Multivariate Quantiles and Multiple-Output Regression Quantiles: from L1 Optimization to Halfspace Depth

Marc Hallin
ECARES, Institut de Recherche en Statistique and
Département de Mathématique, Université Libre de Bruxelles

Davy Paindaveine
ECARES, Institut de Recherche en Statistique and
Département de Mathématique, Université Libre de Bruxelles

Miroslav Siman
ECARES, Université Libre de Bruxelles

ECARES working paper 2008_042

ECARES
ULB - CP 114
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

MULTIVARIATE QUANTILES AND MULTIPLE-OUTPUT REGRESSION QUANTILES: FROM L_1 OPTIMIZATION TO HALFSPACE DEPTH

By Marc Hallin[1,2,3], Davy Paindaveine[3,4], and Miroslav Šiman[4]

Université Libre de Bruxelles

A new multivariate concept of quantile, based on a directional version of Koenker and Bassett's traditional regression quantiles, is introduced for multivariate location and multiple-output regression problems. In their empirical version, those quantiles can be computed efficiently via linear programming techniques. Consistency, Bahadur representation and asymptotic normality results are established. Most importantly, the contours generated by those quantiles are shown to coincide with the classical halfspace depth contours associated with the name of Tukey. This relation does not only allow for efficient depth contour computations by means of parametric linear programming, but also for transferring from the quantile to the depth universe such asymptotic results as Bahadur representations. Finally, linear programming duality opens the way to promising developments in depth-related multivariate rank-based inference.

1. Introduction: Multivariate quantiles and statistical depth. A huge literature has been devoted to the problem of extending to a multivariate setting the fundamental one-dimensional concept of quantile; we refer to [31] for a recent survey and references. An equally huge literature—see [24], [36], and [37] for a comprehensive account—is dealing with the concept of (location) depth. The philosophies

Received 1 January 1; revised 1 January 1; accepted 1 January 1.

[1]Académie Royale de Belgique.

[2]Part of this work was completed while Marc Hallin was visiting the Economics Department at the European University Institute in Florence under a Fernand Braudel Fellowship; the hospitality, stimulating research environment and financial support of EUI are gratefully acknowledged.

[3]Marc Hallin and Davy Paindaveine are also members of ECORE, the recently created association between CORE and ECARES.

[4]Supported by a Mandat d'Impulsion Scientifique of the Belgian Fonds National de la Recherche Scientifique.

AMS 2000 subject classifications: Primary 62H05; secondary 62J05

Keywords and phrases: Multivariate quantiles, Quantile regression, Halfspace depth

underlying those two concepts at first sight are quite different, and even, to some extent, opposite. While quantiles resort to analytical characterizations through inverse distribution functions or L_1 optimization, depth often derives from more geometric considerations such as halfspaces, simplices, ellipsoids, and projections. Both carry advantages and some drawbacks. Analytical definitions usually bring in efficient algorithms and tractable asymptotics. The geometric ones enjoy attractive equivariance properties and intuitive contents, but their probabilistic study and asymptotics are generally trickier, while their implementation, as a rule, leads to heavy combinatorial algorithms; a highly elegant analytical approach to depth has been proposed in [25], but does not help much in that respect.

Yet, beyond those methodological differences, quantiles and depth obviously exhibit close notional kinship. In the univariate case, all definitions basically agree that the depth of a point $x \in \mathbb{R}$ with respect to a probability distribution P with strictly monotone distribution function F should be $\min(F(x), 1 - F(x))$, so that the only points with depth d are $x_d = F^{-1}(d)$ and $x_{1-d} = F^{-1}(1-d)$—the quantiles of orders d and $1 - d$, respectively. Starting with dimension two, no such clear and undisputable relation has been established so far—how could there be one, by the way, as long as no clear and undisputable definition of a multivariate quantile has been agreed upon? Bridging the gap between the two concepts thus would allow for transferring to the depth universe the analytical and algorithmic tools of the quantile approach, while sorting out the many candidates for a sound definition of multivariate quantiles. Therefore, establishing a relation between the quantile and depth philosophies in $\mathbb{R}^k$, if at all possible, is highly desirable.

An important step in that direction has been made very recently in a paper by Kong and Mizera ([22]), based on a directional definition of quantiles (as in, e.g., [1], [3], [4], [6], [8], [11], [21], [32], and [34]). In the Kong and Mizera approach, a quantile of order $\tau \in (0, 1)$, is a point $\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}} \in \mathbb{R}^k$ associated with a direction $\mathbf{u}$, that is, with a point $\mathbf{u}$ on the unit sphere $\mathcal{S}^{k-1}$ (see (2.4) for a precise definition); for given τ, these quantiles naturally yield (see also [34]) contours $\{\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}} : \mathbf{u} \in \mathcal{S}^{k-1}\}$. Those quantile contours, however, are hardly satisfactory, as the authors themselves admit. They lack any reasonable form of equivariance, even with respect to translation, heavily depend on the choice of an origin, and moreover exhibit disturbing *mozzarella shapes*, with a perverse tendency to self-intersection. Finally, the construction of any of them, in principle, involves computing infinitely many univariate quantiles—one for each $\mathbf{u} \in \mathcal{S}^{k-1}$—which of course is impossible in practice. However, associating with each quantile $\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}}$ the hyperplane $\pi_{\mathrm{KM};\tau\mathbf{u}}$ running through $\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}}$ and orthogonal to $\mathbf{u}$, and considering, for fixed τ, their envelope over $\mathbf{u} \in \mathcal{S}^{k-1}$ (see (2.4) and (4.5) for a precise definition of those envelopes) yields regions that happen to coincide—in the population as well as in the empirical case (with obvious notation $\mathbf{q}^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$ and $\pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$)—with the celebrated halfspace depth regions associated with the name of Tukey ([33]).

A relation is thus established between Kong and Mizera's directional quantiles and Tukey's depth, providing a conceptual bridge between the two concepts. That relation provides halfspace depth regions with an interesting quantile interpretation: any given face of a polyhedral empirical Tukey contour of order τ indeed coincides with the τ-quantile hyperplane associated with the corresponding orthogonal direction $\mathbf{u}$. While allowing the quantile philosophy to shed some interpretational light on depth concepts, this unfortunately does not help much in terms of practical computation of empirical contours. For given τ, indeed, there exist infinitely many quantile hyperplanes $\pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$, of which only a finite number contribute to the depth contour of order τ. Unless depth contours are obtained from some other source, the Kong and Mizera approach does not provide any data-based method for identifying or constructing the "effective" $\pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$'s. In particular, it does not allow for importing into the depth universe any of the convenient L_1 or linear programming features of quantiles. Since computing the empirical version of any of their envelopes in principle requires considering all $\mathbf{u}$'s in $\mathcal{S}^{k-1}$, Kong and Mizera suggest sampling $\mathcal{S}^{k-1}$; this however only can yield approximate Tukey regions (the probability, in such sampling, of an actual Tukey contour face being recovered is zero). Summing up, the relation between the Kong and Mizera quantiles and the Tukey depth contours, although conceptually quite satisfactory, does not yield the computational benefits expected from such relation.

Now, Kong and Mizera's directional quantiles $\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}}$ are points in the observation space $\mathbb{R}^k$ while, being orthogonal to $\mathbf{u}$, the associated hyperplanes $\pi_{\mathrm{KM};\tau\mathbf{u}}$ do not carry any additional information. If, instead of points, directional quantiles themselves are characterized as hyperplanes $\pi_{\tau\mathbf{u}}$, but in the Koenker and Bassett L_1 sense (contrary to Kong and Mizera's $\pi_{\mathrm{KM};\tau\mathbf{u}}$'s, these hyperplanes, in general, are not orthogonal to $\mathbf{u}$, hence do not coincide with the $\pi_{\mathrm{KM};\tau\mathbf{u}}$ ones), all the problems faced by the Kong and Mizera construction almost miraculously vanish. The collections $\{\pi_{\tau\mathbf{u}} : \mathbf{u} \in \mathcal{S}^{k-1}\}$ indeed do not depend on the origin, are affine-equivariant, and yield central regions which strictly coincide with the Tukey depth ones. In the empirical case (with obvious notation $\pi_{\tau\mathbf{u}}^{(n)}$), those collections, contrary to their Kong and Mizera counterparts, only contain finitely many hyperplanes (the mapping $\mathbf{u} \mapsto \pi_{\tau\mathbf{u}}^{(n)}$ is piecewise constant on $\mathcal{S}^{k-1}$), among which all those containing a face of the (polyhedral) empirical Tukey contours. From their depth contour nature, these $\pi_{\tau\mathbf{u}}^{(n)}$'s inherit a series of nice properties such as convexity, nestedness, and affine-equivariance. From their definition as Koenker-Bassett quantiles, they receive a probabilistic interpretation (allowing for tractable asymptotics : consistency, Bahadur representations, and asymptotic normality), but, above all, the important benefits of linear programming algorithms, which thereby automatically transfer to depth. Moreover, they readily generalize to the regression setting, yielding polyhedral "hypertubes" wrapping (up to the traditional quantile crossings) a median or deepest regression region, thus extending to the multiple-output context the concept of regression quantiles. A constrained optimization form of the definition also allows for computing Lagrange multipliers with most interesting statistical applications. Finally, by resorting to classical linear programming duality, a concept of directional regression rank score, allowing for multivariate versions of the methods developed in [13], naturally comes into the picture.

From an applied perspective, the possibility of computing Tukey depth contours via parametric linear programming is not a small improvement. To the best of our knowledge, implementable algorithms so far are strictly limited to the two-dimensional case ($k = 2$). Our approach allows for higher dimensions, and we could easily run our algorithms in dimension $k = 5$, for a few hundreds observations.

Some of the basic ideas of this approach to multivariate quantiles were exposed in an unpublished master thesis by Laine ([23]), quoted in [18]. In this paper, we carefully revive Laine's ideas, and systematically develop and prove the main properties of the concept he introduced.

The paper is organized as follows. Section 2 introduces the definitions and main notation to be used throughout. In Section 3, we study the main properties of the new quantiles : from their directional quantile nature, they inherit subgradient characterizations (Section 3.1), equivariance properties (Section 3.2), and quantile-like asymptotics—strong consistency, Bahadur representation, and asymptotic normality (Section 3.3). In Section 4, we establish the equivalence of the quantile contours thus obtained with more traditional halfspace (or Tukey) depth contours, as well as their relation to recent results by Kong and Mizera ([22]). Section 5 is devoted to the computational aspects of our multivariate quantiles, and Section 6 to their extension to tha multiple-output regression context. Section 7 concludes with some perspectives for future research. Proofs are collected in the Appendix.

2. Definition and notation. Consider the k-variate random vector $\mathbf{Z} := (Z_1, \ldots, Z_k)'$. The multivariate quantiles we are proposing are *directional* quantities—more precisely, $(k-1)$-dimensional hyperplanes indexed by non-zero vectors $\boldsymbol{\tau}$ ranging over the (open) unit ball (deprived of the origin) $\mathcal{B}^k := \{\mathbf{z} \in \mathbb{R}^k : 0 < \|\mathbf{z}\| < 1\}$ of $\mathbb{R}^k$. This directional index $\boldsymbol{\tau}$ naturally factorizes into $\boldsymbol{\tau} =: \tau\mathbf{u}$, where $\tau = \|\boldsymbol{\tau}\| \in (0,1)$ and $\mathbf{u} \in \mathcal{S}^{k-1} := \{\mathbf{z} \in \mathbb{R}^k : \|\mathbf{z}\| = 1\}$. Denoting by $\boldsymbol{\Gamma}_\mathbf{u}$ an arbitrary $k \times (k-1)$ matrix of unit vectors such that $(\mathbf{u} \vdots \boldsymbol{\Gamma}_\mathbf{u})$ constitutes an orthonormal basis of $\mathbb{R}^k$, we define the $\boldsymbol{\tau}$-quantile of $\mathbf{Z}$ as the regression τ-quantile hyperplane obtained (in the traditional Koenker and Bassett [20] sense) when regressing $\mathbf{Z}_\mathbf{u} := \mathbf{u}'\mathbf{Z}$ on the marginals of $\mathbf{Z}_\mathbf{u}^\perp := \boldsymbol{\Gamma}_\mathbf{u}'\mathbf{Z}$ and a constant term: the vector $\mathbf{u}$ therefore indicates the direction of the "vertical" axis in the regression, while $\boldsymbol{\Gamma}_\mathbf{u}$ simply provides an orthonormal basis of the vector space orthogonal to $\mathbf{u}$. More precisely, denoting by $x \mapsto \rho_\tau(x) := x(\tau - \mathbb{I}_{[x<0]})$ the usual τ-quantile *check function*, we adopt the following definition.

DEFINITION 2.1. *The $\boldsymbol{\tau}$-quantile of $\mathbf{Z}$ ($\boldsymbol{\tau} := \tau\mathbf{u} \in \mathcal{B}^k$) is any element of the collection $\Pi_{\boldsymbol{\tau}}$ of hyperplanes $\pi_{\boldsymbol{\tau}} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} = \mathbf{b}_{\boldsymbol{\tau}}'\boldsymbol{\Gamma}_\mathbf{u}'\mathbf{z} + a_{\boldsymbol{\tau}}\}$ such that*

$$(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}')' \in \operatorname*{argmin}_{(a,\mathbf{b}')' \in \mathbb{R}^k} \Psi_{\boldsymbol{\tau}}(a, \mathbf{b}), \quad \textit{where} \quad \Psi_{\boldsymbol{\tau}}(a, \mathbf{b}) := \mathrm{E}[\rho_\tau(\mathbf{Z}_\mathbf{u} - \mathbf{b}'\mathbf{Z}_\mathbf{u}^\perp - a)]. \tag{2.1}$$

This definition clearly generalizes the traditional univariate one. For $k = 1$, indeed, hyperplanes of dimension $k-1$ are simply points, $\mathcal{B}^k$ reduces to $(-1,0)\cup(0,1)$, and $\pi_{\boldsymbol{\tau}}$ to a "classical" quantile of order $1 - \|\boldsymbol{\tau}\|$ or $\|\boldsymbol{\tau}\|$, according as $\boldsymbol{\tau}$ is pointing to the left or to the right ($\mathbf{u} = -1$ or $\mathbf{u} = 1$).

Note that the quantile hyperplanes $\pi_{\boldsymbol{\tau}}$ and the "intercepts" $a_{\boldsymbol{\tau}}$ are well-defined in the sense that they only depend on $\boldsymbol{\tau}$, not on the coordinate system associated with the (arbitrary) choice of $\boldsymbol{\Gamma}_\mathbf{u}$. However, the "slope" coefficients $\mathbf{b}_{\boldsymbol{\tau}} = \mathbf{b}_{\boldsymbol{\tau}}(\boldsymbol{\Gamma}_\mathbf{u})$ do depend on $\boldsymbol{\Gamma}_\mathbf{u}$, a dependence we do not stress in the notation unless really necessary.

Each quantile hyperplane $\pi_{\boldsymbol{\tau}}$ (each element $(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}')'$ of $\operatorname{argmin}_{(a,\mathbf{b}')'\in\mathbb{R}^k} \Psi_{\boldsymbol{\tau}}(a, \mathbf{b})$) characterizes a lower (open) *quantile halfspace*

$$H_{\boldsymbol{\tau}}^- = H_{\boldsymbol{\tau}}^-(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} < \mathbf{b}_{\boldsymbol{\tau}}'\boldsymbol{\Gamma}_\mathbf{u}'\mathbf{z} + a_{\boldsymbol{\tau}}\} \tag{2.2}$$

and an upper (closed) *quantile halfspace*

$$H_{\boldsymbol{\tau}}^{+} = H_{\boldsymbol{\tau}}^{+}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} \geq \mathbf{b}_{\boldsymbol{\tau}}'\mathbf{\Gamma}_{\mathbf{u}}'\mathbf{z} + a_{\boldsymbol{\tau}}\}. \tag{2.3}$$

For the sake of comparison, we recall that the upper (closed) directional quantile halfspaces proposed in [22] are simply defined as

$$H_{\mathrm{KM};\tau\mathbf{u}}^{+} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} \geq \mathbf{u}'\mathbf{q}_{\mathrm{KM};\tau\mathbf{u}}\}, \;\; \mathbf{q}_{\mathrm{KM};\tau\mathbf{u}} := q_\tau(\mathbf{u}'\mathbf{Z})\mathbf{u}, \;\; \tau \in (0,1), \;\; \mathbf{u} \in \mathcal{S}^{k-1}, \tag{2.4}$$

where $q_\tau(X)$ stands for the univariate τ-quantile of the random variable X. Directional quantile hyperplanes $\pi_{\mathrm{KM};\tau\mathbf{u}}$ (which are orthogonal to $\mathbf{u}$) and lower halfspaces $H_{\mathrm{KM};\tau\mathbf{u}}^{-}$ are defined accordingly. These quantiles are also studied in [27].

Definition 2.1 requires $\mathbf{Z}$ to have finite first-order moments. Actually, modifying the definition into $(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}')' := \operatorname{argmin}_{(a,\mathbf{b}')' \in \mathbb{R}^k}(\Psi_{\boldsymbol{\tau}}(a, \mathbf{b}) - \Psi_{\boldsymbol{\tau}}(0, \mathbf{0}))$ has no impact on $\pi_{\boldsymbol{\tau}}$, while allowing to relax the moment condition on $\mathbf{Z}_{\mathbf{u}}$; finite first-order moments, however, still are required for $\mathbf{Z}_{\mathbf{u}}^{\perp}$. When $\mathbf{u}$ ranges over $\mathcal{S}^{k-1}$—for instance, when defining quantile contours—we need finite first-order moments for all $\mathbf{Z}_{\mathbf{u}}^{\perp}$'s, hence for $\mathbf{Z}$ itself. For the sake of simplicity, we often adopt the following assumption in the sequel.

Assumption (A). *The distribution of the random vector* $\mathbf{Z}$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^k$*, with a density* (f*, say) that has connected support, and admits finite first-order moments.*

The minimization problem (2.1) may have several solutions, yielding distinct hyperplanes $\pi_{\boldsymbol{\tau}}$. This, however, does not occur under Assumption (A), as shown in the following result, which is a particular case of Theorem 2.1 in [27].

Proposition 2.1. *Let Assumption (A) hold. Then, for any* $\boldsymbol{\tau} \in \mathcal{B}^k$*, the minimizer* $(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}')'$ *in (2.1), hence also the resulting quantile hyperplane* $\pi_{\boldsymbol{\tau}}$*, is unique.*

The family of hyperplanes $\Pi = \{\pi_{\boldsymbol{\tau}} : \boldsymbol{\tau} = \tau\mathbf{u} \in \mathcal{B}^k\}$ can be considered from two different points of view. The *directional* point of view, associated with the fixed-$\mathbf{u}$ subfamilies $\Pi_{\mathbf{u}} := \{\pi_{\boldsymbol{\tau}} : \boldsymbol{\tau} = \tau\mathbf{u}, \; \tau \in (0,1)\}$ is the one emphasized so far in the definition, and provides, for each $\mathbf{u}$, the usual interpretation of a collection of regression quantile hyperplanes. Another point of view is associated with the fixed-τ subfamilies $\Pi_\tau := \{\pi_{\boldsymbol{\tau}} : \boldsymbol{\tau} = \tau\mathbf{u}, \; \mathbf{u} \in \mathcal{S}^{k-1}\}$, which generate *quantile contours*: this point of view is developed in Section 4.

Before turning to the empirical version of our quantiles, let us present an alternative (but strictly equivalent) definition of $\pi_{\boldsymbol{\tau}}$, based on a *constrained* optimization formulation.

DEFINITION 2.2. *The $\boldsymbol{\tau}$-quantile of $\mathbf{Z}$ ($\boldsymbol{\tau} := \tau\mathbf{u} \in \mathcal{B}^k$) is any element of the collection $\Pi_{\boldsymbol{\tau}}$ of hyperplanes $\pi_{\boldsymbol{\tau}} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{z} = a_{\boldsymbol{\tau}}\}$ such that*

$$(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})' \in \underset{(a,\mathbf{c}')' \in \mathcal{M}_{\mathbf{u}}}{\operatorname{argmin}} \Psi^c_\tau(a, \mathbf{c}), \tag{2.5}$$

where $\Psi^c_\tau(a, \mathbf{c}) := \mathrm{E}[\rho_\tau(\mathbf{c}'\mathbf{Z} - a)]$ and $\mathcal{M}_{\mathbf{u}} := \{(a, \mathbf{c}')' \in \mathbb{R}^{k+1} : \mathbf{u}'\mathbf{c} = 1\}$.

Clearly, if $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ is a minimizer of (2.1), then $(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})' := (a_{\boldsymbol{\tau}}, (\mathbf{u} - \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{b}_{\boldsymbol{\tau}})')'$ minimizes the objective function in (2.5); conversely, for any minimizer $(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})'$ of (2.5), $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})' := (a_{\boldsymbol{\tau}}, (-\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{c}_{\boldsymbol{\tau}})')'$ minimizes the objective function in (2.1). The two definitions thus coincide; in particular, the lower and upper quantile halfspaces $\{\mathbf{z} \in \mathbb{R}^k : \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{z} < a_{\boldsymbol{\tau}}\}$ and $\{\mathbf{z} \in \mathbb{R}^k : \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{z} \geq a_{\boldsymbol{\tau}}\}$ associated with the quantile hyperplanes of Definition 2.2 coincide with those of (2.2)-(2.3), and therefore, depending on the context, the notation $H^{\pm}_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}})$, $H^{\pm}_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}})$, or simply $H^{\pm}_{\boldsymbol{\tau}}$ will be used indifferently. Definition 2.1 and Definition 2.2 both have advantages and, in the sequel, we use them both. Definition 2.1 is preferred in this section since it carries all the intuitive contents of our concept; the advantages of Definition 2.2, of an analytical nature, will appear more clearly in Sections 3.1 and 5.

The empirical versions of our quantile hyperplanes and the corresponding lower and upper quantile halfspaces naturally follow as sample analogs of the population concepts. To be more specific, let $\mathbf{Z}^{(n)} := (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ be an n-tuple ($n > k$) of k-dimensional random vectors: we define the *empirical $\boldsymbol{\tau}$-quantile* of $\mathbf{Z}^{(n)}$ as any element of the collection $\Pi^{(n)}_{\boldsymbol{\tau}}$ of hyperplanes $\pi^{(n)}_{\boldsymbol{\tau}} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} = \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}}\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{z} + a^{(n)}_{\boldsymbol{\tau}}\}$ such that (with obvious notation)

$$(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}})' \in \underset{(a,\mathbf{b}')' \in \mathbb{R}^k}{\operatorname{argmin}} \Psi^{(n)}_{\boldsymbol{\tau}}(a, \mathbf{b}), \quad \text{with} \quad \Psi^{(n)}_{\boldsymbol{\tau}}(a, \mathbf{b}) := n^{-1}\sum_{i=1}^n \rho_\tau(\mathbf{Z}_{i\mathbf{u}} - \mathbf{b}'\mathbf{Z}^{\perp}_{i\mathbf{u}} - a), \tag{2.6}$$

or equivalently, of hyperplanes $\pi^{(n)}_{\boldsymbol{\tau}} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}}\mathbf{z} = a^{(n)}_{\boldsymbol{\tau}}\}$ such that

$$(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}})' \in \underset{(a,\mathbf{c}')' \in \mathcal{M}_{\mathbf{u}}}{\operatorname{argmin}} \Psi^{c(n)}_{\boldsymbol{\tau}}(a, \mathbf{c}), \quad \text{with} \quad \Psi^{c(n)}_{\boldsymbol{\tau}}(a, \mathbf{c}) := n^{-1}\sum_{i=1}^n \rho_\tau(\mathbf{c}'\mathbf{Z}_i - a). \tag{2.7}$$

These empirical quantiles—which for given $\mathbf{u}$ clearly coincide with the Koenker and Bassett [20] hyperplanes in the coordinate system $(\mathbf{u} \vdots \boldsymbol{\Gamma}_{\mathbf{u}})$—allow for defining, in an obvious way, the empirical analogs $H_{\boldsymbol{\tau}}^{(n)-}$ and $H_{\boldsymbol{\tau}}^{(n)+}$ of the lower and upper quantile halfspaces in (2.2)-(2.3); see Figures 1 and 2 for an illustration (computational details are provided in Section 5).

Of course, empirical distributions are inherently discrete, so that empirical $\boldsymbol{\tau}$-quantiles and halfspaces in general are not uniquely defined. However, the minimizers of (2.6) (equivalently, of (2.7)), for given $\boldsymbol{\tau}$, are "close to each other", in the sense that the set of minimizers is convex—hence, connected (this readily follows from the fact that the objective functions are convex); this set is shrinking, as $n \to \infty$, to a single point which corresponds to the uniquely defined population quantile, provided that the following assumption is fulfilled (see the asymptotic results of Section 3.3 for details).

ASSUMPTION (A_n). *The observations $\mathbf{Z}_i$, $i = 1, \ldots, n$ are i.i.d. with a common distribution satisfying Assumption (A).*

Finally, note that, since the empirical versions of our quantiles, for given $\mathbf{u}$, are defined as standard single-output quantile regression hyperplanes, they inherit the linear programming features of the Koenker-Bassett theory. This certainly is one of the most important and attractive properties of the proposed quantiles; see Section 5 for details.

3. Multivariate quantiles as directional quantiles. In this section, we describe the "directional" properties of our quantiles. We first derive and discuss the subgradient conditions associated with the optimization problems (2.1) and (2.5), then state the strong equivariance properties of our empirical quantiles. Finally, asymptotic results are presented.

3.1. *Subgradient conditions.* Under Assumption (A), the objective function $\Psi_{\boldsymbol{\tau}}$ appearing in Definition 2.1 is convex and continuously differentiable on $\mathbb{R}^k$. Therefore, our population $\boldsymbol{\tau}$-quantiles can be equivalently defined as the collection of hyperplanes associated with the solutions $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ of the system of equations

$$\mathrm{grad}_{(a,\mathbf{b}')'}\, \Psi_{\boldsymbol{\tau}}(a, \mathbf{b}) = \mathbf{0}. \tag{3.1}$$

These hyperplanes thus are characterized by the relations

$$\begin{aligned} 0 = (\partial_a \Psi_\tau(a, \mathbf{b}))_{(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'} &= \mathrm{P}[\mathbf{u}'\mathbf{Z} < \mathbf{b}'_{\boldsymbol{\tau}} \boldsymbol{\Gamma}'_{\mathbf{u}} \mathbf{Z} + a_{\boldsymbol{\tau}}] - \tau \\ &= \mathrm{P}[\mathbf{Z} \in H^-_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}})] - \tau, \end{aligned} \tag{3.2a}$$

$$\mathbf{0} = (\mathrm{grad}_{\mathbf{b}} \Psi_{\boldsymbol{\tau}}(a, \mathbf{b}))_{(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'} = -\tau\, \mathrm{E}[\boldsymbol{\Gamma}'_{\mathbf{u}} \mathbf{Z}] + \mathrm{E}[\boldsymbol{\Gamma}'_{\mathbf{u}} \mathbf{Z}\, \mathbb{I}_{[\mathbf{Z} \in H^-_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}})]}]. \tag{3.2b}$$

Clearly, (3.2a) provides our multivariate $\boldsymbol{\tau}$-quantiles with a natural probabilistic interpretation, as it keeps the probability of their lower halfspaces equal to $\tau(= \|\boldsymbol{\tau}\|)$. As for (3.2b), it can be rewritten as

$$\boldsymbol{\Gamma}'_{\mathbf{u}} \left[\frac{1}{1-\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^+_{\boldsymbol{\tau}}]}] - \frac{1}{\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^-_{\boldsymbol{\tau}}]}] \right] = \mathbf{0}, \tag{3.3}$$

which—combined with (3.2a)—shows that the probability mass centers $\frac{1}{\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^-_{\boldsymbol{\tau}}]}]$ and $\frac{1}{1-\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^+_{\boldsymbol{\tau}}]}]$ of the lower and upper $\boldsymbol{\tau}$-quantile halfspaces share the same projection on the corresponding quantile hyperplanes $\pi_{\boldsymbol{\tau}}$—equivalently, that the straight line through those probability mass centers is parallel to $\mathbf{u}(= \boldsymbol{\tau}/\tau)$. Note moreover that, quite trivially,

$$(1-\tau) \left(\frac{1}{1-\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^+_{\boldsymbol{\tau}}]}] \right) + \tau \left(\frac{1}{\tau} \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^-_{\boldsymbol{\tau}}]}] \right) = \operatorname{E}[\mathbf{Z}]$$

so that the overall probability mass center also belongs to the same straight line.

Now consider the gradient conditions associated with Definition 2.2, which state that $(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}, \lambda_{\boldsymbol{\tau}})$ are solutions of the system

$$\operatorname{grad}_{(a,\mathbf{c},\lambda)} L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda) = \mathbf{0}, \quad \text{with } L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda) := \Psi^c_{\tau}(a, \mathbf{c}) - \lambda(\mathbf{u}'\mathbf{c} - 1) \tag{3.4}$$

(the Lagrangian function of the problem). Equivalently (indeed, the only points in $\mathbb{R}^{k+2}$ where $(a, \mathbf{c}, \lambda) \mapsto L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda)$ is not continuously differentiable are of the form $(0, \mathbf{0}', \lambda)'$, hence cannot be associated with a minimum of (2.5)), the latter gradient conditions rewrite

$$0 = (\partial_a L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda))_{(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}, \lambda_{\boldsymbol{\tau}})} = \operatorname{P}[\mathbf{c}'_{\boldsymbol{\tau}}\mathbf{Z} < a_{\boldsymbol{\tau}}] - \tau = \operatorname{P}[\mathbf{Z} \in H^-_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}})] - \tau, \tag{3.5a}$$

$$\mathbf{0} = (\operatorname{grad}_{\mathbf{c}} L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda))_{(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}, \lambda_{\boldsymbol{\tau}})} = \tau \operatorname{E}[\mathbf{Z}] - \operatorname{E}[\mathbf{Z}\, \mathbb{I}_{[\mathbf{Z}\in H^-_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}})]}] - \lambda_{\boldsymbol{\tau}}\mathbf{u}, \tag{3.5b}$$

$$0 = (\partial_\lambda L_{\boldsymbol{\tau}}(a, \mathbf{c}, \lambda))_{(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}, \lambda_{\boldsymbol{\tau}})} = 1 - \mathbf{u}'\mathbf{c}_{\boldsymbol{\tau}}. \tag{3.5c}$$

For such a constrained optimization problem, gradient conditions in general are necessary but not sufficient. In this case, however, note that premultiplying both sides of (3.5b) by $\boldsymbol{\Gamma}'_{\mathbf{u}}$ yields (3.2b), which clearly implies that, disregarding the Lagrange multiplier $\lambda_{\boldsymbol{\tau}}$ and (3.5c) to focus on (the coefficients of) the quantile hyperplane $\pi_{\boldsymbol{\tau}}$, the necessary conditions (3.5a)-(3.5b) are no weaker than the necessary and sufficient ones in (3.2a)-(3.2b), hence are necessary and sufficient, too.

Now, we argue that the gradient conditions (3.4) associated with Definition 2.2 are actually richer than those (3.1) associated with the original definition of our quantiles, which is actually one of the main reasons why we also consider that alternative definition. Indeed, (3.5b), which can be rewritten as

$$\frac{1}{1-\tau}\,\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in H_{\boldsymbol{\tau}}^{+}]}] - \frac{1}{\tau}\,\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in H_{\boldsymbol{\tau}}^{-}]}] = \frac{\lambda_{\boldsymbol{\tau}}}{\tau(1-\tau)}\mathbf{u}, \tag{3.6}$$

is clearly more informative than (3.2b)-(3.3), as it clarifies the role of the Lagrange multiplier $\lambda_{\boldsymbol{\tau}}$. Such a multiplier, which in general only measures the impact of the boundary constraint (in this case, constraint (3.5c)), here appears as a functional that is potentially useful for testing (central, elliptical, or spherical) symmetry or for measuring directional outlyingness and tail behavior of the distribution; see Section 7. Moreover, premultiplying (3.5b) with $\mathbf{c}_{\boldsymbol{\tau}}'$ yields $\lambda_{\boldsymbol{\tau}}(\mathbf{c}_{\boldsymbol{\tau}}'\mathbf{u}) = \mathrm{E}[(\tau - \mathbb{I}_{[\mathbf{c}_{\boldsymbol{\tau}}'\mathbf{Z}-a_{\boldsymbol{\tau}}<0]})\mathbf{c}_{\boldsymbol{\tau}}'\mathbf{Z}]$, that is, by using (3.5a) and (3.5c),

$$\lambda_{\boldsymbol{\tau}} = \Psi_{\tau}^{c}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}), \tag{3.7}$$

so that $\lambda_{\boldsymbol{\tau}}$ is nothing but the minimum achieved in (2.5) (equivalently, in (2.1)).

The sample objective functions $\Psi_{\boldsymbol{\tau}}^{(n)}(a,\mathbf{b})$ and $\Psi_{\boldsymbol{\tau}}^{c(n)}(a,\mathbf{c})$ in (2.6)-(2.7) are not continuously differentiable. They however have directional derivatives in all directions, which can be used to formulate fixed-$\mathbf{u}$ subgradient conditions for the empirical $\boldsymbol{\tau}$-quantiles, $\boldsymbol{\tau} = \tau\mathbf{u}$. Focusing first on the constrained optimization problem (2.7), it is easy to show that the coefficients $(a_{\boldsymbol{\tau}}^{(n)}, \mathbf{c}_{\boldsymbol{\tau}}^{(n)\prime})'$ and the corresponding Lagrange multiplier $\lambda_{\boldsymbol{\tau}}^{(n)}$ of any empirical $\boldsymbol{\tau}$-quantile $\pi_{\boldsymbol{\tau}}^{(n)} = \{\mathbf{z}\in\mathbb{R}^k : \mathbf{c}_{\boldsymbol{\tau}}^{(n)\prime}\mathbf{z} = a_{\boldsymbol{\tau}}^{(n)}\}$ must satisfy (letting $r_{i\boldsymbol{\tau}}^{(n)} := \mathbf{c}_{\boldsymbol{\tau}}^{(n)\prime}\mathbf{Z}_i - a_{\boldsymbol{\tau}}^{(n)}$, $i = 1,\ldots,n$)

$$n^{-1}\sum_{i=1}^{n}\mathbb{I}_{[r_{i\boldsymbol{\tau}}^{(n)}<0]} \le \tau \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{[r_{i\boldsymbol{\tau}}^{(n)}\le 0]}, \tag{3.8a}$$

$$\begin{aligned} -n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i^{-}\mathbb{I}_{[r_{i\boldsymbol{\tau}}^{(n)}=0]} \le \tau\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\right] \\ -\left[n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\mathbb{I}_{[r_{i\boldsymbol{\tau}}^{(n)}<0]}\right] - \lambda_{\boldsymbol{\tau}}^{(n)}\mathbf{u} \le \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i^{+}\mathbb{I}_{[r_{i\boldsymbol{\tau}}^{(n)}=0]}, \quad \text{and} \end{aligned} \tag{3.8b}$$

$$0 = 1 - \mathbf{u}'\mathbf{c}_{\boldsymbol{\tau}}^{(n)}, \tag{3.8c}$$

where $\mathbf{z}^{+} := (\max(z_1,0),\ldots,\max(z_k,0))'$ and $\mathbf{z}^{-} := (-\min(z_1,0),\ldots,-\min(z_k,0))'$. These necessary conditions are obtained by imposing that directional derivatives in each of the $2(k+2)$ semi-axial directions of the $(a,\mathbf{c}',\lambda)'$-space be nonnegative at $(a_{\boldsymbol{\tau}}^{(n)}, \mathbf{c}_{\boldsymbol{\tau}}^{(n)\prime}, \lambda_{\boldsymbol{\tau}}^{(n)})'$.

For $n \gg k$, we clearly may interpret (3.8a) and (3.8b) as an approximate version of their population analogs (3.5a) and (3.5b), roughly with the same consequences (condition (3.8c) simply restates our boundary constraint). More specifically, (3.8a) indicates that

$$\frac{N}{n} \leq \tau \leq \frac{N+Z}{n}, \quad \text{hence} \quad \frac{P}{n} \leq 1-\tau \leq \frac{P+Z}{n}, \tag{3.9}$$

where N, P, and Z are the numbers of negative, positive, and zero values, respectively, in the residual series $r^{(n)}_{i\boldsymbol{\tau}}$, $i = 1, \ldots, n$. This implies that, for non-integer values of $n\tau$, empirical $\boldsymbol{\tau}$-quantile hyperplanes have to go through some of the $\mathbf{Z}_i$'s. Actually, if the data points are in general position (which of course holds with probability one under Assumption (A_n)), there exists a sample $\boldsymbol{\tau}$-quantile hyperplane $\pi^{(n)}_{\boldsymbol{\tau}}$ which fits exactly k observations; (3.9) then holds with $Z = k$ (see Sections 2.2.1 and 2.2.2 of [18]). Note that the inequalities in (3.8a)-(3.8b) (hence also in (3.9)) must be strict if the sample $\boldsymbol{\tau}$-quantile is to be uniquely defined. Finally, as we will see in (5.2) below, the value of $\lambda^{(n)}_{\boldsymbol{\tau}}$, parallel to the population case, is the minimal one that can be achieved in (2.7), hence also in (2.6).

For the unconstrained definition of our empirical quantiles in (2.6), necessary and sufficient subgradient conditions can be obtained by applying Theorem 2.1 of [18], since (2.6) is nothing but a standard single-output quantile regression optimization problem. Assuming that the data points are in general position and defining, for any k-tuple of indices $h = (i_1, \ldots, i_k)$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$\mathbb{Y}_{\mathbf{u}}(h) := \mathbb{Z}'(h)\mathbf{u} \quad \text{and} \quad \mathbb{X}_{\mathbf{u}}(h) := (\mathbf{1}_k \,\vdots\, \mathbb{Z}'(h)\boldsymbol{\Gamma}_{\mathbf{u}}), \tag{3.10}$$

where $\mathbb{Z}(h) := (\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_k})$ and $\mathbf{1}_k = (1, \ldots, 1)' \in \mathbb{R}^k$, Koenker's result, in the present context, states that $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})' = (\mathbb{X}_{\mathbf{u}}(h))^{-1}\mathbb{Y}_{\mathbf{u}}(h)$ (we just pointed out that, under such conditions, there always exists a quantile hyperplane fitting exactly k observations) is a solution of (2.6) if and only if

$$-\tau\mathbf{1}_k \leq \boldsymbol{\xi}_{\boldsymbol{\tau}}(h) \leq (1-\tau)\mathbf{1}_k, \tag{3.11}$$

where

$$\boldsymbol{\xi}_{\boldsymbol{\tau}}(h) := (\mathbb{X}'_{\mathbf{u}}(h))^{-1} \sum_{i \notin h} (\tau - \mathbb{I}_{[r_i<0]}) \begin{pmatrix} 1 \\ \boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Z}_i \end{pmatrix}, \tag{3.12}$$

with $r_i := \mathbf{u}'\mathbf{Z}_i - \mathbf{b}'_{\boldsymbol{\tau}}\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Z}_i - a_{\boldsymbol{\tau}}$. Again, this solution is unique if and only if the inequalities in (3.11) are strict; see [18]. As for the constrained case, it follows from the linear programming theory that $(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})'$ are the coefficients of a $\boldsymbol{\tau}$-quantile hyperplane iff (3.12) holds with $r_i := \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{Z}_i - a_{\boldsymbol{\tau}}$ (still with a unique solution when the inequalities are strict).

We stress that no conditions (in particular, no moment conditions) are required here; only, the data points are assumed to be in general position.

3.2. *Equivariance properties.* In this short section, we study how the proposed $\boldsymbol{\tau}$-quantiles behave under various transformations—in particular, the affine ones—of the underlying distribution, and state the equivariance properties of the corresponding finite-sample concepts.

For the sake of simplicity, results for population quantiles here are stated under Assumption (A); more general statements can be derived, however, by simply taking into account—typically in the same fashion as in Theorem 3.1 below—the possible non-unicity of the resulting $\boldsymbol{\tau}$-quantiles (see Proposition 2.1). It is then easy to check that, with obvious notation, for any invertible $k \times k$ matrix $\mathbf{M}$ and any k-vector $\mathbf{d}$,

$$\pi_{\tau\mathbf{Mu}/\|\mathbf{Mu}\|}(\mathbf{MZ} + \mathbf{d}) = \mathbf{M}\,\pi_{\tau\mathbf{u}}(\mathbf{Z}) + \mathbf{d}. \tag{3.13}$$

In other words, the quantile hyperplane of $\mathbf{MZ}+\mathbf{d}$ with order τ in the direction given by $\mathbf{Mu}$ coincides with the image, under the affine transformation $\mathbf{z} \mapsto \mathbf{Mz} + \mathbf{d}$, of the quantile hyperplane of $\mathbf{Z}$ with the same order τ but for direction $\mathbf{u}$. In particular, for translations of $\mathbf{Z}$, we have

$$\pi_{\tau\mathbf{u}}(\mathbf{Z} + \mathbf{d}) = \pi_{\tau\mathbf{u}}(\mathbf{Z}) + \mathbf{d},$$

for any k-vector $\mathbf{d}$, which confirms that our concept of multivariate quantiles is *not localized* at any point of the k-dimensional Euclidean space; this was not so clear in Section 2 since the center of the unit sphere $\mathcal{S}^{k-1}$ (the origin of $\mathbb{R}^k$) seems to play an important role in their definitions). This is in sharp contrast with other directional quantile contours that are defined with respect to some location center, such as those of [22] (under the terminology *quantile biplots*) and [34].

Note that for any $\tau \in (0, 1)$ and any $\mathbf{u} \in \mathcal{S}^{k-1}$,

$$\pi_{(1-\tau)\mathbf{u}}(\mathbf{Z}) = \pi_{\tau(-\mathbf{u})}(\mathbf{Z}), \tag{3.14}$$

with the corresponding upper and lower quantile halfspaces exchanged: $\mathrm{int}H^{\pm}_{(1-\tau)\mathbf{u}}(\mathbf{Z}) = \mathrm{int}H^{\mp}_{\tau(-\mathbf{u})}(\mathbf{Z})$. Clearly, there is no general link between $\pi_{\tau(-\mathbf{u})}(\mathbf{Z})$ and $\pi_{\tau\mathbf{u}}(\mathbf{Z})$ unless the distribution of $\mathbf{Z}$ is centrally symmetric with respect to some k-vector $\boldsymbol{\theta}$.

Now, as shown by the following result, the sample versions of our quantiles are *equivariant*.

THEOREM 3.1. *For any* $\tau \in (0,1)$, $\mathbf{u} \in \mathcal{S}^{k-1}$, *invertible* $k \times k$ *matrix* $\mathbf{M}$ *and any* k*-vector* $\mathbf{d}$, *we have*

$$\Pi^{(n)}_{\tau \mathbf{Mu}/\|\mathbf{Mu}\|}(\mathbf{MZ}_1 + \mathbf{d}, \ldots, \mathbf{MZ}_n + \mathbf{d}) = \mathbf{M}\Big[\Pi^{(n)}_{\tau\mathbf{u}}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)\Big] + \mathbf{d},$$

$$\Pi^{(n)}_{\tau\mathbf{u}}(\mathbf{Z}_1 + \mathbf{d}, \ldots, \mathbf{Z}_n + \mathbf{d}) = \Pi^{(n)}_{\tau\mathbf{u}}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) + \mathbf{d},$$

and

$$\Pi^{(n)}_{(1-\tau)\mathbf{u}}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) = \Pi^{(n)}_{\tau(-\mathbf{u})}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n),$$

where equalities have to be understood as set equalities (see page 7 for the notation).

We skip the proof, which is straightforward.

3.3. *Asymptotic results.* This section derives, under Assumption (A_n) above, strong consistency, asymptotic normality, and Bahadur-type representation results for sample $\boldsymbol{\tau}$-quantiles and related quantities.

Under Assumption (A), the population $\boldsymbol{\tau}$-quantiles $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ and $(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})'$ always are uniquely defined (Proposition 2.1), unlike their sample counterparts $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}})'$ and $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}})'$; in the sequel, the latter notation will be used for arbitrary sequences of solutions to (2.6) and (2.7), respectively.

Strong consistency of our sample $\boldsymbol{\tau}$-quantiles, namely the fact that $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}})'$ converges to $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ almost surely as $n \to \infty$, holds under Assumption (A_n); this follows, e.g., from [14], Section 2.3. Asymptotic normality and Bahadur-type representation results, however, require stronger assumptions. More precisely, we will need the following reinforcement of Assumption (A_n).

ASSUMPTION (A'_n). *The observations* $\mathbf{Z}_i$, $i = 1, \ldots, n$ *are i.i.d. with a common distribution that is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^k$, *with a density* (f, *say*) *that has a connected support, and admits finite* second-*order moments. Moreover, there exist some constants* $r > k - 2$ *and* $C, s > 0$ *such that* f *satisfies*

$$|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leq C\|\mathbf{z}_1 - \mathbf{z}_2\|^s \left(1 + \left\|\frac{\mathbf{z}_1 + \mathbf{z}_2}{2}\right\|^2\right)^{-(3+r+s)/2},$$

for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^k$.

As we show in the Appendix (see the proof of Theorem 3.2), Assumption (A'_n) implies that the (strictly convex) function $(a, \mathbf{b}')' \mapsto \Psi_{\boldsymbol{\tau}}(a, \mathbf{b})$ (see Definition 2.1) is twice differentiable at $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$, with Hessian matrix

$$\begin{aligned}(3.15)\quad \mathbf{H}_{\boldsymbol{\tau}} &:= \int_{\mathbb{R}^{k-1}} \begin{pmatrix} 1 & \mathbf{x}' \\ \mathbf{x} & \mathbf{x}\mathbf{x}' \end{pmatrix} f((a_{\boldsymbol{\tau}} + \mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x})\mathbf{u} + \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{x})\, d\mathbf{x} \\ &= \mathbf{J}'_{\mathbf{u}} \left(\int_{\mathbf{u}^{\perp}} \begin{pmatrix} 1 & \mathbf{z}' \\ \mathbf{z} & \mathbf{z}\mathbf{z}' \end{pmatrix} f((a_{\boldsymbol{\tau}} - \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{z})\mathbf{u} + \mathbf{z})\, d\sigma(\mathbf{z}) \right) \mathbf{J}_{\mathbf{u}} =: \mathbf{J}'_{\mathbf{u}}\mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}},\end{aligned}$$

where $\mathbf{u}^{\perp} := \{\mathbf{z} \in \mathbb{R}^k : \mathbf{u}'\mathbf{z} = 0\}$ and $\mathbf{J}_{\mathbf{u}}$ denotes the $(k+1) \times k$ block-diagonal matrix with diagonal blocks 1 and $\boldsymbol{\Gamma}_{\mathbf{u}}$. Strict convexity of course guarantees that $\mathbf{H}_{\boldsymbol{\tau}}$ is positive semidefinite. The asymptotic results in Theorem 3.2 below furthermore require

Assumption ($\mathrm{B}_{\boldsymbol{\tau}}$). $\mathbf{H}_{\boldsymbol{\tau}}$ is positive definite.

Letting $\boldsymbol{\xi}_{i,\boldsymbol{\tau}}(a, \mathbf{b}) := -(\tau - \mathbb{I}_{[\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Z}_i - a < 0]})\, \dot{\mathbf{Z}}_i$ and $\boldsymbol{\xi}^c_{i,\boldsymbol{\tau}}(a, \mathbf{c}) := -(\tau - \mathbb{I}_{[\mathbf{c}'\mathbf{Z}_i - a < 0]})\, \dot{\mathbf{Z}}_i$, where $\dot{\mathbf{Z}}_i := (1, \mathbf{Z}'_i)'$, we have

$$\begin{aligned}\mathbf{V}_{\boldsymbol{\tau}} &:= \mathrm{Var}[\mathbf{J}'_{\mathbf{u}}\boldsymbol{\xi}_{1,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}})] \\ &= \mathbf{J}'_{\mathbf{u}} \begin{pmatrix} \tau(1-\tau) & \tau(1-\tau)\mathrm{E}[\mathbf{Z}'] \\ \tau(1-\tau)\mathrm{E}[\mathbf{Z}] & \mathrm{Var}[(\tau - \mathbb{I}_{[\mathbf{Z}_i \in H^-_{\boldsymbol{\tau}}]})\mathbf{Z}] \end{pmatrix} \mathbf{J}_{\mathbf{u}} \\ &= \mathbf{J}'_{\mathbf{u}}\mathrm{Var}[\boldsymbol{\xi}^c_{1,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}})]\mathbf{J}_{\mathbf{u}} =: \mathbf{J}'_{\mathbf{u}}\mathbf{V}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}}.\end{aligned}$$

We are then ready to state an asymptotic normality and Bahadur-type representation result for our sample $\boldsymbol{\tau}$-quantile coefficients, which is the main result of this section.

Theorem 3.2. *Let Assumptions (A'_n) and ($\mathrm{B}_{\boldsymbol{\tau}}$) hold. Then, as $n \to \infty$,*

$$(3.16)\qquad \sqrt{n} \begin{pmatrix} a^{(n)}_{\boldsymbol{\tau}} - a_{\boldsymbol{\tau}} \\ \mathbf{b}^{(n)}_{\boldsymbol{\tau}} - \mathbf{b}_{\boldsymbol{\tau}} \end{pmatrix} = -\frac{1}{\sqrt{n}} \mathbf{H}^{-1}_{\boldsymbol{\tau}} \mathbf{J}'_{\mathbf{u}} \sum_{i=1}^{n} \boldsymbol{\xi}_{i,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) + o_{\mathrm{P}}(1)$$

$$(3.17)\qquad \xrightarrow{\mathcal{L}} \mathcal{N}_k(\mathbf{0}, \mathbf{H}^{-1}_{\boldsymbol{\tau}}\mathbf{V}_{\boldsymbol{\tau}}\mathbf{H}^{-1}_{\boldsymbol{\tau}}).$$

Equivalently, with $\mathbf{P}_k = \mathrm{diag}(1, -\mathbf{I}_k)$,

$$(3.18)\qquad \sqrt{n} \begin{pmatrix} a^{(n)}_{\boldsymbol{\tau}} - a_{\boldsymbol{\tau}} \\ \mathbf{c}^{(n)}_{\boldsymbol{\tau}} - \mathbf{c}_{\boldsymbol{\tau}} \end{pmatrix} = -\frac{1}{\sqrt{n}} \mathbf{P}_k(\mathbf{H}^c_{\boldsymbol{\tau}})^{-} \sum_{i=1}^{n} \boldsymbol{\xi}^c_{1,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}) + o_{\mathrm{P}}(1)$$

$$(3.19)\qquad \xrightarrow{\mathcal{L}} \mathcal{N}_{k+1}(\mathbf{0}, \mathbf{P}_k(\mathbf{H}^c_{\boldsymbol{\tau}})^{-}\mathbf{V}^c_{\boldsymbol{\tau}}(\mathbf{H}^c_{\boldsymbol{\tau}})^{-}\mathbf{P}'_k),$$

where $(\mathbf{H}^c_{\boldsymbol{\tau}})^-$ *denotes the Moore-Penrose pseudoinverse of* $\mathbf{H}^c_{\boldsymbol{\tau}}$*. Moreover,*

$$\sqrt{n}\,(\lambda^{(n)}_{\boldsymbol{\tau}} - \lambda_{\boldsymbol{\tau}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\rho_\tau(\mathbf{c}'_{\boldsymbol{\tau}}\mathbf{Z}_i - a_{\boldsymbol{\tau}}) - \lambda_{\boldsymbol{\tau}}\right) + o_{\mathrm{P}}(1) \tag{3.20}$$

$$\xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Var}[\rho_\tau(\mathbf{c}'_{\boldsymbol{\tau}}\mathbf{Z}_1 - a_{\boldsymbol{\tau}})]). \tag{3.21}$$

As $\rho_\tau(\cdot)$ is a nonnegative function, the distribution of $\sqrt{n}\,(\lambda^{(n)}_{\boldsymbol{\tau}} - \lambda_{\boldsymbol{\tau}})$ is likely to be skewed for finite n (see (3.20)), which can be partly corrected via a normalizing transformation such as that from [9]. Also, the proof of the above theorem can be easily generalized to derive the asymptotic joint distribution of vectors of the form $(a^{(n)}_{\boldsymbol{\tau}_1}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}_1}, \ldots, a^{(n)}_{\boldsymbol{\tau}_J}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}_J})'$, $J \in \mathbb{N}_0$.

Theorem 3.2 of course paves the way to inference about $\boldsymbol{\tau}$-quantiles; in particular, it allows to build confidence zones for them. Testing linear restrictions on $\boldsymbol{\tau}$-quantiles coefficients—that is, testing null hypotheses of the form $\mathcal{H}_0 : (a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})' \in \mathcal{M}(a_0, \mathbf{b}_0, \boldsymbol{\Upsilon}) := \{(a_0, \mathbf{b}'_0)' + \boldsymbol{\Upsilon}\mathbf{v} : \mathbf{v} \in \mathbb{R}^\ell\}$ (indexed by some k-vector $(a_0, \mathbf{b}'_0)'$ and some full-rank $k \times \ell$ matrix $\boldsymbol{\Upsilon}$, $\ell < k$)—can be achieved in the same way as in [26]. Defining and studying such tests requires a detailed investigation of the asymptotic behavior of the constrained estimators

$$(\tilde{a}^{(n)}_{\boldsymbol{\tau}}, \tilde{\mathbf{b}}^{(n)\prime}_{\boldsymbol{\tau}})' := \underset{(a,\mathbf{b}')' \in \mathcal{M}(a_0,\mathbf{b}_0,\boldsymbol{\Upsilon})}{\mathrm{argmin}} \Psi^{(n)}_{\boldsymbol{\tau}}(a, \mathbf{b}),$$

which is beyond the scope of this work.

4. Multivariate quantiles as depth contours. Turning to the *contour* nature of our multivariate quantiles, we first define the (population and sample) quantile regions and contours that naturally follow from Definitions 2.1-2.2 and their empirical counterparts, and state their basic properties. We then establish the strong connections between those regions/contours and Tukey's classical *halfspace depth regions/contours*.

4.1. *Quantile regions.* The proposed quantile regions are obtained by taking, for some fixed $\tau(= \|\boldsymbol{\tau}\|)$, the "upper envelope" of our $\boldsymbol{\tau}$-quantile hyperplanes. More precisely, for any $\tau \in (0, 1)$, we define our τ-quantile region $R(\tau)$ as

$$R(\tau) := \bigcap_{\mathbf{u} \in \mathcal{S}^{k-1}} \bigcap \{H^+_{\tau\mathbf{u}}\} \tag{4.1}$$

where $\bigcap\{H^+_{\tau\mathbf{u}}\}$ stands for the intersection of the collection $\{H^+_{\tau\mathbf{u}}\}$ of all (closed) upper $(\tau\mathbf{u})$-quantile halfspaces (2.3); for $\tau = 0$, we simply let $R(\tau) := \mathbb{R}^k$. The corresponding τ-quantile contour then is defined as the boundary $\partial R(\tau)$ of $R(\tau)$. At this stage, it is already clear that those τ-quantile regions are *closed* and *convex* (since they are obtained by intersecting closed halfspaces). As we will see below, they are also *nested*: $R(\tau_1) \subset R(\tau_2)$ if $\tau_1 \geq \tau_2$.

Empirical quantile regions $R^{(n)}(\tau)$ are obtained by replacing in (4.1) the population quantile halfspaces $H^+_{\tau\mathbf{u}}$ with their sample counterparts $H^{(n)+}_{\tau\mathbf{u}}$, yielding, parallel to (4.1),

$$R^{(n)}(\tau) := \bigcap_{\mathbf{u}\in\mathcal{S}^{k-1}} \bigcap \{H^{(n)+}_{\tau\mathbf{u}}\}, \tag{4.2}$$

for any $\tau \in (0,1)$, with $R^{(n)}(0) := \mathbb{R}^k$. Since they result from intersecting *finitely many* halfspaces, these empirical quantile regions are *closed convex polyhedral* sets, the faces of which all are part of are quantile hyperplanes of order τ. Another important property of our sample regions, which readily follows from Theorem 3.1, is that, for any invertible $k \times k$ matrix $\mathbf{M}$ and any k-vector $\mathbf{d}$,

$$R^{(n)}(\tau; \mathbf{M}\mathbf{Z}_1 + \mathbf{d}, \ldots, \tau; \mathbf{M}\mathbf{Z}_n + \mathbf{d}) = \mathbf{M}R^{(n)}(\tau; \mathbf{Z}_1, \ldots, \mathbf{Z}_n) + \mathbf{d}.$$

As the population regions, in view of (3.13), satisfy $R(\tau; \mathbf{M}\mathbf{Z}+\mathbf{d}) = \mathbf{M}R(\tau; \mathbf{Z})+\mathbf{d}$ for any such $\mathbf{M}\mathbf{Z}$ and $\mathbf{d}$, the empirical regions (4.2) may be considered *affine-equivariant.*

4.2. *Connection with halfspace depth regions.* Recall that the *halfspace* or *Tukey depth* ([33]) of $\mathbf{z} \in \mathbb{R}^k$ with respect to the probability distribution P is defined as $HD(\mathbf{z}, \mathrm{P}) := \inf\{\mathrm{P}[H] : H \text{ is a closed halfspace containing } \mathbf{z}\}$. The halfspace depth region $D(\tau)$ of order $\tau \in [0,1]$ associated with P then collects all points of the k-dimensional Euclidean space with depth at least τ, that is,

$$D(\tau) = D_{\mathrm{P}}(\tau) := \{\mathbf{z} \in \mathbb{R}^k : HD(\mathbf{z}, \mathrm{P}) \geq \tau\}. \tag{4.3}$$

Clearly, $D(0) = \mathbb{R}^k$. Also, it can be shown (see, e.g., Proposition 6 in [29]) that, for any $\tau > 0$,

$$D(\tau) = \bigcap \{H : H \text{ is a closed halfspace with } \mathrm{P}[\mathbf{Z} \in H] > 1 - \tau\}. \tag{4.4}$$

The empirical version $D^{(n)}(\tau)$ of $D(\tau)$, as usual, is obtained by replacing, in (4.3) and (4.4), the probability measure P with the empirical measure associated with the observed n-tuple $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ at hand. As shown by the following results, the population halfspace depth regions, under Assumption (A), coincide with the quantile regions $R(\tau)$ defined in (4.1), and so do—almost surely under Assumption (A_n)—their empirical counterparts $D^{(n)}(\tau)$, whenever their interior is not empty, with the empirical quantile regions $R^{(n)}(\tau)$ (see the Appendix for the proofs).

THEOREM 4.1. *Under Assumption (A), $R(\tau) = D(\tau)$ for all $\tau \in [0, 1)$.*

THEOREM 4.2. *Assume that the $n(\geq k+1)$ data points are in general position. Then, for any $\ell \in \{1, 2, \dots, n-k\}$ such that $D^{(n)}(\frac{\ell}{n})$ has a non-empty interior, we have that $R^{(n)}(\tau) = D^{(n)}(\frac{\ell}{n})$ for all positive τ in $[\frac{\ell-1}{n}, \frac{\ell}{n})$.*

Theorem 4.1 of course implies that, under Assumption (A), all results on halfspace depth regions $D(\tau)$ also apply to the $R(\tau)$ regions. It follows that the $R(\tau)$'s are *compact*; the supremum of all τ's such that $R(\tau) \neq \emptyset$ belongs to $[\frac{1}{k+1}, \frac{1}{2}]$, and takes value $\frac{1}{2}$ iff the distribution of $\mathbf{Z}$ is *angularly symmetric* in the sense that there exists some k-vector $\boldsymbol{\theta}$ such that $\frac{\mathbf{Z}-\boldsymbol{\theta}}{\|\mathbf{Z}-\boldsymbol{\theta}\|}$ and $-\frac{\mathbf{Z}-\boldsymbol{\theta}}{\|\mathbf{Z}-\boldsymbol{\theta}\|}$ share the same distribution (see [5], [29], and [30]). This implies that, under Assumption (A), we also may restrict to $\tau \in [0, 1/2]$. As for Theorem 4.2, note that the restriction to halfspace depth regions with non-empty interiors is not really restrictive, since it only applies to the deepest regions.

Beyond that, Theorems 4.1 and 4.2, by showing that the halfspace depth regions coincide with the upper envelope of *directional* quantile halfspaces, and that the faces of the polyhedral empirical depth contours are parts of empirical quantile hyperplanes, provide depth contours with a straightforward quantile-based interpretation. Above all, these two theorems bring to the halfspace depth context the extremely efficient computational features of linear programming. This important issue is briefly discussed in Section 5; we refer to [28] for details. See Figure 3 for two- and three-dimensional illustrations.

Kong and Mizera in [22] establish somewhat similar results for the directional upper quantile halfspace $H^+_{\mathrm{KM};\tau\mathbf{u}}$ defined in (2.4). They show indeed that

$$D(\tau) = R_{\mathrm{KM}}(\tau) := \bigcap_{\mathbf{u} \in \mathcal{S}^{k-1}} \{H^+_{\mathrm{KM};\tau\mathbf{u}}\} \quad \text{for any } \tau \tag{4.5}$$

and that

$$D^{(n)}(\tfrac{\ell}{n}) = R^{(n)}_{\mathrm{KM}}(\tau) \quad \text{for any } \tau \in [\tfrac{\ell-1}{n}, \tfrac{\ell}{n}) \tag{4.6}$$

(see [22] and [27] for different proofs of this latter equality), where $R^{(n)}_{\mathrm{KM}}(\tau)$ stands for the empirical version of $R_{\mathrm{KM}}(\tau)$, obtained by replacing P with the empirical measure of a sample of size n.

These two results at first sight look completely equivalent to those of Theorems 4.1 and 4.2. They also establish a close connection between depth and directional quantiles—here, the Kong and Mizera ones. That connection, however, is much less exploitable than the former one. It does provide the faces of the polyhedral empirical depth regions $D^{(n)}(\tau)$ with a neat and interesting quantile interpretation: each face of $D^{(n)}(\tau)$ indeed is part of the Kong and Mizera quantile hyperplane $\pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}_0}$, where $\mathbf{u}_0$ stands for the unit vector orthogonal to that face and pointing to the interior of $D^{(n)}(\tau)$. Unless the depth region $D^{(n)}(\tau)$ is available from some other source, this is not really helpful, though. Contrary to the collection $\{\pi^{(n)}_{\tau\mathbf{u}}\}$, which is strictly finite, the collection $\{\pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}\}$, for fixed τ, contains infinitely many hyperplanes (one for each $\mathbf{u} \in \mathcal{S}^{k-1}$). And, since the definition of the upper envelopes of halfspaces $H^{(n)+}_{\mathrm{KM};\tau\mathbf{u}}$ involves an infinite number of such $H^{(n)+}_{\mathrm{KM};\tau\mathbf{u}}$'s, (4.6), contrary to Theorem 4.2, does not readily provide a feasible computation of $D^{(n)}(\tau)$. It is crucial to understand, in that respect, that our quantile halfspaces $H^{(n)+}_{\tau\mathbf{u}}$ are piecewise constant functions of $\mathbf{u}$, in sharp contrast with their Kong and Mizera counterparts $H^{(n)+}_{\mathrm{KM};\tau\mathbf{u}}$: since $\partial H^{+}_{\mathrm{KM};\tau\mathbf{u}}$ is orthogonal to $\mathbf{u}$ for any direction $\mathbf{u}$, there are uncountably many such upper halfspaces in any neighborhood of any fixed direction $\mathbf{u}$, even in the empirical case. To palliate this, Kong and Mizera ([22]) propose to sample the unit sphere $\mathcal{S}^{k-1}$, which leads to approximate envelopes, that only approximately satisfy (4.6); the same strategy is adopted in [34].

5. Computational aspects. In this section, we briefly discuss various computational issues related to the proposed quantiles; the reader is referred to [28] for details. We first restrict to the computation of (fixed-$\mathbf{u}$) directional quantiles and related quantities such as the corresponding Lagrange multipliers $\lambda^{(n)}_{\boldsymbol{\tau}}$ in (3.8b), then consider the computation of (fixed-τ) quantile contours.

5.1. *Computing directional quantiles.* As we have seen in the previous sections, the constrained formulation (2.5) of the definition of our directional quantiles is richer than the unconstrained one (2.1), since it introduces Lagrange multipliers, which bear highly relevant information (that can be exploited for statistical inference ; see Section 7). It is therefore natural to focus on the computation of the sample quantiles $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}})'$ in (2.7) first.

The problem of finding $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}})'$ can be reformulated as the linear program (P)

$$\min_{(a,\mathbf{c}',\mathbf{r}'_+,\mathbf{r}'_-)' \in \mathbb{R}\times\mathbb{R}^k\times\mathbb{R}^n\times\mathbb{R}^n} \tau\mathbf{1}'_n\mathbf{r}_+ + (1-\tau)\mathbf{1}'_n\mathbf{r}_-$$

subject to

$$\mathbf{u}'\mathbf{c} = 1, \quad \mathbb{Z}_n'\mathbf{c} - a\mathbf{1}_n - \mathbf{r}_+ + \mathbf{r}_- = \mathbf{0}, \quad \mathbf{r}_+ \geq \mathbf{0}, \quad \mathbf{r}_- \geq \mathbf{0}, \tag{5.1}$$

where we set $\mathbb{Z}_n := (\mathbf{Z}_1 \ldots \mathbf{Z}_n)$ and wrote $\mathbf{r}_\pm := ((r_1)_\pm, \ldots, (r_n)_\pm)'$, $r_+ := \max(r, 0)$ and $r_- := \max(-r, 0)$. Associated with problem (P) is the dual problem (D)

$$\max_{(\lambda_D, \boldsymbol{\mu}')' \in \mathbb{R}\times\mathbb{R}^n} \lambda_D, \text{ subject to } \mathbf{1}_n'\boldsymbol{\mu} = 0, \ \lambda_D\mathbf{u} + \mathbb{Z}_n\boldsymbol{\mu} = \mathbf{0}_m, \ -\tau\mathbf{1}_n \leq \boldsymbol{\mu} \leq (1-\tau)\mathbf{1}_n,$$

where λ_D and $\boldsymbol{\mu}$ are Lagrange multipliers corresponding to the first and second equality constraint in (5.1), respectively. Both (P) and (D) have at least one feasible solution (and therefore also an optimal one). This dual formulation leads to a natural multiple-output generalization of the powerful concept of regression rank scores introduced in [13], allowing for a depth-related form of rank-based inference in this context. This promising line of investigation is not considered here, and left for future research.

We need not worry about the possible non-unicity of the optimal solutions of (P) since, as we have seen in Section 3.3, any sequence of such solutions converges (under Assumption (A_n)) to the unique population coefficient vector $(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}')'$ almost surely as $n \to \infty$. In practice, one could compute $(a_{\boldsymbol{\tau}}^{(n)}, \mathbf{c}_{\boldsymbol{\tau}}^{(n)\prime})'$ by means of standard quantile regression of $\mathbf{0}_n$ on $(\mathbf{1}_n|\mathbb{Z}_n')$ with an extra pseudo-observation consisting of response C and corresponding design row $(0, C\mathbf{u}')$ for some sufficiently large constant C, which, in the limit, guarantees that the boundary constraint $\mathbf{u}'\mathbf{c}_{\boldsymbol{\tau}}^{(n)} = 1$ is satisfied; see [2] for another application of the same trick.

Now, since λ_D and $\lambda_{\boldsymbol{\tau}}^{(n)}$ are Lagrange multipliers associated with the same constraint, the optimal value λ_D of (D) satisfies

$$\lambda_D = n\lambda_{\boldsymbol{\tau}}^{(n)}$$

where, in view of (3.8b), $\lambda_{\boldsymbol{\tau}}^{(n)}$ has a clear meaning. Besides, due to the Strong Duality Theorem, the optimal values of the objective functions in (P) and (D) coincide. Therefore, $\lambda_{\boldsymbol{\tau}}^{(n)}$ is always unique and one has, with $\Psi_{\boldsymbol{\tau}}^{c(n)}$ defined in (2.7),

$$\lambda_{\boldsymbol{\tau}}^{(n)} = \Psi_{\boldsymbol{\tau}}^{c(n)}(a_{\boldsymbol{\tau}}^{(n)}, \mathbf{c}_{\boldsymbol{\tau}}^{(n)}) > 0 \tag{5.2}$$

(except for the rare case of exact fit where $\lambda_{\boldsymbol{\tau}}^{(n)} = 0$), which holds for all optimal solutions to (P) and (D). In other words, $\lambda_{\boldsymbol{\tau}}^{(n)}$ can be obtained from solving (P) as a by-product.

Most importantly, (5.2) allows us to focus on computing our $\boldsymbol{\tau}$-quantiles through the unconstrained problem (2.6) without any loss of generality because we may simply set $\lambda_{\boldsymbol{\tau}}^{(n)} = \Psi_{\boldsymbol{\tau}}^{(n)}(a_{\boldsymbol{\tau}}^{(n)}, \mathbf{b}_{\boldsymbol{\tau}}^{(n)})$. This approach is of course advantageous because it falls directly into the realm of quantile regression, as the problem of finding the sample $\boldsymbol{\tau}$-quantiles in (2.6) can be viewed as looking for standard—that is, single-output—regression quantiles in the regression of $\mathbf{Z}_{\mathbf{u}}$ on the marginals of $\mathbf{Z}_{\mathbf{u}}^{\perp}$ and a constant (in the notation of Section 2).

Needless to say, this interpretation has a large number of implications. Above all, it offers fast, powerful and sophisticated tools for computing sample $\boldsymbol{\tau}$-quantiles (along with the corresponding Lagrange multiplier $\lambda_{\boldsymbol{\tau}}^{(n)}$) in any fixed direction $\mathbf{u}$ and possibly for all τ's at once, with $\boldsymbol{\tau} = \tau\mathbf{u} = \|\boldsymbol{\tau}\|\mathbf{u}$ as usual. In particular, there is an excellent package for advanced quantile regression analysis in R (see [19]) and the key function for computing quantile regression estimates is also freely available for MATLAB, for example from Roger Koenker's homepage at *http://www.econ.uiuc.edu/~roger/research/rq/rq.html*.

5.2. *Computing quantile contours.* As the previous subsection shows that the computation of $H_{t\mathbf{u}}^{(n)+}$ is pretty straightforward, we now turn to the problem of how to aggregate efficiently, for fixed τ, the information associated with the various directional quantile halfspaces in order to compute the regions $R^{(n)}(\tau)$ defined in (4.2). The issue of course lies in the proper identification of the finite set of upper quantile halfspaces which are relevant for the computation of $R^{(n)}(\tau)$. Solving this problem, by finding *efficiently* solutions to (P) for all directions $\mathbf{u} \in \mathcal{S}^{k-1}$ and for any given $\tau \neq \frac{\ell}{n}, \ell \in \{0, 1, \ldots, n\}$ (in view of Theorem 4.2, we can restrict to such τ's without any loss of generality) can be achieved by turning to *parametric programming.*

For any fixed τ as above and under Assumption (A_n), parametric programming indeed reveals that $\mathbb{R}^k$ can be almost surely segmented into a finite number of non-degenerate cones $\mathcal{C}_i(\tau)$, $i = 1, 2, \ldots, N_C$, such that

$$\begin{aligned}
(a_{\tau\mathbf{u}}^{(n)}, \mathbf{c}_{\tau\mathbf{u}}^{(n)\prime}) &= (a_i, \mathbf{c}_i')/\mathbf{t}_i'\mathbf{u} \\
\lambda_{\tau\mathbf{u}}^{(n)} &= \lambda_i/\mathbf{t}_i'\mathbf{u} \\
\mu_{j,\tau\mathbf{u}}^{(n)} &= \begin{cases} \mathbf{v}_{ij}'\mathbf{u}/\mathbf{t}_i'\mathbf{u} \in [-\tau, 1-\tau] & \text{if } r_j = 0 \\ -\tau & \text{if } r_j > 0 \\ 1-\tau & \text{if } r_j < 0, \end{cases}
\end{aligned}$$

with $r_j := \mathbf{c}_j'\mathbf{Z}_j - a_j$, for any $\mathbf{u} \in \mathcal{C}_i(\tau) \cap \mathcal{S}^{k-1}$, $i = 1, 2, \ldots, N_C$ and $j = 1, \ldots, n$. Each cone $\mathcal{C}_i(\tau)$ then corresponds to one optimal basis $\mathbb{B}_i = \mathbb{B}_{i,\mathbf{u}}$ that uniquely determines constant scalars and vectors λ_i, a_i, $\mathbf{c}_i$, $\mathbf{v}_{ij}$, and $\mathbf{t}_i$ and guarantees that $\mathbf{t}_i'\mathbf{u} > 0$ for any $\mathbf{u} \in \mathcal{C}_i(\tau) \cap \mathcal{S}^{k-1}$. Consequently, each cone $\mathcal{C}_i(\tau)$ corresponds to exactly one quantile hyperplane, and any statistic $S_\mathbf{u}$ of the form

$$S_\mathbf{u} = g_1(\lambda_\mathbf{u}, a_\mathbf{u}, \mathbf{c}_\mathbf{u})/g_2(\lambda_\mathbf{u}, a_\mathbf{u}, \mathbf{c}_\mathbf{u})$$

is piecewise constant on the unit sphere whenever $g_1(\lambda, a, \mathbf{c})$ and $g_2(\lambda, a, \mathbf{c})$ are homogenous functions of the same order. Figure 4 provides such cones for a bivariate dataset.

It remains to note that we may investigate all the cones $\mathcal{C}_i(\tau)$'s by passing through them counter-clockwise when $k = 2$. In general, we can use the breadth-first search algorithm and always consider all such $\mathcal{C}_i(\tau)$'s that are adjacent to a cone treated in the previous step and have not been considered yet. If $\mathcal{C}_j(\tau)$ and $\mathcal{C}_i(\tau)$ are adjacent cones with point $\mathbf{u}_f$ inside their common facet, then $\mathbb{B}_{j,\mathbf{u}_f}$ (and consequently also $\mathbb{B}_{j,\mathbf{u}}$) may be found from the primal feasible basis $\mathbb{B}_{i,\mathbf{u}_f}$ by only a few iterations of the primal simplex algorithm at most.

Moreover, a careful reading of the proof of Theorem 4.2 reveals (see the remark right after the proof) that a single fixed-τ collection of quantile hyperplanes $\{\pi_{\tau\mathbf{u}}^{(n)} : \mathbf{u} \in \mathcal{S}^{k-1}\}$ typically contains all hyperplanes relevant for the computation of k consecutive Tukey depth contours. Technical details are provided in [28]. A Matlab implementation of the procedure is available from the authors. That implementation was used to generate all the illustrations in this paper.

6. Multiple-output quantile regression. Our approach to multivariate quantiles also allow to define *multiple-output regression quantiles* enjoying all nice properties of their classical single-output counterparts.

Consider the multiple-output regression problem in which the m-variate response $\mathbf{Y} := (Y_1, \ldots, Y_m)'$ is to be regressed on the vector of regressors $\mathbf{X} := (X_1, \ldots, X_p)'$, where $X_1 = 1$ a.s. and the other X_j's are random. In the sequel, we let $\mathbf{X} =: (1, \mathbf{W}')'$, so that $\{(\mathbf{w}', \mathbf{y}')' : \mathbf{w} \in \mathbb{R}^{p-1}, \mathbf{y} \in \mathbb{R}^m\} = \mathbb{R}^{p-1} \times \mathbb{R}^m$ is the natural space for considering fitted regression "objects". Multiple-output regression quantiles, in that context, can be obtained by applying Definition 2.1 to the k-dimensional random vector $\mathbf{Z} := (\mathbf{W}', \mathbf{Y}')'$, $k = p + m - 1$, *with the important restriction that the direction* $\mathbf{u}$ *should be taken in the response space only*, that is, $\mathbf{u} \in \mathcal{S}_{p-1}^{m-1} := \{\mathbf{0}_{p-1}\} \times \mathcal{S}^{m-1} \subset \mathcal{S}^{k-1}$. This directly yields the following definition.

DEFINITION 6.1. *For any $\boldsymbol{\tau} = \tau\mathbf{u}$, with $\tau \in (0,1)$ and $\mathbf{u} = (\mathbf{0}'_{p-1}, \mathbf{u}'_{\mathbf{y}})' \in \mathcal{S}^{m-1}_{p-1}$, the regression $\boldsymbol{\tau}$-quantile of $\mathbf{Y}$ with respect to $\mathbf{X} = (1, \mathbf{W}')'$ is defined as any element of the collection $\Pi_{\boldsymbol{\tau}}$ of hyperplanes $\pi_{\boldsymbol{\tau}} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p+m-1} : \mathbf{u}'_{\mathbf{y}}\mathbf{y} = \mathbf{b}'_{\boldsymbol{\tau}}\boldsymbol{\Gamma}'_{\mathbf{u}}(\mathbf{w}', \mathbf{y}')' + a_{\boldsymbol{\tau}}\}$ such that*

$$(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})' \in \underset{(a,\mathbf{b}')' \in \mathbb{R}^{p+m-1}}{\operatorname{argmin}} \Psi_{\boldsymbol{\tau}}(a, \mathbf{b}), \tag{6.1}$$

where, denoting by $\boldsymbol{\Gamma}_{\mathbf{u}}$ an arbitrary $(p+m-1) \times (p+m-2)$ matrix such that $(\mathbf{u} \vdots \boldsymbol{\Gamma}_{\mathbf{u}})$ is orthogonal, we let $\Psi_{\boldsymbol{\tau}}(a, \mathbf{b}) := \mathrm{E}[\rho_\tau(\mathbf{u}'_{\mathbf{y}}\mathbf{Y} - \mathbf{b}'\boldsymbol{\Gamma}'_{\mathbf{u}}(\mathbf{W}', \mathbf{Y}')' - a)]$.

Although—similarly as in Definition 2.1—the choice of $\boldsymbol{\Gamma}_{\mathbf{u}}$ has no impact on the directional regression quantile $\pi_{\boldsymbol{\tau}}$, it is here natural to take $\boldsymbol{\Gamma}_{\mathbf{u}}$ of the form

$$\boldsymbol{\Gamma}_{\mathbf{u}} = \begin{pmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{\mathbf{u}_{\mathbf{y}}} \end{pmatrix},$$

where $\mathbf{I}_{p-1}$ denotes the $(p-1)$-dimensional identity matrix and the $m \times (m-1)$ matrix $\boldsymbol{\Gamma}_{\mathbf{u}_{\mathbf{y}}}$ is such that $(\mathbf{u}_{\mathbf{y}} \vdots \boldsymbol{\Gamma}_{\mathbf{u}_{\mathbf{y}}})$ is orthogonal. The directional regression quantiles in Definition 6.1 then take the form

$$\pi_{\boldsymbol{\tau}} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p+m-1} : \mathbf{u}'_{\mathbf{y}}\mathbf{y} = \mathbf{b}'_{\boldsymbol{\tau}\mathbf{y}}\boldsymbol{\Gamma}'_{\mathbf{u}_{\mathbf{y}}}\mathbf{y} + \mathbf{b}'_{\boldsymbol{\tau}\mathbf{w}}\mathbf{w} + a_{\boldsymbol{\tau}}\},$$

with $\mathbf{b}_{\boldsymbol{\tau}} = (\mathbf{b}'_{\boldsymbol{\tau}\mathbf{w}}, \mathbf{b}'_{\boldsymbol{\tau}\mathbf{y}})'$. Clearly, an equivalent definition of multiple-output regression quantiles can be obtained by extending Definition 2.2 in the same fashion.

Now, as in the location case, each quantile hyperplane $\pi_{\boldsymbol{\tau}}$ characterizes a lower (open) and upper (closed) regression quantile halfspaces defined as

$$H^-_{\boldsymbol{\tau}} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p+m-1} : \mathbf{u}'_{\mathbf{y}}\mathbf{y} < \mathbf{b}'_{\boldsymbol{\tau}\mathbf{y}}\boldsymbol{\Gamma}'_{\mathbf{u}_{\mathbf{y}}}\mathbf{y} + \mathbf{b}'_{\boldsymbol{\tau}\mathbf{w}}\mathbf{w} + a_{\boldsymbol{\tau}}\} \tag{6.2}$$

and

$$H^+_{\boldsymbol{\tau}} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p+m-1} : \mathbf{u}'_{\mathbf{y}}\mathbf{y} \geq \mathbf{b}'_{\boldsymbol{\tau}\mathbf{y}}\boldsymbol{\Gamma}'_{\mathbf{u}_{\mathbf{y}}}\mathbf{y} + \mathbf{b}'_{\boldsymbol{\tau}\mathbf{w}}\mathbf{w} + a_{\boldsymbol{\tau}}\}, \tag{6.3}$$

respectively. Most importantly, for fixed $\tau (= \|\boldsymbol{\tau}\|) \in (0,1)$, (multiple-output) τ-*quantile regression regions* are obtained by taking the "upper envelope" of our regression $\boldsymbol{\tau}$-quantile hyperplanes. More precisely, for any $\tau \in (0,1)$, we define regression τ-quantile regions $R_{\mathrm{regr}}(\tau)$ as

$$R_{\mathrm{regr}}(\tau) := \bigcap_{\mathbf{u} \in \mathcal{S}^{m-1}_{p-1}} \bigcap \{H^+_{\tau\mathbf{u}}\} \tag{6.4}$$

(with corresponding regression contours $\partial R_{\mathrm{regr}}(\tau)$), where $H^{+}_{\tau\mathbf{u}}$ denotes the (closed) upper regression ($\tau\mathbf{u}$)-quantile halfspace in (6.3). Unlike the location case ($p = 1$), these regression regions may be non-nested—an m-dimensional form of the familiar *regression quantile crossing phenomenon*.

Finite-sample versions of all regression concepts above are obtained, similarly as in the location case (Section 2), as the natural sample analogs of the corresponding population concepts; see Figures 5 and 6 for an illustration. From a numerical point of view, Section 5.2, with obvious minor changes, still describes how to compute the resulting regression quantile regions $R^{(n)}_{\mathrm{regr}}(\tau)$, with m and $\mathbf{u}_y$ substituded for k and $\mathbf{u}$.

7. Final comments. This work defines a new concept of multivariate quantile based on L_1 optimization ideas and clarifies the quantile nature of halfspace depth contours, while providing an extremely efficient way to compute the latter. The same concept readily allows for an extension of quantile regression to the multiple-output context, thus paving the way to a multiple-output generalization of the many tools and techniques that have been based on the standard (single-output) Koenker and Bassett concept of quantile regression. This final section quickly discusses several open problems, of high practical relevance, that could now be considered.

First of all, Section 6 only very briefly indicates how our multivariate quantiles extend to the context of multiple-output regression; that extension clearly calls for a more detailed study, covering standard asymptotic issues (limiting distributions, Bahadur representations) as well as robustness aspects (breakdown points and influence functions) and nonlinear quantile regression.

The regression rank score perspectives (associated with linear programming duality) sketched in Section 5.1 also look extremely promising, possibly leading to the development of a full body of multivariate, depth-related, methods of rank-based inference.

Finally, as mentioned in the Introduction and in Section 3.1, various concepts introduced in this paper can be quite useful for inference. As an example, note that the symmetry (central, elliptical, or spherical) structure of P is reflected in the mappings

$$\mathbf{u} \mapsto \lambda_{\tau\mathbf{u}}\mathbf{u}/\lambda^{(\infty)}_{\tau} \quad \text{and} \quad \mathbf{u} \mapsto \|\mathbf{c}_{\tau\mathbf{u}}\|\mathbf{u}/c^{(\infty)}_{\tau},$$

with $\lambda^{(\infty)}_{\tau} := \sup_{\mathbf{u}\in\mathcal{S}^{k-1}} \lambda_{\tau\mathbf{u}}$ and $c^{(\infty)}_{\tau} := \sup_{\mathbf{u}\in\mathcal{S}^{k-1}} \|\mathbf{c}_{\tau\mathbf{u}}\|$, as illustrated (with the corresponding empirical quantities, of course) in Figure 7. A test of the hypothesis that the density of $\mathbf{Z}$ is, e.g., spherically symmetric thus could be based on (the empirical version $T^{(n)} := T(\mathrm{P}_n)$ of) a functional of the form

$$T(\mathrm{P}) := \int_0^1 \int_{\mathcal{S}^{k-1}} \delta\Big(\frac{\lambda_{\tau\mathbf{u}}}{\lambda_\tau^{(\infty)}}, 1\Big)\, d\sigma(\mathbf{u})\, W(\tau)\, d\tau,$$

where $\delta(\cdot,\cdot)$ denotes some distance (such as Cramér-von Mises), W some positive weight function over $(0,1)$, and σ the uniform measure over $\mathcal{S}^{k-1}$.

APPENDIX A

PROOF OF THEOREM 3.2. The quantity $\boldsymbol{\eta}_{i,\boldsymbol{\tau}}(a,\mathbf{b}) := \mathbf{J}'_{\mathbf{u}}\boldsymbol{\xi}_{i,\boldsymbol{\tau}}(a,\mathbf{b})$ is a subgradient for $(a,\mathbf{b}) \mapsto \rho_\tau(\mathbf{Z}_{i\mathbf{u}} - \mathbf{b}'\mathbf{Z}^\perp_{i\mathbf{u}} - a)$ since, for all $(a,\mathbf{b}')', (a_0,\mathbf{b}'_0)' \in \mathbb{R}^k$, we have that

$$\begin{aligned}\rho_\tau(\mathbf{Z}_{i\mathbf{u}} - \mathbf{b}'\mathbf{Z}^\perp_{i\mathbf{u}} - a) - \rho_\tau(\mathbf{Z}_{i\mathbf{u}} - \mathbf{b}'_0\mathbf{Z}^\perp_{i\mathbf{u}} - a_0) - (a-a_0, \mathbf{b}'-\mathbf{b}'_0)\,\boldsymbol{\eta}_{i,\boldsymbol{\tau}}(a_0,\mathbf{b}_0)\\ = (\mathbb{I}_{[\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'_0\boldsymbol{\Gamma}'_{i\mathbf{u}}\mathbf{Z}_i - a_0<0]} - \mathbb{I}_{[\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'\boldsymbol{\Gamma}'_{i\mathbf{u}}\mathbf{Z}_i - a<0]})(\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'\boldsymbol{\Gamma}'_{i\mathbf{u}}\mathbf{Z}_i - a) \geq 0,\end{aligned}$$

irrespective of the value of $\mathbf{Z}_i$. Hence, interchanging differentiation and expectation (which is justified in a standard way) shows that $(a,\mathbf{b}')' \mapsto \Psi_{\boldsymbol{\tau}}(a,\mathbf{b})$ (see Definition 2.1) satisfies $\mathrm{grad}\,\Psi_{\boldsymbol{\tau}}(a,\mathbf{b}) = \mathrm{grad}\,\mathrm{E}[\rho_\tau(\mathbf{Z}_{i\mathbf{u}} - \mathbf{b}'\mathbf{Z}^\perp_{i\mathbf{u}} - a)] = \mathrm{E}[\boldsymbol{\eta}_{i,\boldsymbol{\tau}}(a,\mathbf{b})]$; see (3.2a)-(3.2b). Therefore,

$$\begin{aligned}&\mathrm{grad}\,\Psi_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}} + \Delta_a, \mathbf{b}_{\boldsymbol{\tau}} + \boldsymbol{\Delta}_{\mathbf{b}}) - \mathrm{grad}\,\Psi_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) - \mathbf{H}_{\boldsymbol{\tau}}(\Delta_a, \boldsymbol{\Delta}'_{\mathbf{b}})'\\ &\quad= \int_{\mathbb{R}^{k-1}} \int_{a_{\boldsymbol{\tau}}+\mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x}}^{(a_{\boldsymbol{\tau}}+\Delta_a)+(\mathbf{b}_{\boldsymbol{\tau}}+\boldsymbol{\Delta}_{\mathbf{b}})'\mathbf{x}} (f(z\mathbf{u} + \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{x}) - f((a_{\boldsymbol{\tau}} + \mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x})\mathbf{u} + \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{x}))\,(1,\mathbf{x}')'\,dz\,d\mathbf{x},\end{aligned}$$

and Assumption (A'_n) yields that

$$\begin{aligned}&\|\mathrm{grad}\,\Psi_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}} + \Delta_a, \mathbf{b}_{\boldsymbol{\tau}} + \boldsymbol{\Delta}_{\mathbf{b}}) - \mathrm{grad}\,\Psi_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) - \mathbf{H}_{\boldsymbol{\tau}}(\Delta_a, \boldsymbol{\Delta}'_{\mathbf{b}})'\|\\ &\quad\leq C\int_{\mathbb{R}^{k-1}} \left| \int_{a_{\boldsymbol{\tau}}+\mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x}}^{(a_{\boldsymbol{\tau}}+\Delta_a)+(\mathbf{b}_{\boldsymbol{\tau}}+\boldsymbol{\Delta}_{\mathbf{b}})'\mathbf{x}} \frac{|z - (a_{\boldsymbol{\tau}} + \mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x})|^s\,\|(1,\mathbf{x}')'\|}{(1 + \|\frac{1}{2}(z + a_{\boldsymbol{\tau}} + \mathbf{b}'_{\boldsymbol{\tau}}\mathbf{x})\mathbf{u} + \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{x}\|^2)^{(3+r+s)/2}}\,dz \right| d\mathbf{x}\\ &\quad\leq C\int_{\mathbb{R}^{k-1}} |\Delta_a + \boldsymbol{\Delta}'_{\mathbf{b}}\mathbf{x}|\,\frac{|\Delta_a + \boldsymbol{\Delta}'_{\mathbf{b}}\mathbf{x}|^s\,\|(1,\mathbf{x}')'\|}{\|(1,\mathbf{x}')'\|^{3+r+s}}\,d\mathbf{x}\\ &\quad\leq C\,\|(\Delta_a, \boldsymbol{\Delta}'_{\mathbf{b}})'\|^{1+s} \int_{\mathbb{R}^{k-1}} \|(1,\mathbf{x}')'\|^{-(r+1)}\,d\mathbf{x} = o(\|(\Delta_a, \boldsymbol{\Delta}'_{\mathbf{b}})'\|),\end{aligned}$$

as $\|(\Delta_a, \boldsymbol{\Delta}'_{\mathbf{b}})'\| \to 0$. This shows that $(a,\mathbf{b}')' \mapsto \Psi_{\boldsymbol{\tau}}(a,\mathbf{b})$ is twice differentiable at $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ with Hessian matrix $\mathbf{H}_{\boldsymbol{\tau}}$. Since, moreover, Assumption (A'_n) clearly ensures that $\mathrm{E}[\|\boldsymbol{\eta}_{i,\boldsymbol{\tau}}(a,\mathbf{b})\|^2] < \infty$ for all $(a,\mathbf{b}')' \in \mathbb{R}^k$, Theorem 4 of [26] applies, which establishes (3.16). Of course, (3.17) results from (3.16) by the multivariate CLT.

Recall that, under Assumption (A), the unique solution of (2.1) can be written as $(a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})' := (a_{\boldsymbol{\tau}}, (-\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{c}_{\boldsymbol{\tau}})')'$, where $(a_{\boldsymbol{\tau}}, \mathbf{c}'_{\boldsymbol{\tau}})'$ denotes the unique solution of (2.5). Similarly, any solution $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}})'$ of (2.6) is related to some solution $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{c}^{(n)\prime}_{\boldsymbol{\tau}})'$ of (2.7) via the relation $(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau}})' = (a^{(n)}_{\boldsymbol{\tau}}, (-\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{c}^{(n)}_{\boldsymbol{\tau}})')'$. This allows for rewriting (3.16) as

$$\sqrt{n}\,\mathbf{P}_k\mathbf{J}'_{\mathbf{u}}\begin{pmatrix} a^{(n)}_{\boldsymbol{\tau}} - a_{\boldsymbol{\tau}} \\ \mathbf{c}^{(n)}_{\boldsymbol{\tau}} - \mathbf{c}_{\boldsymbol{\tau}} \end{pmatrix} = -\frac{1}{\sqrt{n}}\,\mathbf{H}^{-1}_{\boldsymbol{\tau}}\mathbf{J}'_{\mathbf{u}}\sum_{i=1}^{n}\boldsymbol{\xi}^{c}_{i,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}) + o_{\mathrm{P}}(1), \tag{A.1}$$

as $n \to \infty$. By first premultiplying both sides of (A.1) with $\mathbf{P}_k\mathbf{J}_{\mathbf{u}}$, then using $\boldsymbol{\Gamma}_{\mathbf{u}}\boldsymbol{\Gamma}'_{\mathbf{u}} = \mathbf{I}_k - \mathbf{u}\mathbf{u}'$ (which follows from the orthogonality of $(\mathbf{u} \vdots \boldsymbol{\Gamma}_{\mathbf{u}})$) and $\mathbf{u}'\mathbf{c}^{(n)}_{\boldsymbol{\tau}} = 1 = \mathbf{u}'\mathbf{c}_{\boldsymbol{\tau}}$, we obtain

$$\sqrt{n}\begin{pmatrix} a^{(n)}_{\boldsymbol{\tau}} - a_{\boldsymbol{\tau}} \\ \mathbf{c}^{(n)}_{\boldsymbol{\tau}} - \mathbf{c}_{\boldsymbol{\tau}} \end{pmatrix} = -\frac{1}{\sqrt{n}}\,\mathbf{P}_k\mathbf{J}_{\mathbf{u}}\mathbf{H}^{-1}_{\boldsymbol{\tau}}\mathbf{J}'_{\mathbf{u}}\sum_{i=1}^{n}\boldsymbol{\xi}^{c}_{i,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{c}_{\boldsymbol{\tau}}) + o_{\mathrm{P}}(1),$$

as $n \to \infty$. Lemma A.1 below therefore establishes (3.18). Again, the multivariate CLT then trivially yields (3.19).

Finally, applying Theorem 6 in [26] (more precisely, applying the version (a) of Statement (3.8) in that theorem) with $L = \mathbf{I}_k$ and $c = (a_{\boldsymbol{\tau}}, \mathbf{b}'_{\boldsymbol{\tau}})'$ yields

$$n\Psi^{(n)}_{\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) - n\Psi^{(n)}_{\boldsymbol{\tau}}(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)}_{\boldsymbol{\tau}}) - \frac{1}{2n}\sum_{i,j=1}^{n}\boldsymbol{\xi}'_{i,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}})\mathbf{J}_{\mathbf{u}}\mathbf{H}^{-1}_{\boldsymbol{\tau}}\mathbf{J}'_{\mathbf{u}}\boldsymbol{\xi}_{j,\boldsymbol{\tau}}(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) = o_{\mathrm{P}}(1), \tag{A.2}$$

as $n \to \infty$. Note that the third term is clearly $O_{\mathrm{P}}(1)$ as $n \to \infty$. The result then follows by dividing both sides of (A.2) by $\sqrt{n}$, and by using the identities $\lambda^{(n)}_{\boldsymbol{\tau}} = \Psi^{(n)}_{\boldsymbol{\tau}}(a^{(n)}_{\boldsymbol{\tau}}, \mathbf{b}^{(n)}_{\boldsymbol{\tau}})$ (see the end of Section 5.1) and $\mathbf{u}'\mathbf{z} - \mathbf{b}'_{\boldsymbol{\tau}}\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{z} - a_{\boldsymbol{\tau}} = \mathbf{c}'_{\boldsymbol{\tau}}\mathbf{z} - a_{\boldsymbol{\tau}}$ for all $\mathbf{z} \in \mathbb{R}^k$. Since (3.7) entails $\lambda_{\boldsymbol{\tau}} = \mathrm{E}[\rho_\tau(\mathbf{c}'_{\boldsymbol{\tau}}\mathbf{Z}_i - a_{\boldsymbol{\tau}})]$, the CLT yields (3.21). ☐

In order to complete the proof of Theorem 3.2, it is sufficient to establish the following lemma.

LEMMA A.1. *The matrix* $\mathbf{G}_{\boldsymbol{\tau}} := \mathbf{J}_{\mathbf{u}}(\mathbf{J}'_{\mathbf{u}}\mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}})^{-1}\mathbf{J}'_{\mathbf{u}}$ *is the Moore-Penrose pseudoinverse of* $\mathbf{H}^{c}_{\boldsymbol{\tau}}$, *that is,* $\mathbf{G}_{\boldsymbol{\tau}}$ *is such that (i)* $\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}} = \mathbf{G}_{\boldsymbol{\tau}}$, *(ii)* $\mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^{c}_{\boldsymbol{\tau}} = \mathbf{H}^{c}_{\boldsymbol{\tau}}$, *(iii)* $(\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^{c}_{\boldsymbol{\tau}})' = \mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^{c}_{\boldsymbol{\tau}}$, *and (iv)* $(\mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}})' = \mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}}$.

PROOF OF LEMMA A.1. (i) This directly follows from trivial computations. (ii) Let $\mathbf{K}_{\mathbf{u}}$ be the invertible matrix $(\mathbf{J}_{\mathbf{u}} \vdots \dot{\mathbf{u}})$, where $\dot{\mathbf{u}} := (0, \mathbf{u}')'$. Clearly, $(\mathbf{H}^{c}_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^{c}_{\boldsymbol{\tau}} - \mathbf{H}^{c}_{\boldsymbol{\tau}})\mathbf{J}_{\mathbf{u}} = \mathbf{0}$, and the definition of $\mathbf{H}^{c}_{\boldsymbol{\tau}}$ implies that $\dot{\mathbf{u}}$ belongs to the null space of $\mathbf{H}^{c}_{\boldsymbol{\tau}}$.

Hence, $(\mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^c_{\boldsymbol{\tau}} - \mathbf{H}^c_{\boldsymbol{\tau}})\mathbf{K}_{\mathbf{u}} = \mathbf{0}$, which establishes the result. (iii)-(iv) Since $\mathbf{J}'_{\mathbf{u}}\mathbf{J}_{\mathbf{u}} = \mathbf{I}_k$, $(\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^c_{\boldsymbol{\tau}} - \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}})\mathbf{J}_{\mathbf{u}} = \mathbf{J}_{\mathbf{u}} - \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}}(\mathbf{J}'_{\mathbf{u}}\mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}})^{-1} = \mathbf{0}$; the last equality follows, as in the proof of Part (ii), by showing that $(\mathbf{J}_{\mathbf{u}} - \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}}(\mathbf{J}'_{\mathbf{u}}\mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{J}_{\mathbf{u}})^{-1})'\mathbf{K}_{\mathbf{u}} = \mathbf{0}$. Now, as we also have that $(\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^c_{\boldsymbol{\tau}} - \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}})\dot{\mathbf{u}} = \mathbf{0}$, we conclude that $(\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^c_{\boldsymbol{\tau}} - \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}})\mathbf{K}_{\mathbf{u}} = \mathbf{0}$, hence that $\mathbf{G}_{\boldsymbol{\tau}}\mathbf{H}^c_{\boldsymbol{\tau}} = \mathbf{H}^c_{\boldsymbol{\tau}}\mathbf{G}_{\boldsymbol{\tau}}$. This establishes (iii)-(iv) since both $\mathbf{H}^c_{\boldsymbol{\tau}}$ and $\mathbf{G}_{\boldsymbol{\tau}}$ are symmetric. □

PROOF OF THEOREM 4.1. Under Assumption (A), it directly follows from (4.4) that, for any $\tau \in (0,1)$ (note that Theorem 4.1 trivially holds for $\tau = 0$), $D(\tau) = \bigcap \{H : H \text{ is a closed halfspace with } P[\mathbf{Z} \in H] \geq 1-\tau\}$. Consequently, by noting that any $H^+_{\mathrm{KM};\tau\mathbf{u}}$, $\mathbf{u} \in \mathcal{S}^{k-1}$ (see (2.4)) satisfies $\mathrm{P}[\mathbf{Z} \in H^+_{\mathrm{KM};\tau\mathbf{u}}] = 1-\tau$ under Assumption (A), it follows from (4.5) that

$$\begin{aligned} D(\tau) &\subset \bigcap \{H : H \text{ is a closed halfspace with } P[\mathbf{Z} \in H] = 1-\tau\} \\ &\subset \bigcap_{\mathbf{u} \in \mathcal{S}^{k-1}} \{H^+_{\mathrm{KM};\tau\mathbf{u}}\} = D(\tau), \end{aligned}$$

which entails that, still under Assumption (A),

$$D(\tau) = \bigcap \{H : H \text{ is a closed halfspace with } \mathrm{P}[\mathbf{Z} \in H] = 1-\tau\}. \tag{A.3}$$

Now, since (3.2a) (equivalently, (3.5a)) implies that any closed quantile halfspace $H^+_{\tau\mathbf{u}}$, $\mathbf{u} \in \mathcal{S}^{k-1}$, satisfies $\mathrm{P}[\mathbf{Z} \in H^+_{\tau\mathbf{u}}] = 1-\tau$, (A.3) yields that $D(\tau) \subset R(\tau)$. To show $D(\tau) \supset R(\tau)$, consider an arbitrary closed halfspace H with $\mathrm{P}[\mathbf{Z} \in H] = 1-\tau$. Then $H = H^+_{\tau\mathbf{u}}$, with

$$\mathbf{u} := \frac{\frac{1}{1-\tau}\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in H]}] - \frac{1}{\tau}\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in \mathbb{R}^k\setminus H]}]}{\|\frac{1}{1-\tau}\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in H]}] - \frac{1}{\tau}\mathrm{E}[\mathbf{Z}\,\mathbb{I}_{[\mathbf{Z}\in \mathbb{R}^k\setminus H]}]\|},$$

so that $R(\tau) \subset D(\tau)$; see (3.6) and (A.3) again. □

PROOF OF THEOREM 4.2. We start with some remarks on sample halfspace depth regions. By (4.4), $D^{(n)}(\frac{\ell}{n})$, for any $\ell \in \{1, 2, \ldots, n-k\}$, coincides with the intersection of all closed halfspaces containing at least $n-\ell+1$ observations. Actually, one can restrict to closed halfspaces containing exactly $n-\ell+1$ observations (see [10], page 1805). Also, it can be shown (see [12]) that $D^{(n)}(\frac{\ell}{n})$—provided that its interior is not the empty set—is bounded by hyperplanes containing at least k points that span a $(k-1)$-dimensional subspace of $\mathbb{R}^k$.

Now, fix $\ell \in \{1, 2, \ldots, n-k\}$ such that $D^{(n)}(\frac{\ell}{n})$ has indeed a non-empty interior. Consider an arbitrary closed halfspace H containing exactly $n - \ell + 1$ data points, among which exactly k ($\mathbf{Z}_i$, $i \in h = \{i_1, \ldots, i_k\}$, say) sit in ∂H—and actually span ∂H, since the data points are assumed to be in general position. It follows from the results stated in the previous paragraph that $D^{(n)}(\frac{\ell}{n})$, under the assumptions of Theorem 4.2, coincides with the intersection of all such halfspaces.

Letting $s_\tau(n, k, \ell) := (n - k - \ell + 1)\tau + (\ell - 1)(\tau - 1)$, define then

$$\mathbf{u} = \frac{\mathbf{T}_D - s_\tau(n,k,\ell)\mathbf{T}_{on}}{\|\mathbf{T}_D - s_\tau(n,k,\ell)\mathbf{T}_{on}\|}, \tag{A.4}$$

where

$$\mathbf{T}_D := \tau \sum_{\mathbf{Z}_i \in H \setminus \partial H} \mathbf{Z}_i + (\tau - 1) \sum_{\mathbf{Z}_i \notin H} \mathbf{Z}_i \quad \text{and} \quad \mathbf{T}_{on} := \frac{1}{k} \sum_{\mathbf{Z}_i \in \partial H} \mathbf{Z}_i.$$

Taking $\mathbf{\Gamma_u}$ as in Definition 2.1, one of course has $\mathbf{\Gamma}'_\mathbf{u}\mathbf{T}_D = s_\tau(n,k,\ell)\mathbf{\Gamma}'_\mathbf{u}\mathbf{T}_{on}$. Hence, writing $(a_h, \mathbf{b}'_h)'$ for the unique solution of

$$\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'\mathbf{\Gamma}'_\mathbf{u}\mathbf{Z}_i - a = 0, \quad i \in h,$$

we obtain

$$\begin{aligned}
&\sum_{i \in \{1,\ldots,n\} \setminus h} (\tau - \mathbb{I}_{[\mathbf{u}'\mathbf{Z}_i - \mathbf{b}'_h\mathbf{\Gamma}'_\mathbf{u}\mathbf{Z}_i - a_h < 0]}) \begin{pmatrix} 1 \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{Z}_i \end{pmatrix} \\
&\qquad = \tau \sum_{\mathbf{Z}_i \in H \setminus \partial H} \begin{pmatrix} 1 \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{Z}_i \end{pmatrix} + (\tau - 1) \sum_{\mathbf{Z}_i \notin H} \begin{pmatrix} 1 \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{Z}_i \end{pmatrix} \\
&\qquad = \begin{pmatrix} s_\tau(n,k,\ell) \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{T}_D \end{pmatrix} = s_\tau(n,k,\ell) \begin{pmatrix} 1 \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{T}_{on} \end{pmatrix}.
\end{aligned}$$

Since (see (3.10))

$$\frac{1}{k} \mathbb{X}'_\mathbf{u}(h)\mathbf{1}_k = \begin{pmatrix} 1 \\ \mathbf{\Gamma}'_\mathbf{u}\mathbf{T}_{on} \end{pmatrix},$$

this implies that, with the same notation as in the end of Section 3.1, we have

$$\boldsymbol{\xi}_{\tau\mathbf{u}}(h) = \frac{s_\tau(n,k,\ell)}{k} \mathbf{1}_k,$$

hence that the subgradient conditions (3.11) are satisfied for any $\tau \in [\frac{\ell-1}{n}, \frac{\ell+k-1}{n}]$. It follows that, for any such τ, H coincides with the upper quantile halfspace $H_{\tau\mathbf{u}}^{(n)+}$ associated with some $\pi_{\tau\mathbf{u}}^{(n)} \in \Pi_{\tau\mathbf{u}}^{(n)}$, where $\mathbf{u}$ is as in (A.4), so that

$$R^{(n)}(\tau) := \bigcap_{\mathbf{u}\in\mathcal{S}^{k-1}} \bigcap \{H_{\tau\mathbf{u}}^{(n)+}\} \subset D^{(n)}(\tfrac{\ell}{n}), \tag{A.5}$$

for any *positive* $\tau \in [\frac{\ell-1}{n}, \frac{\ell}{n})$; one should indeed avoid the value $\tau = 0$ for which $R^{(n)}(\tau)$ is not defined as the upper envelope of quantile halfspaces.

Now, fix $\tau \in (0, \frac{\ell}{n})$. Then, according to (3.9), all upper sample quantile halfspaces $H_{\tau\mathbf{u}}^{(n)+}$ generating $R^{(n)}(\tau)$ contain $P + Z \geq \lceil n(1-\tau)\rceil = n - \lfloor n\tau \rfloor \geq n - \ell + 1$ observations. Hence, $D^{(n)}(\frac{\ell}{n}) \subset R^{(n)}(\tau)$ for any such τ. This, jointly with (A.5), establishes the result. □

Most interestingly, the proof of Theorem 4.2 actually shows that, for any $\tau \in (0,1)$, the set $\{\pi_{\tau\mathbf{u}}^{(n)} : \mathbf{u} \in \mathcal{S}^{k-1}, \pi_{\tau\mathbf{u}}^{(n)}$ contains k data points$\}$ coincides with the collection of all hyperplanes passing through k observations and cutting off at most $\lfloor n\tau \rfloor$ and at least $\lceil n\tau \rceil - k$ data points. Consequently, as stated at the end of Section 5.2, the set of τ-quantile hyperplanes in all directions provides enough material to compute not only one, but $\min(k + \eta_{n\tau}, \lfloor n\tau \rfloor + 1)$ Tukey depth contours at a time, where η_x is one if x is an integer and zero otherwise.

Acknowledgement. This research originates in the unpublished Master thesis of Benoît Laine ([23]) defended at the Université libre de Bruxelles (advisor Marc Hallin), devoted to a pioneering exploration of the subject. We are grateful to Ivan Mizera for encouraging Benoît Laine, inviting him to Edmonton, and sharing with him his expertise in depth problems. We also greatly benefited from many inspiring discussions with Ivan Mizera, Roger Koenker, Steve Portnoy, and Bob Serfling, to whom we express our warmest thanks.

REFERENCES

[1] Abdous, B., and Theodorescu, R. (1992). Note on the spatial quantile of a random vector. *Statist. Probab. Lett.* **13** 333–336.

[2] Bassett, G. W., Koenker, R., and Kordas, G. (2004). Pessimistic portfolio allocation and Choquet expected utility. *J. Fin. Econometrics* **2** 477–492.

[3] Breckling, J., and Chambers, R. (1988). M-quantiles. *Biometrika* **75** 761–771.

[4] Breckling, J., Kokic, P., and Lübke, O. (2001). A note on multivariate M-quantiles. *Statist. Probab. Lett.* **55** 39–44.

[5] Cascos, I., and López-Díaz, M. (2005). Integral trimmed regions. *J. Multivariate Anal.* **96** 404–424.

[6] Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Ann. Inst. Statist. Math.* **53** 380–403.

[7] Chakraborty, B. (2003). On multivariate quantile regression. *J. Statist. Plann. Inference* **110** 109–132.

[8] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.

[9] Chen, W. W., and Deo, R. S. (2004). Power transformations to induce normality and their applications. *J. R. Stat. Soc. Ser. B* **66** 117–130.

[10] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

[11] Dudley, R. M., and Koltchinskii, V. I. (1992). The spatial quantiles. Unpublished manuscript.

[12] Fukuda, K., and Rosta, V. (2005). Data depth and maximum feasible subsystems. In Avis, D., Hertz, A., and Marcotte, O., editors, *Data Depth and Combinatorial Optimization*, 37–67. Springer.

[13] Gutenbrunner, C, and Jurečková. J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305–330.

[14] Haberman, S. J. (1989). Concavity and estimation. *Ann. Statist.* **17** 1631–1661.

[15] He, X., and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504.

[16] Hettmansperger, T. P., Nyblom, J., and Oja, H. (1992). On multivariate notions of sign and rank. In *L_1-Statistical Analysis and Related Methods*, ed. Y. Dodge, Amsterdam: North Holland, 267–278.

[17] Hodges, J. L., Jr. (1955). A bivariate sign test. *Ann. Math. Statist.* **26** 523–527.

[18] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York, 1st edition.

[19] Koenker, R. (2007). *Quantile regression in R: A vignette*. Freely downloadable from CRAN: http://cran.r-project.org.

[20] Koenker, R., and Basset, G. J. (1978). Regression quantiles. *Econometrica* **46** 33–50.

[21] Koltchinskii, V. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477.

[22] Kong, L., and Mizera, I. (2008). Quantile tomography: Using quantiles with multivariate data. Submitted.

[23] Laine, B. (2001). *Depth contours as multivariate quantiles: a directional approach*. Master's thesis (advisor: M. Hallin), Université Libre de Bruxelles.

[24] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–840.

[25] Mizera, I. (2002). On depth and deep points : a calculus. *Ann. Statist.* **30** 1681–1736.

[26] Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.* **20** 1514–1533.

[27] Paindaveine, D., and Šiman, M. (2008a). Projection pursuit quantile regression. Manuscript in preparation.

[28] Paindaveine, D., and Šiman, M. (2008b). Computing multidimensional regression quantile

contours. Manuscript in preparation.

[29] Rousseeuw, P. J., and Ruts, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244.

[30] Rousseeuw, P. J., and Struyf, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference* **122** 161–173.

[31] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica* **56** 214–232.

[32] Serfling, R. (2008). A Mahalanobis multivariate quantile function. Submitted.

[33] Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians (Vancouver, B. C., 1974), Vol. 2, Quebec*: Canad. Math. Congress 523–531.

[34] Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.* **103** 397–409.

[35] Wei, Y., Pere, A., Koenker, R., and He, X. (2005). Quantile regression methods for reference growth charts. *Statistics in Medicine* **25** 1369–1382.

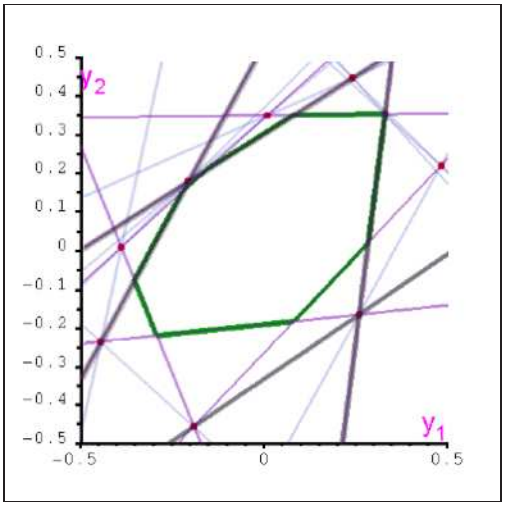
[36] Zuo, Y., and Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.* **28** 461–482.

[37] Zuo, Y., and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28** 483–499.

E.C.A.R.E.S., Institut de Recherche en Statistique,
and Département de Mathématique
Université Libre de Bruxelles
Campus de la Plaine CP 210
B-1050 Bruxelles
Belgium
E-mail: mhallin@ulb.ac.be

E.C.A.R.E.S., Institut de Recherche en Statistique,
and Département de Mathématique
Université Libre de Bruxelles
Avenue F.D. Roosevelt, 50, CP114
B-1050 Bruxelles
Belgium
E-mail: dpaindav@ulb.ac.be
url: http://homepages.ulb.ac.be/~dpaindav

E.C.A.R.E.S. and Institut de Recherche en Statistique
Université Libre de Bruxelles
Avenue F.D. Roosevelt, 50, CP114
B-1050 Bruxelles
Belgium



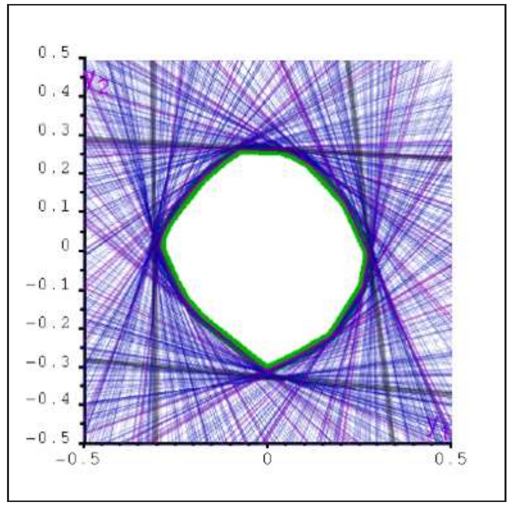


Figure 1: The left plot contains $n = 9$ (red) points drawn from $U([-.5, .5]^2)$ and provides all τ-quantile hyperplanes for $\tau = .2$ in magenta (in black for the four semiaxial directions); these hyperplanes define a central region (green contour), which, in Section 4, is shown to coincide with a Tukey depth region. The right plot provides the same information for $n = 499$ (invisible) points drawn from the same population distribution.

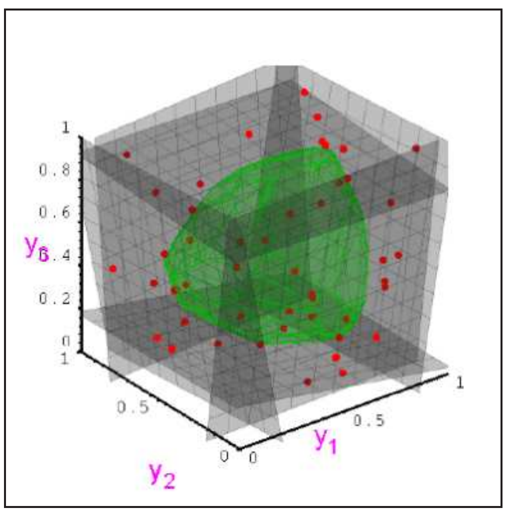


Figure 2: This plot provides the six semiaxial τ-quantile hyperplanes (in black) for $\tau = .1$, computed from $n = 49$ (red) points drawn from $U([0, 1]^3)$, along with the corresponding central region (in green).

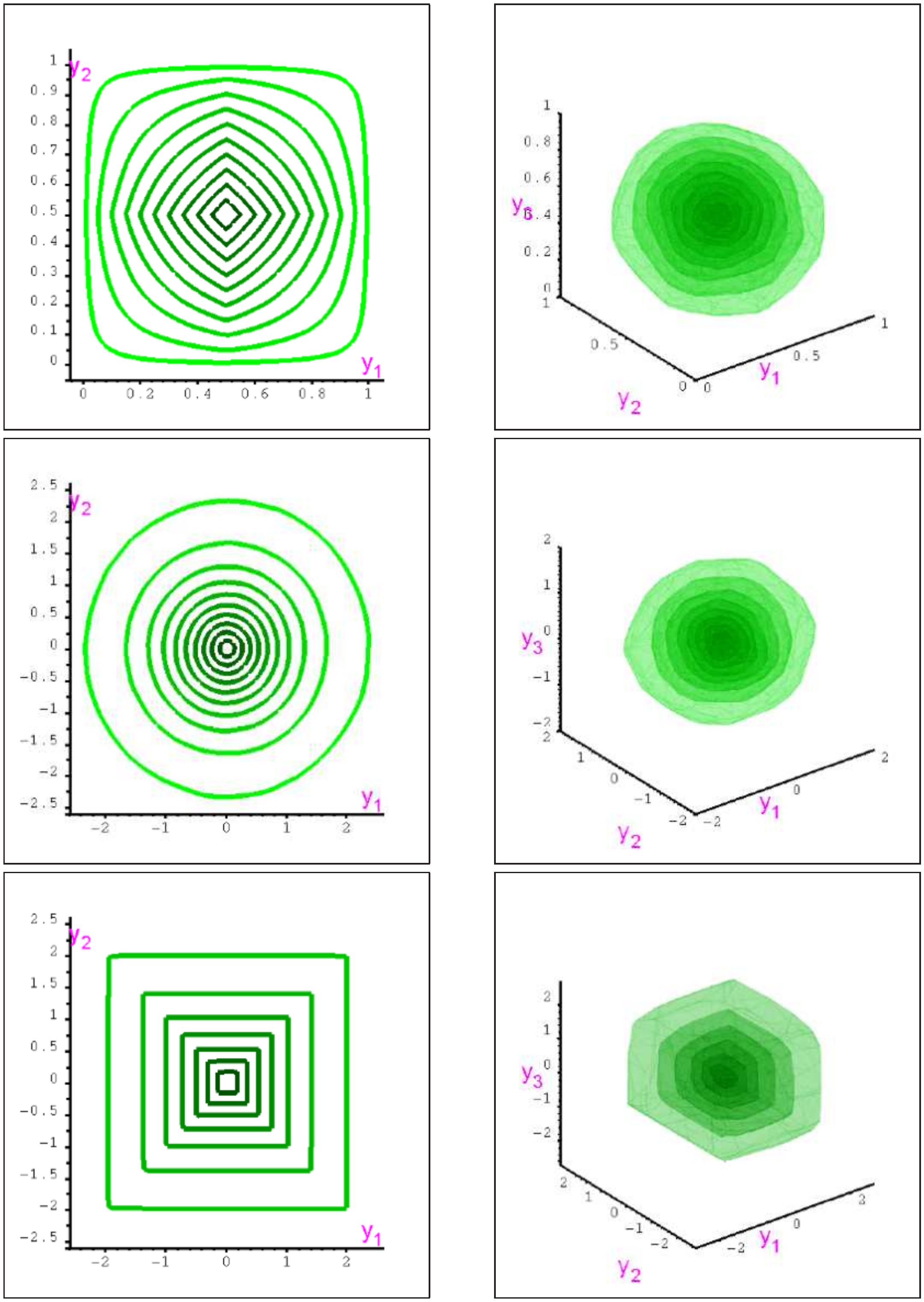


Figure 3: Tukey contours $D^{(n)}(\tau)$ (in green) obtained from $U([0,1]^k)$ (top), $N(0,1)^k$ (center), and t_1^k (bottom) for $k = 2$ (left), with $n = 49,999$ and $\tau \in \{.01, .05, .10, .15, .20, \ldots, .45\}$, and for $k = 3$ (right), with $n = 399$ and $\tau \in \{.05, .10, .15, .20, \ldots, .40\}$. Only the contours in the plotting range are displayed.

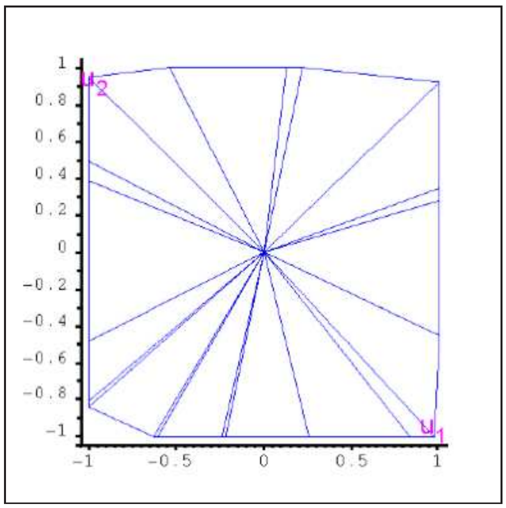


Figure 4: The twenty cones $\mathcal{C}_i(\tau)$ (in blue) obtained for $\tau = .1$ via parametric programming from $n = 19$ points drawn from $U([-.5, .5]^2)$, indicating that twenty distinct τ-quantile hyperplanes have been found in that case.

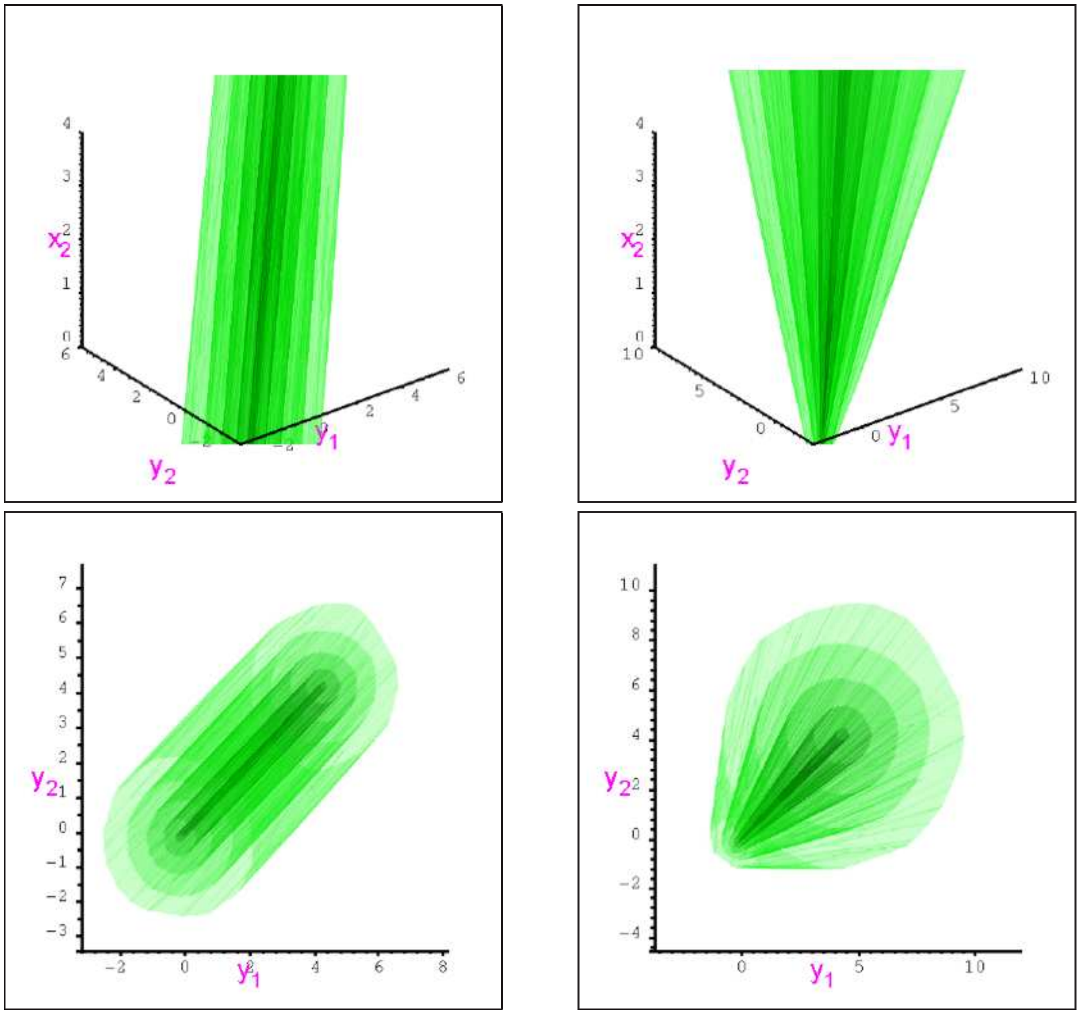


Figure 5: Two different views on the regression τ-quantile contours (in green) from 9,999 data points for $\tau \in \{.01, .05, .15, .30, .45\}$ in a homoscedastic ($(Y_1, Y_2)' = 4(X_2, X_2)' + \varepsilon$; left) and a heteroscedastic ($(Y_1, Y_2)' = 4(X_2, X_2)' + \sqrt{X_2}\varepsilon$; right) bivariate-output regression setting, respectively, where $X_2 \sim U([0, 1])$ and $\varepsilon \sim N(0, 1)^2$.

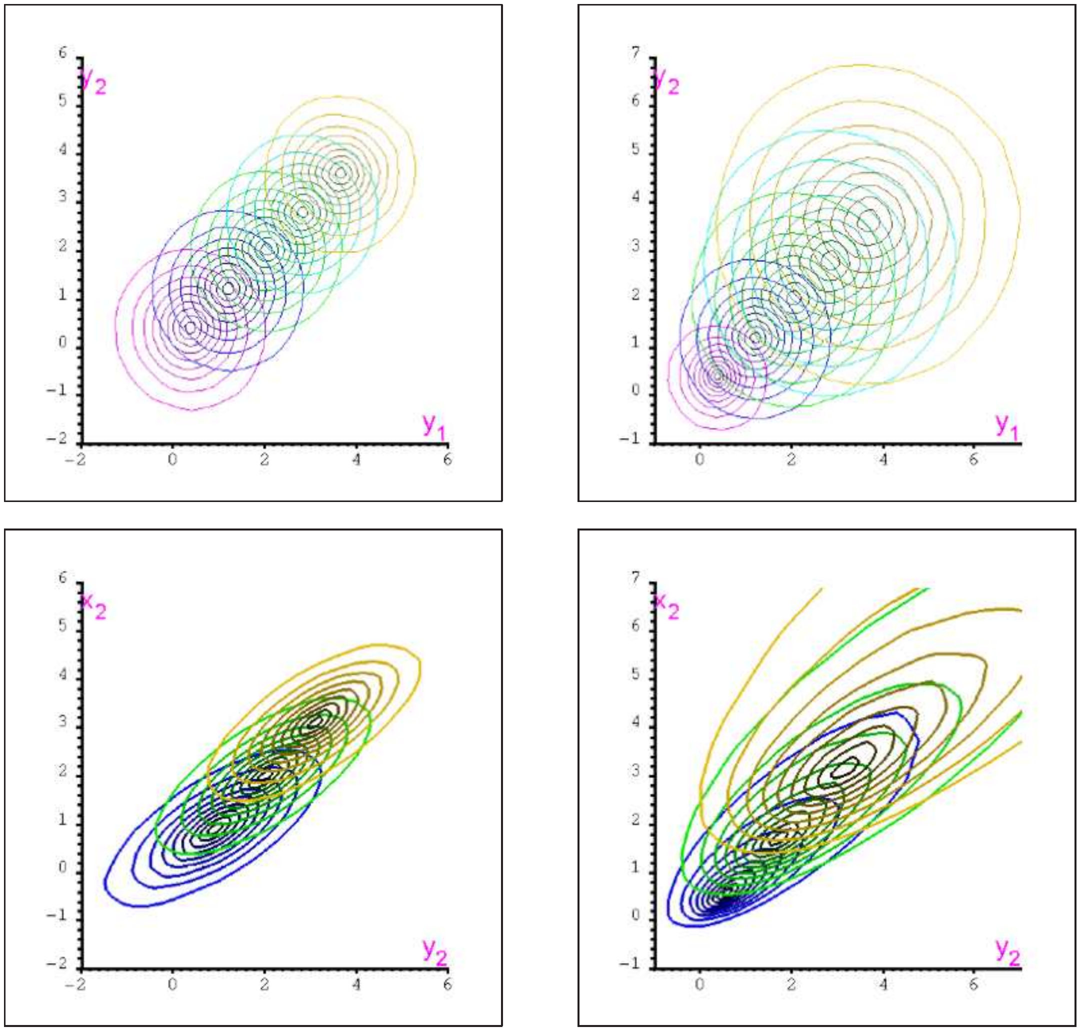


Figure 6: Various cuts of the regression τ-quantile "hypertube" regions from the same two models (left and right, respectively) as in Figure 5 with $n = 9,999$ observations. The top plots provide regression τ-quantile cuts, $\tau \in \{.05, .10, .15, .20, \ldots, .45\}$, through 10% (magenta), 30% (blue), 50% (green), 70% (cyan) and 90% (yellow) empirical quantiles of X_2; the bottom ones show regression τ-quantile cuts for the same τ values, and through 25% (blue), 50% (green) and 75% (yellow) empirical quantiles of Y_1. Their centers provide information about trend and their shapes and sizes shed light on variability.

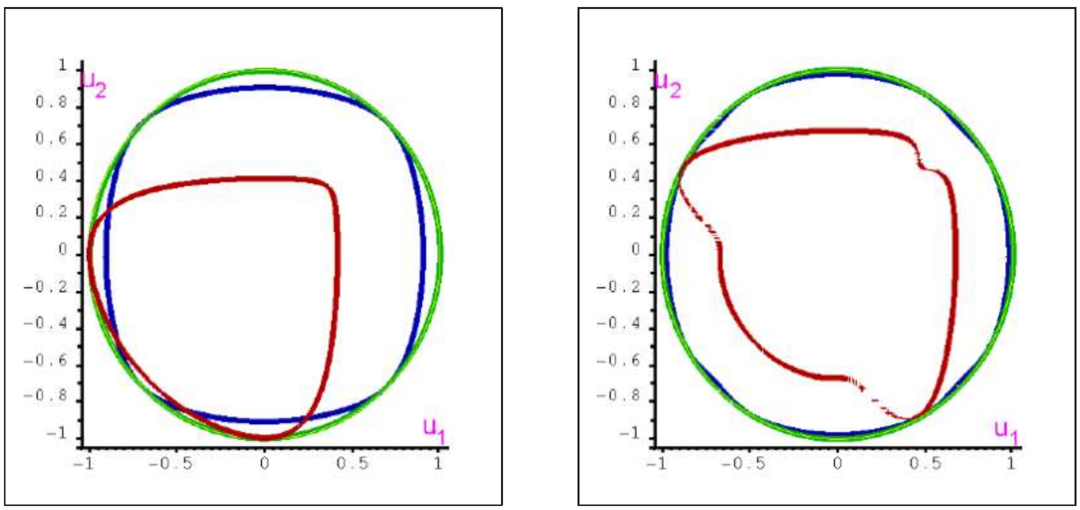


Figure 7: Plots (polar coordinates) of the mappings $\mathbf{u} \mapsto \lambda^{(n)}_{\tau\mathbf{u}}\mathbf{u}/(\sup_{\mathbf{v}\in\mathcal{S}^1}\lambda^{(n)}_{\tau\mathbf{v}})$ (left) and $\mathbf{u} \mapsto \|\mathbf{c}^{(n)}_{\tau\mathbf{u}}\|\mathbf{u}/(\sup_{\mathbf{v}\in\mathcal{S}^1}\|\mathbf{c}^{(n)}_{\tau\mathbf{v}}\|)$ (right), for $\tau = 0.1$ and $n = 49,999$ points drawn from $N(0,1)^2$ (green), $U([-.5,.5]^2)$ (blue), and $(Exp(1)-1)^2$ (red) populations, respectively; see Section 7. The resulting shapes clearly reflect the axes of symmetry of the underlying distributions.