}_{\alpha,x}^c \), particularly for the estimation of the first component \( a_{\alpha,x} \) of \( q_{\alpha,x} \). In most cases, the proposed estimator \( q_{\alpha,x}^N \) actually also dominates the local bilinear one \( q_{\alpha,x}^\ell \) (the only cases where the opposite holds relate to the estimation of \( c_{\alpha,x} \) and the difference of performance is then really small). It should be noted that the quantization-based estimator \( q_{\alpha,x}^N \) is local constant in nature, which makes it remarkable that it behaves well in terms of ISE compared to its local bilinear kernel competitor.

### 4.2.3 Comparison of sample quantile regions

As explained in HLPS15, the regression quantile regions \( R_{\alpha,x} \) in (2.3) are extremely informative about the conditional distribution of the response, which makes it desirable to obtain well-behaved estimations of these regions. That is why we now compare the sample regions obtained from the proposed quantization-based quantile estimators with the kernel ones from HLPS15. Irrespective of the quantile coefficient estimators \( \hat{q}_{\alpha u,x} = (\hat{a}_{\alpha u,x}, \hat{c}_{\alpha u,x})' \) used, the corresponding sample regions are obtained as

\[
\hat{R}_{\alpha,x} := \cap_{u \in S^{m-1}} \{ y \in \mathbb{R}^m : u'y \geq \hat{c}_{\alpha u,x}' \Gamma_y u + \hat{a}_{\alpha u,x} \},
\]

where \( S^{m-1}_F \) is a finite subset of \( S^{m-1} \); compare with (2.3).

We considered a random sample of size \( n = 999 \) from Model (M1) and computed, for the various estimation methods, \( \hat{R}_{\alpha,x} \) for \( \alpha = 0.2, 0.4 \) and for \( x = -1.89, -1.83, -1.77, \ldots, 1.89 \); in each case, \( S_F^{m-1} = S_F^1 \) is made of 360 equispaced directions in \( S^1 \). In each model, we did not select \( h \) following the data-driven procedure mentioned in Section 4.2.1, but chose it equal to 0.37, as proposed in HLPS15, Figure 3. For the quantization-based estimators, \( N \) was selected according to the data-driven method from Section 4.1 (with \( B = 100, K = 360, N = \{5, 10, 15, 20\} \), and still \( x_1 = -1.89, x_2 = -1.83, \ldots, x_J = 1.89) \), which led to the optimal value \( N = 10 \). The resulting sample quantile regions, obtained from the quantization-based method and from the local constant and local bilinear kernel ones, are plotted in Figure 3.
Figure 2: Boxplots, for $\alpha = 0.2, 0.4$ and $u = (0, 1)'$, of $\text{ISE}_a^\alpha$ (left) and of $\text{ISE}_c^\alpha$ (right) for various conditional quantile estimators obtained from 500 independent random samples according to Models (M1) (top), (M2) (middle) and (M3) (bottom), with size $n = 999$. The estimators considered are the quantization-based estimator $\hat{q}_{\alpha,x}^{N,n}$ (in blue), the local bilinear estimator $\hat{q}_{\alpha,x}^\ell$ (in purple) and the local constant estimator $\hat{q}_{\alpha,x}^c$ (in red).
For comparison purposes, the population quantile regions $R_{\alpha,x}$ are also reported there. We observe that the quantization-based and local bilinear methods provide quantile regions that are nice and close to the population ones. They succeed in particular in catching the underlying heteroscedasticity. Clearly, they perform better than the local constant methods close to the boundary of the covariate range. While the local (bi)linear methods, as already mentioned, are known to exhibit good boundary behaviour, it is surprising that the quantization-based method also behaves well in this respect, since this method is of a local constant nature. Finally, it should be noted that, unlike the smoothing parameter of the local constant/bilinear methods (namely, $h$), that of the quantization-based method (namely, $N$) was chosen in a fully data-driven way.

5 Summary

In this paper, we defined new nonparametric estimators of the multiple-output regression quantiles proposed in HLPS15. The main idea, that was already used in the single-output context in Charlier et al. (2015a,b), is to perform localization through the concept of optimal quantization rather than through standard kernel methods. We derived consistency results that generalize to the multiple-output context those obtained in Charlier et al. (2015a). Moreover, the good empirical efficiency properties of quantization-based quantiles showed in Charlier et al. (2015b) extend to the multiple-output context. In particular, the proposed quantization-based sample quantiles, that are local constant in nature, outperform their kernel-based counterparts, both in terms of integrated square errors and in terms of visual inspection of the corresponding quantile regions. The proposed quantiles actually perform as well as (and sometimes even strictly dominate) the local bilinear kernel estimators from HLPS15.

The data-driven selection procedure we proposed for the smoothing parameter $N$ involved in the quantization-based method allows to make the estimation fully automatic. Our estimation procedure was actually implemented in R and the code is available from the authors on simple request.

A Proof of Theorem 2.1

The proof requires several lemmas. First, recall that $G_{a,c}(x) = \mathbb{E}[\rho_{\alpha}(Y_u - c'Y_u^\perp - a)|X = x]$ and consider the corresponding quantized quantity $\tilde{G}_{a,c}(\tilde{x}) = \mathbb{E}[\rho_{\alpha}(Y_u - c'Y_u^\perp - a)|\tilde{X}^N = \tilde{x}]$. Since $q_{a,x} = (a_{a,x},c_{a,x})'$ and $q_{a,x}^N = (\tilde{a}^N_{a,x},(\tilde{c}^N_{a,x})')'$ are defined as the vectors achieving the minimum of $G_{a,c}(x)$ and $\tilde{G}_{a,c}(\tilde{x})$ respectively, we naturally start controlling the distance between $\tilde{G}_{a,c}(\tilde{x})$ and $G_{a,c}(x)$. This is achieved below in Lemma A.5, whose proof requires the following
Figure 3: (a)-(c) The sample quantile regions $\hat{R}_{\alpha,x}$, for $\alpha = 0.2, 0.4$ and $x = -1.89, -1.83, -1.77, \ldots, 1.89$, computed from a random sample of size $n = 999$ from Model (M1) by using (a) the quantization-based method, (b) the local constant kernel method, and (c) the local bilinear kernel one. (d) The corresponding population quantile regions $R_{\alpha,x}$. 
preliminary lemmas. Throughout this appendix, C is a constant that may vary from line to line.

**Lemma A.1.** Let Assumption (A) hold and fix $\alpha = \alpha u \in \mathbb{B}^m$, $a \in \mathbb{R}$ and $c \in \mathbb{R}^{m-1}$. Then for $x_1, x_2 \in \mathbb{R}^d$, $|G_{a,c}(x_1) - G_{a,c}(x_2)| \leq \max(\alpha, 1 - \alpha) \sqrt{1 + |c|^2} (|M_{a,c}|_{\text{Lip}} + |M_{b,c}|_{\text{Lip}} \|\varepsilon\|_1)|x_1 - x_2|.$

**Proof.** For $x_1, x_2 \in \mathbb{R}^d$, we have

$$|G_{a,c}(x_1) - G_{a,c}(x_2)|$$

$$= |E[\rho_a(Y_u - c'Y_u - a)|X = x_1] - E[\rho_a(Y_u - c'Y_u - a)|X = x_2]|$$

$$= |E[\rho_a((u - \Gamma_u c)'M(X, \varepsilon) - a)|X = x_1] - E[\rho_a((u - \Gamma_u c)'M(X, \varepsilon) - a)|X = x_2]|$$

$$= |E[\rho_a((u - \Gamma_u c)'M(x_1, \varepsilon) - a) - \rho_a((u - \Gamma_u c)'M(x_2, \varepsilon) - a)|,$$

where we used the independence of $X$ and $\varepsilon$. Using the fact that $\rho_a$ is a Lipschitz function with Lipschitz constant $[\rho_a]_{\text{Lip}} := \max(\alpha, 1 - \alpha)$, then the Cauchy-Schwarz inequality, this yields

$$|G_{a,c}(x_1) - G_{a,c}(x_2)| \leq [\rho_a]_{\text{Lip}} E[|((u - \Gamma_u c)'(M(x_1, \varepsilon) - M(x_2, \varepsilon))|]$$

$$\leq [\rho_a]_{\text{Lip}} E[[u - \Gamma_u c]'|M(x_1, \varepsilon) - M(x_2, \varepsilon)|]$$

$$\leq [\rho_a]_{\text{Lip}} E[[u - \Gamma_u c]'(1, -c') E[|M_{a,x_1} - M_{a,x_2} + (M_{b,x_1} - M_{b,x_2})\varepsilon]|$$

$$\leq [\rho_a]_{\text{Lip}} E[[u - \Gamma_u c]'\varepsilon]$$

$$\leq [\rho_a]_{\text{Lip}} E[[u - \Gamma_u c]'\varepsilon]$$

$$\leq [\rho_a]_{\text{Lip}} E[[u - \Gamma_u c]'\varepsilon]$$

by Assumptions (A)(iii)-(v).

The following lemma shows that, under the assumptions considered, the regularity property (2.9) extends from the error density $f^\varepsilon(\cdot)$ to the conditional density $f^{Y|X=x}(\cdot)$.

**Lemma A.2.** Let Assumptions (A) and (B) hold and fix $x \in S_X$. Then, for some constants $C > 0$, $r > m - 1$ and $s > 0$, we have

$$|f^{Y|X=x}(y_1) - f^{Y|X=x}(y_2)| \leq C|y_1 - y_2|^r(1 + \frac{1}{2}|y_1 + y_2|^2)^{-\frac{3+r+s}{2}}, \quad (A.1)$$

for all $y_1, y_2 \in \mathbb{R}^m$.

**Proof.** Assumption (A) allows to rewrite the conditional density of $Y$ given $X = x$ as

$$f^{Y|X=x}(y) = \frac{1}{|\text{det}(M_{b,x})|} f^\varepsilon(M_{b,x}^{-1}(y - M_{a,x})), \quad \forall y \in \mathbb{R}^m.$$

Hence, we have

$$|f^{Y|X=x}(y_1) - f^{Y|X=x}(y_2)|$$

$$= \frac{1}{|\text{det}(M_{b,x})|} \left|f^\varepsilon(M_{b,x}^{-1}(y_1 - M_{a,x})) - f^\varepsilon(M_{b,x}^{-1}(y_2 - M_{a,x}))\right|,$$

16
Now, Assumption (B) entails
\[ |f^Y|X=x(y_1) - f^Y|X=x(y_2)| \leq \frac{C|\det(1 + \frac{1}{2} M_{b,x}^{-1}(y_1 + y_2) - 2M_{a,x})|^2}{|\det(M_{b,x})|(1 + \frac{1}{2} |M_{b,x}^{-1}(y_1 + y_2 - 2M_{a,x})|^2)^{(3+r+s)/2}} \]
\[ \leq C|(y_1 - y_2)|^s \left(1 + \frac{1}{2} |M_{b,x}^{-1}(y_1 + y_2 - 2M_{a,x})|^2\right)^{-(3+r+s)/2}, \]
where the second inequality comes from the compactness of \( S_X \) and the continuity of the mapping \( x \mapsto M_{b,x} \). The result then follows from the fact that
\[ \frac{1 + \frac{1}{2} |y_1 + y_2|^2}{1 + \frac{1}{2} |M_{b,x}^{-1}(y_1 + y_2 - 2M_{a,x})|^2} = \frac{1 + \frac{1}{2} |M_{b,x}^{-1}(y_1 + y_2 - 2M_{a,x}) + 2M_{a,x}|^2}{1 + \frac{1}{2} |M_{b,x}^{-1}(y_1 + y_2 - 2M_{a,x})|^2} \leq C, \]
where we used again the continuity of \( x \mapsto M_{a,x} \) and \( x \mapsto M_{b,x} \), and the compactness of \( S_X \). \( \square \)

We will also need the following result belonging to linear algebra.

**Lemma A.3.** For \( p > q \geq 1 \), let \( V = (v_1 \ldots v_q) \) be a \( p \times q \) full-rank matrix and \( H \) be a \( q \)-dimensional vector subspace of \( \mathbb{R}^p \). Then, there exists a \( p \times q \) matrix \( U = (u_1 \ldots u_q) \) whose columns form an orthonormal basis of \( H \) and such that \( I_q + U'^tV \) is invertible.

**Proof.** We fix \( p \geq 2 \) and we prove the result by induction on \( q \) between \( q = 1 \) and \( q = p - 1 \). We start with the case \( q = 1 \) and take \( U = (u_1) \), where \( u_1 \) is an arbitrary unit \( p \)-vector in \( H \). If \( \det(1 + U'^tV) = 1 + u_1'^t v_1 = 0 \), then we may alternatively take \( U_* = (-u_1) \), which provides \( \det(1 + U_*'^tV) = 1 - u_1'^t v_1 = 2 \neq 0 \). Assume then that the result holds for \( q \) (with \( q < p - 1 \)) and let us prove it for \( q + 1 \). Pick an arbitrary \( p \times (q + 1) \) matrix \( U = (u_1 \ldots u_{q+1}) \) whose columns form an orthonormal basis of the \((q+1)\)-dimensional vector subspace \( H \) of \( \mathbb{R}^p \). Assume that \( \det(I_{q+1} + U'^tV) = 0 \), where \( V = (v_1 \ldots v_{q+1}) \) is the given \( p \times (q + 1) \) full-rank matrix. Letting \( U_{-1} = (u_2 \ldots u_{q+1}) \) and \( V_{-1} = (v_2 \ldots v_{q+1}) \), an expansion of the determinant along the first row provides
\[ 0 = \det(I_{q+1} + U'^tV) = (u_1'^t v_1 + 1) \det(I_q + U'^t_{-1}V_{-1}) + \sum_{i=2}^{q+1} (-1)^{i+1} u_1'^t v_i \det(W_i), \]
for some \( q \times q \) matrices \( W_2, \ldots, W_m \). With \( U_* = (-u_1, u_2 \ldots u_{q+1}) \), we then have
\[ \det(I_{q+1} + U_*'^tV) = (u_1'^t v_1 + 1) \det(I_q + U_1'^t V_{-1}) - \sum_{i=2}^{q+1} (-1)^{i+1} u_1'^t v_i \det(W_i) \]
\[ = 2 \det(I_q + U_1'^t V_{-1}) - \det(I_{q+1} + U'^tV) \]
\[ = 2 \det(I_q + U_1'^t V_{-1}). \]
The induction hypothesis guarantees that $U_{-1}$ can be chosen such that $\det(I_q + U_{-1}'V_{-1})$ is non-zero, which establishes the result.

We can now calculate explicitly the gradient and the Hessian matrix of the function $(a, c) \mapsto G_{a,c}(x)$ for any $x$ in the support $S_X$ of $X$, and derive some important properties of this Hessian matrix.

**Lemma A.4.** Let Assumptions (A) and (B) hold. Then (i) $(a, c) \mapsto G_{a,c}(x)$ is twice differentiable at any $x \in S_X$, with gradient vector

$$
\nabla G_{a,c}(x) = \begin{pmatrix} \nabla_a G_{a,c}(x) \\ \nabla_c G_{a,c}(x) \end{pmatrix} = \begin{pmatrix} P[u'Y < a + c'\Gamma_u Y]X = x - \alpha \\ E[\Gamma_u Y (I_{[u'Y < a + c'\Gamma_u Y]} - \alpha)]X = x \end{pmatrix}
$$

(A.2)

and Hessian matrix

$$
H_{a,c}(x) = \int_{\mathbb{R}^{m-1}} \begin{pmatrix} 1 & t' \\ t & tt' \end{pmatrix} f_Y(x = x)((a + c't)u + \Gamma_u t) \, dt;
$$

(ii) for any $(a, c, x) \in \mathbb{R} \times \mathbb{R}^{m-1} \times S_X$, $H_{a,c}(x)$ is positive definite; (iii) $(a, c, x) \mapsto H_{a,c}(x)$ is continuous over $\mathbb{R} \times \mathbb{R}^{m-1} \times S_X$.

**Proof.** (i) Let

$$
\eta_\alpha(a, c) = (I_{[u'Y - c'\Gamma_u Y - a < 0]} - \alpha) \begin{pmatrix} 1 \\ \Gamma_u Y \end{pmatrix}.
$$

For any $(a, c', (a_0, c_0')) \in \mathbb{R}^m$, we then have

$$
\rho_\alpha(u'Y - c'\Gamma_u Y - a) - \rho_\alpha(u'Y - c_0'\Gamma_u Y - a_0) - (a - a_0, c' - c_0')\eta_\alpha(a_0, c_0) = (u'Y - c'\Gamma_u Y - a)\left( I_{[u'Y - c'\Gamma_u Y - a < 0]} - I_{[u'Y - c_0'\Gamma_u Y - a_0 < 0]} \right) \geq 0,
$$

(A.3)

so that $\eta_\alpha(a, c)$ is a subgradient for $(a, c) \mapsto \rho_\alpha(u'Y - c'\Gamma_u Y - a)$. Hence,

$$
\nabla G_{a,c}(x) = \nabla_a E[\rho_\alpha(u'Y - c'\Gamma_u Y - a)]X = x = E[\eta_\alpha(a, c)]X = x,
$$

(A.4)

which readily provides (A.2). Let us now show that

$$
|\nabla G_{a+\Delta_a, c+\Delta_c}(x) - \nabla G_{a,c}(x) - H_{a,c}(x)(\Delta_a, \Delta_c')| = o\left(|(\Delta_a, \Delta_c')|\right)
$$

as $(\Delta_a, \Delta_c') \to 0$. From (A.4) and the identity

$$
\int_{a+c't}^{(a+\Delta_a)+(c+\Delta_c)t} \begin{pmatrix} 1 \\ t \end{pmatrix} \, dz = \begin{pmatrix} 1 & t' \\ t & tt' \end{pmatrix} \begin{pmatrix} \Delta_a \\ \Delta_c \end{pmatrix},
$$

we then have

18
we obtain

\[\nabla G_{a+\Delta_a,c+\Delta_c}(x) - \nabla G_{a,c}(x) - H_{a,c}(x)(\Delta_a, \Delta'_c) \]

\[= E[\eta_\alpha(a + \Delta_a, c + \Delta_c) - \eta_\alpha(a, c) | X = x] \]

\[= \int_{\mathbb{R}^m-1} \left( \begin{array}{c}
1 \\
t \\
t t'
\end{array} \right) \frac{\Delta_a + \Delta'_c}{\Delta_c} f^Y|X=x((a + c't)u + \Gamma_ut) \, dt \]

\[= \int_{\mathbb{R}^m-1} \int_{\mathbb{R}} \left( \begin{array}{c}
1 \\
(a - (x + \Delta_a)t - (a + \Delta_a) < 0) - \mathbb{I}[a - (c + \Delta_c)t - (a + \Delta_a) < 0]
\end{array} \right) \frac{1}{t} f^Y|X=x((a + c't)u + \Gamma_ut) \, dz \, dt \]

\[= \int_{\mathbb{R}^m-1} \int_{a+c't} (a + c') t \left( \begin{array}{c}
1 \\
(a + c') t
\end{array} \right) f^Y|X=x((a + c't)u + \Gamma_ut) \, dz \, dt \]

\[= \int_{\mathbb{R}^m-1} \int_{a+c't} (a + c') t \left( \begin{array}{c}
1 \\
(a + c') t
\end{array} \right) f^Y|X=x((a + c't)u + \Gamma_ut) \} dz \, dt. \]

Now, by Lemma A.2, one has, for any \( z \) between \( a + c't \) and \( (a + \Delta_a) + (c + \Delta_c)'t \),

\[|f^Y|X=x((z u + \Gamma_ut) - f^Y|X=x((a + c't)u + \Gamma_ut)| \]

\[\leq \frac{C|z - a - c'|^s}{(1 + \frac{1}{2}|(z + a + c't)u + 2\Gamma_ut|^2)^{1+s}} \leq \frac{C|\Delta_a + \Delta'_c|^s}{|(1, t')|^{1+s}}. \]

This entails

\[|\nabla G_{a+\Delta_a,c+\Delta_c}(x) - \nabla G_{a,c}(x) - H_{a,c}(x)(\Delta_a, \Delta'_c)| \leq C \int_{\mathbb{R}^m-1} \frac{|\Delta_a + \Delta'_c|^1}{(|1, t'|)^{1+s}} dt \]

\[\leq C(|\Delta_a, \Delta'_c|^1 + s) \int_{\mathbb{R}^m-1} \frac{1}{(|1, t'|)^{1+s}} dt = o(|(\Delta_a, \Delta'_c)|), \]

as \( (\Delta_a, \Delta'_c)' \to 0 \). Therefore, \( (a, c) \mapsto G_{a,c}(x) \) is twice continuously differentiable at any \( x \in S_X \), with Hessian matrix \( H(G_{a,c}(x)) \). Eventually, Assumption (A)(iii) implies that

\[H_{a,c}(x) = \frac{1}{|\det(M_{b,x})|} \int_{\mathbb{R}^m-1} \left( \begin{array}{c}
1 \\
t \\
t t'
\end{array} \right) f^x(M^{-1}_{b,x}((a + c't)u + \Gamma_ut - M_{a,x})) dt. \quad (A.5)\]

(ii) Positive definiteness then readily follows from (A.5) and Assumption (B).

(iii) Every entry of \( H_{a,c}(x) \) is an integral involving an integrand of the form (see (A.5))

\[g_{i,j}(a, c, x, t) = \frac{t^i_j t^j_i}{|\det(M_{b,x})|} \left( f^x(M^{-1}_{b,x}((a + c't)u + \Gamma_ut - M_{a,x})) - f^x((1, t')) \right) \]

\[+ \frac{t^i_j t^j_i}{|\det(M_{b,x})|} f^x((1, t')) =: g_{i,j}^l(a, c, x, t) + g_{i,j}^u(a, c, x, t), \]
where $\delta_i, \delta_j \in \{0, 1\}$. Clearly, for any $t$, $(a, c, x) \mapsto g^I_{i,j}(a, c, x, t)$ and $(a, c, x) \mapsto g^H_{i,j}(a, c, x, t)$ are continuous. Therefore, in view of Theorem 8.5 in Briane and Pagès (2012), it is sufficient to prove that there exist integrable functions $h^I_{i,j}, h^H_{i,j} : \mathbb{R}^{m-1} \to \mathbb{R}^+$ such that
\[
|g^I_{i,j}(a, c, x, t)| \leq h^I_{i,j}(t) \quad \text{and} \quad |g^H_{i,j}(a, c, x, t)| \leq h^H_{i,j}(t) \quad \text{for any} \ (a, c, x, t).
\]
Since Assumptions (A)(ii)-(iii) ensure that $\det(M_{b,x})$ stays away from 0 for any $x \in S_X$, we can take $t \mapsto h^I_{i,j}(t) := t^\delta_i \delta_j f^\varepsilon((1, t'))/(\inf_{x \in S_X} |\det(M_{b,x})|)$, whose integrability follows from the fact that $f^\varepsilon(\cdot)$ is bounded and $\varepsilon$ has finite second-order moments. Now, Lemma A.2 and Assumptions (A)(ii)-(iii) readily entail that there exist $r > m - 1$ and $s > 0$ such that
\[
|g^I_{i,j}(a, c, x, t)| = |t^\delta_i \delta_j| \times |f^Y|_{X=x}(a + c't)u + \Gamma_u t - M_{b,x}(1, t') - M_{a,x}|^s \leq |t^\delta_i \delta_j| (1 + \frac{1}{2}|a + c't)u + \Gamma_u t + M_{b,x}(1, t') + M_{a,x}|^{(3+r+s)/2} \leq C|t^\delta_i \delta_j|(1 + |t|^s)(1 + \frac{1}{2}|t + \Gamma'_u M_{b,x}(1, t') + \Gamma'_u M_{a,x}|^2)^{-\frac{(3+r+s)}{2}} \leq C|t^\delta_i \delta_j|(1 + |t|^s)(1 + \frac{1}{2}|(I_{m-1} + \Gamma'_u A_x) t + \Gamma'_u B_x|^2)^{(3+r+s)/2},
\]
where the matrices $A_x := (M_{b,x})_2$ and $B_x := (M_{b,x})_1 - M_{a,x}$ are based on the partition $M_{b,x} = ((M_{b,x})_1, (M_{b,x})_2)$ into an $m \times 1$ matrix $(M_{b,x})_1$ and an $m \times (m - 1)$ matrix $(M_{b,x})_2$. Lemma A.3 implies that it is always possible to choose $\Gamma_u$ in such a way that $I_{m-1} + \Gamma'_u A_x$ is invertible. Consequently, one may proceed as in the proof of Lemma A.2 and write
\[
\frac{1 + |t|^2}{1 + \frac{1}{2}|(I_{m-1} + \Gamma'_u A_x) t + \Gamma'_u B_x|^2} = \frac{1 + |(I_{m-1} + \Gamma'_u A_x)^{-1}[(I_{m-1} + \Gamma'_u A_x) t + \Gamma'_u B_x] - (I_{m-1} + \Gamma'_u A_x)^{-1}\Gamma'_u B_x|^2}{C + 1 + \frac{1}{2}|(I_{m-1} + \Gamma'_u A_x) t + \Gamma'_u B_x|^2} \leq C,
\]
where we used the fact that $x \mapsto A_x$ and $x \mapsto B_x$ are continuous functions defined over the compact set $S_X$. Therefore, (A.6) provides $|g^I_{i,j}(a, c, x, t)| \leq C|t^\delta_i \delta_j|(1 + |t|^s)(1 + |t|^2)^{-\frac{(3+r+s)}{2}} =: h^I_{i,j}(t)$, where $h^I_{i,j}(\cdot)$ is integrable over $\mathbb{R}^{m-1}$ (since $r > m - 1$).

The proof of Theorem 2.1 still requires the following lemma.

**Lemma A.5.** Let Assumptions (A) and (B) hold, fix $\alpha \in \mathbb{R}^m$, and write $\tilde{x} = \tilde{x}^N = \text{Proj}_{\gamma_N}(x)$ for any $x$. Then,

(i) for any compact set $K(\subset \mathbb{R}^{m-1})$, $\sup_{x \in S_X} \sup_{a \in \mathbb{R}, c \in K} |\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(x)| \to 0$ as $N \to \infty$;

(ii) $\sup_{x \in S_X} |\min_{(a,c) \in \mathbb{R}} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c) \in \mathbb{R}} G_{a,c}(x)| \to 0$ as $N \to \infty$.

20
Proof. (i) Fix \( a \in \mathbb{R} \) and \( c \in K \). First note that \( \tilde{X}^N = \tilde{x} \) is equivalent to \( X \in C_{\tilde{x}} \), where we let \( C_x = C_x^N = \{ z \in S_X : \text{Proj}_{\gamma_N}(z) = \tilde{x} \} \). Hence, one has

\[
\left| E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} \right| - E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} \right| 
\leq \sup_{z \in C_x} |E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = z| - E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} ||,
\]

which provides

\[
|\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(\tilde{x})| 
\leq \left| E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} \right| - E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} || 
+ \left| E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} \right| - E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} || 
\leq 2 \sup_{z \in C_x} |E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = z| - E[\rho_a(Y_u - c'Y_{u,\perp}^+ - a)]X = \tilde{x} || 
\leq 2 \sup_{z \in C_x} |G_{a,c}(z) - G_{a,c}(\tilde{x})| 
\leq 2 \max(\alpha, 1 - \alpha) \sqrt{1 + |c|^2} (\|M_a\|_{\text{Lip}} + \|M_c\|_{\text{Lip}} \|e\|_1) \sup_{z \in C_x} |z - \tilde{x}|,
\]

where we used Lemma A.1. It directly follows that, for some \( C \) that does not depend on \( N \),

\[
sup_{x \in S_X} \sup_{a \in \mathbb{R}, c \in K} |\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(\tilde{x})| \leq C \sup_{x \in S_X} |z - \tilde{x}| =: C \sup_{x \in S_X} R(C_x);
\]

the quantity \( R(C_x) \) is the “radius” of the cell \( C_x \). The result then follows from the fact that \( \sup_{x \in S_X} R(C_x) \to 0 \) as \( N \to \infty \); see Lemma A.2(ii) in Charlier et al. (2015a).

(ii) For simplicity of notations, we write \( \tilde{a} = \tilde{a}_{\alpha,x}^N \) and \( \tilde{c} = \tilde{c}_{\alpha,x}^N \). From Lemma A.4(ii),

\[
v'H_{a,c}(x)v > 0
\]

for any \( x \in S_X \) and any \( v \in S^{m-1} = \{ x \in \mathbb{R}^m : |x| = 1 \} \). The compactness assumption on \( S_X \) and the continuity of \( x \mapsto H_{a,c}(x) \) (Lemma A.4(iii)) yield that

\[
\inf_{x \in S_X} \inf_{v \in S^{m-1}} v'H_{a,c}(x)v > 0.
\]

This, jointly with Part (i) of the result, implies that there exists a positive integer \( N_0 \) and a compact set, \( K_\alpha(\subset \mathbb{R}^m) \) say, such that, for all \( N \geq N_0 \) and for all \( x \in S_X, \tilde{a}_{\alpha,x}^N \) and \( \tilde{c}_{\alpha,x}^N \) belong to \( K_\alpha \). In particular, for all \( N \geq N_0 \) and for all \( x \in S_X, \tilde{e}_{\alpha,x}^N \) and \( \tilde{c}_{\alpha,x}^N \) belong to a compact set \( K_\alpha^C(\subset \mathbb{R}^{n-1}) \). Then, with \( \mathbb{I}^+ = \|_{[\min(a,c') \in \mathbb{R} \geq \min(a,c') \in \mathbb{R} \min G_{a,c}(\tilde{x})]} \), we have

\[
\min_{(a,c') \in \mathbb{R}^m} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c') \in \mathbb{R}^m} G_{a,c}(\tilde{x}) \mathbb{I}^+ = (\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(\tilde{x})) \mathbb{I}^+ 
\leq (\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(\tilde{x})) \mathbb{I}^+ \leq \sup_{a \in \mathbb{R}, c \in K_\alpha^C} |\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(\tilde{x})| \mathbb{I}^+.
\]
Similarly, with $\mathbb{I}_- := 1 - \mathbb{I}_+$, we have

$$
\left| \min_{(a,c)'} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c)'} G_{a,c}(x) \right| \mathbb{I}_- = \left| (G_{a\alpha,x,c \alpha,x}(x) - \tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x})) \mathbb{I}_- \right|
\leq \left( \tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) - \tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) \right) \mathbb{I}_- \leq \sup_{a \in \mathbb{R}, c \in K^c_{\alpha}} |\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(x)| \mathbb{I}_-.
$$

(A.8)

From (A.7)-(A.8), we readily obtain

$$
\left| \min_{(a,c)'} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c)'} G_{a,c}(x) \right| \leq \sup_{a \in \mathbb{R}, c \in K^c_{\alpha}} |\tilde{G}_{a,c}(\tilde{x}) - G_{a,c}(x)|.
$$

The result then directly follows from Part (i) of the result.

We can now prove Theorem 2.1.

**Proof of Theorem 2.1.** Write again $\tilde{x} = \tilde{x}^N = \text{Proj}_{\mathbb{R}^N}(x)$ and fix the same integer $N_0$ and the same compact sets $K_\alpha$ and $K^c_\alpha$ as in the proof of Lemma A.5. Then, for $x \in S_X$ and $N \geq N_0$, one has

$$
|\tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) - G_{a\alpha,x,c \alpha,x}(x)|
\leq |\tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) - G_{\tilde{a},\tilde{c}}(\tilde{x})| + |G_{a\alpha,x,c \alpha,x}(x) - G_{a\alpha,x,c \alpha,x}(x)|
\leq \sup_{a \in \mathbb{R}, c \in K^c_{\alpha}} |G_{a,c}(x) - \tilde{G}_{a,c}(\tilde{x})| + \left| \min_{a,c} \tilde{G}_{a,c}(\tilde{x}) - \min_{a,c} G_{a,c}(x) \right|
\leq \sup_{x \in S_X} \sup_{a \in \mathbb{R}, c \in K^c_{\alpha}} |G_{a,c}(x) - \tilde{G}_{a,c}(\tilde{x})| + \sup_{x \in S_X} \left| \min_{a,c} \tilde{G}_{a,c}(\tilde{x}) - \min_{a,c} G_{a,c}(x) \right|.
$$

Therefore, Lemma A.5 implies that, as $N \to \infty$,

$$
\sup_{x \in S_X} |\tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) - G_{a\alpha,x,c \alpha,x}(x)| \to 0.
$$

(A.9)

Performing a second-order expansion about $q_{a\alpha,x} = (a_{\alpha,x}, c_{\alpha,x})'$ provides

$$
G_{\tilde{a},\tilde{c}}(\tilde{x}) - G_{a\alpha,x,c \alpha,x} = \frac{1}{2} (\tilde{q}_{a\alpha,x} - q_{a\alpha,x})' H^N_{x}(x) (\tilde{\tilde{q}}_{a\alpha,x} - q_{a\alpha,x}),
$$

with $H^N_{x} := H_{a\alpha,x,c \alpha,x}(x)$, where $q^N_{a\alpha,x} = (a^N_{\alpha,x}, (c^N_{\alpha,x})') = \theta q_{\alpha,x} + (1 - \theta) \tilde{q}_{a\alpha,x}$, for some $\theta \in (0,1)$. Write $H^N_{x} = O_{x} \Lambda^N_{x} O'_{x}$, where $O_{x}$ is an $m \times m$ orthogonal matrix and where $\Lambda^N_{x} = \text{diag}(\lambda^N_{1,x}, \ldots, \lambda^N_{m,x})$ collects the eigenvalues of $H^N_{x}$ in decreasing order. We then have

$$
G_{\tilde{a},\tilde{c}}(\tilde{x}) - G_{a,c}(x) = \frac{1}{2} (\tilde{q}_{a\alpha,x} - q_{a\alpha,x})' H^N_{x}(x) (\tilde{\tilde{q}}_{a\alpha,x} - q_{a\alpha,x})
= \frac{1}{2} \sum_{j=1}^{m} \lambda^N_{j,x} \left( (O_{x}(q^N_{a\alpha,x} - q_{a\alpha,x}) \right)_j^2 \geq \frac{\lambda^N_{m,x}}{2} \sum_{i=1}^{m} \left( (O_{x}(q^N_{a\alpha,x} - q_{a\alpha,x}) \right)_j^2
= \frac{\lambda^N_{m,x}}{2} \left| O_{x}(\tilde{q}^N_{a\alpha,x} - q_{a\alpha,x}) \right|^2 \geq \frac{\lambda^N_{m,x}}{2} \left| \tilde{q}^N_{a\alpha,x} - q_{a\alpha,x} \right|^2.
$$

22
Hence,

\[
\sup_{x \in S_X} |\tilde{q}_{\alpha,x}^N - q_{\alpha,x}|^2 \leq 2 \left( \inf_{N \geq N_0} \inf_{x \in S_X} \lambda_{m,x}^N \right)^{-1} \sup_{x \in S_X} |G_{a,e}(x) - G_{a,c}(x)|.
\]  

(A.10)

The result then follows from (A.9) and from the fact

\[
\inf_{N \geq N_0} \inf_{x \in S_X} \lambda_{m,x}^N = \inf_{N \geq N_0} \inf_{x \in S_X} \inf_{v \in S^{n-1}} v' H_{a,c}(x) = \sup_{x \in S_X} \inf_{v \in S^{m-1}} v' H_{a,c}(x) v > 0,
\]

which results from Lemma A.4(ii)-(iii) and the compactness of \(S_X, S^{m-1}\) and \(K_\alpha\). \(\square\)

B Proof of Theorem 3.1

Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be independent copies of \((X, Y)\). Recall that \(\gamma^N\) denotes an optimal quantization grid of size \(N\) for the random \(d\)-vector \(X\) and that \(\hat{\gamma}^{N,n}\) stands for the grid provided by the CLVQ algorithm on the basis of \(X_1, \ldots, X_n\). Below, we will write \((\hat{x}_1^N, \ldots, \hat{x}_N^N)\) and \((\tilde{x}_1^{N,n}, \ldots, \tilde{x}_N^{N,n})\) for the grid points of \(\gamma^N\) and \(\hat{\gamma}^{N,n}\), respectively.

Throughout this section, we assume that the empirical quantization of \(X\), based on \(X_1, \ldots, X_n\) converges almost surely towards the population one, i.e., \(\hat{X}^{N,n} = \text{Proj}_{\hat{\gamma}^{N,n}}(X) \rightarrow \hat{X}^N = \text{Proj}_{\hat{\gamma}^N}(X)\) almost surely as \(n \rightarrow \infty\). This is justified by classical results in quantization about the convergence in \(n\) of the CLVQ algorithm when \(N\) is fixed; see Pagès (1998).

The proof of Theorem 3.1 requires Lemmas B.1-B.2 below.

**Lemma B.1.** Let Assumption (A)' hold. Fix \(N \in \mathbb{N}_0\) and \(x \in S_X\). Write \(\bar{x} = \hat{x}^N = \text{Proj}_{\hat{\gamma}^N}(x)\) and \(\hat{x} = \hat{x}^{N,n} = \text{Proj}_{\hat{\gamma}^{N,n}}(x)\). Then, with \(\bar{X}_i = \hat{X}_i^N = \text{Proj}_{\hat{\gamma}^{N,n}}(X_i), i = 1, \ldots, n\), we have

(i) \(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{[\hat{X}_i^N = \hat{x}^N]} \xrightarrow{a.s.,n \rightarrow \infty} P[\hat{X}_i^N = \hat{x}]\);

(ii) after possibly reordering the \(\tilde{x}_i^{N,n}\)'s, \(\tilde{x}_i^{N,n} \xrightarrow{a.s.,n \rightarrow \infty} \tilde{x}_i^N, i = 1, \ldots, N\) (hence, \(\hat{\gamma}^{N,n} \xrightarrow{a.s.,n \rightarrow \infty} \gamma^N\)).

A proof is given in Charlier et al. (2015a).

**Lemma B.2.** Let Assumptions (A) and (B) hold. Fix \(\alpha = a\mathbf{u} \in B^m, x \in S_X\) and \(N \in \mathbb{N}_0\). Let \(K \subset \mathbb{R}^m\) be compact and define

\[
\tilde{G}_{a,e}(\hat{x}) = \frac{1}{n} \mathbb{E} \left[ u' Y_i - c' T_i u Y_i - a \right] [X_i^N = \hat{x}].
\]

Then

(i) \(\sup_{(a,c') \in K} |\tilde{G}_{a,e}(\hat{x}) - \tilde{G}_{a,e}(\hat{x})| = o_p(1)\) as \(n \rightarrow \infty\);
(ii) $|\min_{(a,c')\in \mathbb{R}^{m}} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c')\in \mathbb{R}^{m}} \tilde{G}_{a,c}(\tilde{x})| = o_p(1)$ as $n \to \infty$;

(iii) $|\tilde{G}_{a,N,a}(\tilde{x}) - \tilde{G}_{a,N,a}(\tilde{x})| = o_p(1)$ as $n \to \infty$.

Proof. (i) Since

$$\tilde{G}_{a,c}(\tilde{x}) = E[\rho_{a}(u'Y - c'\Gamma_{u}Y - a)|X^{N} = \bar{x}] = \frac{E[\rho_{a}(u'Y - c'\Gamma_{u}Y - a)]_{X^{N} = \bar{x}}}{P[X^{N} = \bar{x}]}$$

it is sufficient, in view of Lemma B.1(i), to show that

$$\sup_{(a,c')\in K} \left| \frac{1}{n} \sum_{i=1}^{n} \rho_{a}(u'Y_{i} - c'\Gamma_{u}Y_{i} - a)I[X_{i}^{N} = \bar{x}] - E[\rho_{a}(u'Y - c'\Gamma_{u}Y - a)I[X^{N} = \bar{x}] \big| X^{N} = \bar{x}] \right| = o_p(1),$$

as $n \to \infty$. It is natural to decompose it as

$$\sup_{(a,c')\in K} \left| \frac{1}{n} \sum_{i=1}^{n} \rho_{a}(u'Y_{i} - c'\Gamma_{u}Y_{i} - a)I[X_{i}^{N} = \bar{x}] - E[\rho_{a}(u'Y - c'\Gamma_{u}Y - a)I[X^{N} = \bar{x}] \big| X^{N} = \bar{x}] \right| \leq \sup_{(a,c')\in K} |T_{ac1}| + \sup_{(a,c')\in K} |T_{ac2}|,$$

with

$$T_{ac1} := \frac{1}{n} \sum_{i=1}^{n} \rho_{a}(u'Y_{i} - c'\Gamma_{u}Y_{i} - a)(I[X_{i}^{N} = \bar{x}] - I[X^{N} = \bar{x}]),$$

and

$$T_{ac2} := \frac{1}{n} \sum_{i=1}^{n} \rho_{a}(u'Y_{i} - c'\Gamma_{u}Y_{i} - a)I[X^{N} = \bar{x}] - E[\rho_{a}(u'Y - c'\Gamma_{u}Y - a)I[X^{N} = \bar{x}] \big| X^{N} = \bar{x}],$$

with $X_{i}^{N} = \text{Proj}_{\gamma}(X_{i}, i = 1, \ldots, n)$.

We start by considering $T_{ac2}$. Since $x \mapsto M_{a,x}$ and $x \mapsto M_{b,x}$ are continuous functions defined over the compact set $S_{X}$, one has that, for all $(a,c')' \in K$,

$$\rho_{a}(u'Y - c'\Gamma_{u}Y - a)I[X^{N} = \bar{x}] \leq \max(\alpha, 1 - \alpha)|u'Y - c'\Gamma_{u}Y - a| \leq \max(\alpha, 1 - \alpha)|(u - \Gamma_{u}c)'Y - a|$$

$$\leq \max(\alpha, 1 - \alpha)\left[|u - \Gamma_{u}c| \left(\sup_{x \in S_{X}} |M_{a,x}| + |\varepsilon| \sup_{x \in S_{X}} ||M_{b,x}|| \right) + |a| \right]$$

$$\leq C_{1}|\varepsilon| + C_{2}, \quad \text{(B.1)}$$

for some constants $C_{1}, C_{2}$ that do not depend on $(a,c')'$. Since Assumption (A)(v) ensures that $E[|\varepsilon|] < +\infty$ (recall that $p = 2$ here), the uniform law of large numbers (see, e.g., Theorem 16(a) in Ferguson, 1996) then implies that

$$\sup_{(a,c')\in K} |T_{ac2}| = o_p(1), \quad \text{as } n \to \infty. \quad \text{(B.2)}$$
It remains to treat $T_{ac1}$. Let $\ell_n := \{ i = 1, \ldots, n : \mathbb{I}[x_i = \hat{x}] \neq \mathbb{I}[\tilde{x}_i = \hat{x}] \}$ be the set collecting the indices of observations that are projected on the same point as $x$ for $\gamma^N$ but not for $\hat{\gamma}^{N,n}$, or on the same point as $\hat{x}$ for $\hat{\gamma}^{N,n}$ but not for $\gamma^N$. Proceeding as in (B.1) then shows that, for any $(a, c')' \in K$,
\[
|T_{ac1}| \leq \frac{1}{n} \sum_{i \in \ell_n} |\rho_0(u'Y_i - c'\Gamma_uY_i - a)| \leq \frac{\#\ell_n}{n} \sum_{i \in \ell_n} (C_1 + C_2|\varepsilon_i|) =: S_1 \times S_2,
\]
say. Lemma B.1(ii) implies that $S_1 = o_p(1)$ as $n \to \infty$. Regarding $S_2$, the independence between $\#\ell_n$ and the $\varepsilon_i$'s (which follows from the fact that $\#\ell_n$ is measurable with respect to the $X_i$'s) entails that $E[S_2] = O(1)$ as $n \to \infty$, hence that $S_2 = O_P(1)$ as $n \to \infty$. Therefore,
\[
\sup_{(a, c')' \in K} |T_{ac1}| \leq S_1S_2 = o_p(1) \quad \text{as } n \to \infty,
\]
which, jointly with (B.2), establishes the result.

(ii) For simplicity, we write $\hat{\mathbf{q}} = (\hat{a}, \hat{c}')'$ and $\tilde{\mathbf{q}} = (\hat{a}, \hat{c}')'$ instead of $\hat{q}_{a,x}^N = (\hat{a}_{a,x}, \hat{c}_{a,x}')'$ and $\tilde{q}_{a,x}^{N,n} = (\tilde{a}_{a,x}, \tilde{c}_{a,x}'))'$, respectively. First fix $\delta > 0$ and $\eta > 0$, and choose $n_1$ and $R$ large enough to have $|\tilde{q}| \leq R$ and $P[|\tilde{q}| > R] < \eta/2$ for any $n \geq n_1$. This is possible since $\hat{\mathbf{q}}$ is nothing but the sample Hallin et al. (2010) quantile of a number of $Y_i$'s that increases to infinity (so that, with arbitrary large probability for $n$ large, $|\tilde{q}|$ cannot exceed $2\sup_{x \in S_X} |q_{a,x}|$). Define $K_R := \{ y \in \mathbb{R}^m : |y| \leq R \}$. Then, with $\mathbb{I}_+ := \{ \min_{(a, c')' \in \mathbb{R}^m} \hat{G}_{a,c}(\hat{x}) \geq \min_{(a, c')' \in \mathbb{R}^m} \tilde{G}_{a,c}(\hat{x}) \}$, we have
\[
\left| \min_{(a, c')' \in \mathbb{R}^m} \hat{G}_{a,c}(\hat{x}) \right| - \min_{(a, c')' \in \mathbb{R}^m} \tilde{G}_{a,c}(\hat{x}) \mathbb{I}_+ \leq (\hat{G}_{a,c}(\hat{x}) - \tilde{G}_{a,c}(\hat{x})) \mathbb{I}_+, \tag{B.3}
\]
for $n \geq n_1$. Similarly, with $\mathbb{I}_- := 1 - \mathbb{I}_+$, we have that, under $|\tilde{q}| \leq R$,
\[
\left| \min_{(a, c')' \in \mathbb{R}^m} \hat{G}_{a,c}(\hat{x}) \right| - \min_{(a, c')' \in \mathbb{R}^m} \tilde{G}_{a,c}(\hat{x}) \mathbb{I}_- \leq (\hat{G}_{a,c}(\hat{x}) - \tilde{G}_{a,c}(\hat{x})) \mathbb{I}_-, \tag{B.4}
\]
still for $n \geq n_1$. By combining (B.3) and (B.4), we obtain that, under $|\tilde{q}| \leq R$,
\[
\left| \min_{(a, c')' \in \mathbb{R}^m} \hat{G}_{a,c}(\hat{x}) \right| - \min_{(a, c')' \in \mathbb{R}^m} \tilde{G}_{a,c}(\hat{x}) \leq \sup_{(a, c')' \in K_R} |\hat{G}_{a,c}(\hat{x}) - \tilde{G}_{a,c}(\hat{x})|,
\]
for $n \geq n_1$. Therefore, for any such $n$, we get
\[
P \left[ \min_{(a, c')' \in \mathbb{R}^m} \hat{G}_{a,c}(\hat{x}) - \min_{(a, c')' \in \mathbb{R}^m} \tilde{G}_{a,c}(\hat{x}) > \delta \right] \leq P \left[ \min_{a,c} \hat{G}_{a,c}(\hat{x}) - \min_{a,c} \tilde{G}_{a,c}(\hat{x}) > \delta, |\tilde{q}| \leq R \right] + P[|\tilde{q}| > R]
\]
\[
\leq P \left[ \sup_{(a, c')' \in K_R} |\hat{G}_{a,c}(\hat{x}) - \tilde{G}_{a,c}(\hat{x})| > \delta \right] + \frac{\eta}{2}.
\]
From Part (i) of the lemma, we conclude that, for \( n \) large enough,
\[
P\left[ \left| \min_{(a,c') \in \mathbb{R}^m} \tilde{G}_{a,c}(\tilde{x}) - \min_{(a,c') \in \mathbb{R}^m} \tilde{G}_{a,c}(\tilde{x}) \right| > \delta \right] < \eta',
\]
as was to be shown.

(iii) This proof proceeds in the same way as for (ii). We start with picking \( N_1 \) and \( R \) large enough so that \( P[|\hat{q}| > R] < \eta/2 \) for any \( N \geq N_1 \), with \( \eta \) fixed. This yields
\[
P\left[ |\tilde{G}_{a,c}(\tilde{x}) - \tilde{G}_{a,c}(\tilde{x})| > \delta, |\hat{q}| \leq M \right] + \frac{\eta}{2} \tag{B.5}
\]
Note then that
\[
P\left[ |\tilde{G}_{a,c}(\tilde{x}) - \tilde{G}_{a,c}(\tilde{x})| > \delta, |\hat{q}| \leq M \right] \\
\leq P\left[ |\tilde{G}_{a,c}(\tilde{x}) - \tilde{G}_{a,c}(\tilde{x})| > \delta/2, |\hat{q}| \leq M \right] + P\left[ |\tilde{G}_{a,c}(\tilde{x}) - \tilde{G}_{a,c}(\tilde{x})| > \delta/2, |\hat{q}| \leq M \right] \\
\leq P\left[ \sup_{(a,c') \in K_R} |\tilde{G}_{a,c}(\tilde{x}) - \tilde{G}_{a,c}(\tilde{x})| > \delta/2 \right] + P\left[ \min_{a,c} \tilde{G}_{a,c}(\tilde{x}) - \min_{a,c} \tilde{G}_{a,c}(\tilde{x}) > \delta/2 \right] \\
=: p_1^{(n)} + p_2^{(n)},
\]
say. Parts (i) and (ii) of the lemma imply that \( p_1^{(n)} \) and \( p_2^{(n)} \) can be made arbitrarily small for \( n \) large enough. Combining this with (B.5) yields the result.

We can now prove Theorem 3.1.

**Proof of Theorem 3.1.** Under the assumptions considered, the function \( (a,c')' \mapsto \tilde{G}_{a,c}(\tilde{x}) \) has a unique minimizer (that is the Hallin et al. (2010) \( \alpha \)-quantile of the distribution of \( Y \) conditional on \( \tilde{X}^N = \tilde{x} \)). Therefore, the convergence in probability of \( \tilde{G}_{\tilde{a},\tilde{c}}^{N,n}(\tilde{x}) \) to \( \tilde{G}_{\tilde{a},\tilde{c}}(\tilde{x}) \) (Lemma B.2(iii)) implies the convergence in probability of the corresponding arguments (note indeed that the function \( (a,c')' \mapsto \tilde{G}_{a,c}(\tilde{x}) \) does not depend on \( n \)).

**References**


