Received: 23 April 2020 | Revised: 15 February 2022 | Accepted: 2 May 2022

DOI: 10.1111/sjos.12603

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Multivariate quantiles with both overall and directional probability interpretation

Daniel Hlubinka[1] | **Lukáš Kotík**[2] | **Miroslav Šiman**[3]

[1]Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

[2]National Radiation Protection Institute, Czech Republic

[3]The Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, Czech Republic

Correspondence
Daniel Hlubinka, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, CZ-186 75 Prague, Czech Republic.
Email: hlubinka@karlin.mff.cuni.cz

Funding information
Grantová Agentura České Republiky, Grant/Award Number: GA17-07384S; Grantová Agentura, Univerzita Karlova, Grant/Award Number: PRIMUS/17/SCI/03

Abstract

The article introduces multivariate quantiles (or reference regions) that have both overall and directional probability interpretation and need not be necessarily convex. They are defined by means of univariate conditional quantiles along the rays starting at a suitable central point. Their basic properties are investigated, their sample estimators and regression extensions are proposed, and their use is illustrated with both simulated and real data.

KEYWORDS

directional quantile, kernel estimation, multivariate quantile, nonconvex quantile, quantile regression, reference region, tolerance region

1 | INTRODUCTION

Univariate quantiles have already become important pillars of modern statistical theory. Unfortunately, none of their numerous multivariate extensions has become similarly useful, widespread, and universally acceptable due to the lack of natural ordering in vector spaces. The most fruitful multivariate quantile generalizations are usually based on statistical data depth, norm minimization or M-estimation, functional inversions, gradients, or generalized quantile processes and their properties; see Serfling (2002) for a partial survey. The various directional multivariate quantile concepts are especially appealing because they typically generalize to the multidimensional case not only the (inter)quantile intervals as quantile regions, but also the univariate quantiles themselves (usually as multidimensional points, vectors, or hyperplanes); see, for example, Chaudhuri (1996), Koltchinskii (1997), Chakraborty (2001), Cheng and De Gooijer (2007), Wei (2008), Hallin et al. (2010a,b), Paindaveine and Šiman (2011), Kong and Mizera (2012) and Hallin et al. (2015). Although these directional quantiles are often equipped with some direct probability interpretation related to their quantile level, such an interpretation of the resulting quantile regions is usually unclear and distribution-dependent. The only exceptions in this regard can probably be found in the two works closely related to this one, namely in Wei (2008) and Petersen (2003). They therefore deserve further attention.

Assume a standardized random vector $\boldsymbol{X} \in \mathbb{R}^p$ and a quantile level $\tau \in (0, 1)$. Wei (2008) considers all unit directions $\boldsymbol{u} \in \mathbb{R}^p$, $\|\boldsymbol{u}\| = 1$, and corresponding lines $a(\boldsymbol{u}) = \{t\boldsymbol{u},\ t \in \mathbb{R}\}$ containing the origin. Each of them determines the closed directional τ-quantile interval, say $[l_\tau(\boldsymbol{u}), u_\tau(\boldsymbol{u})]$, which by definition consists of all points $\boldsymbol{x} \in a(\boldsymbol{u})$ satisfying both

$$\mathsf{P}\left(\boldsymbol{u}'\boldsymbol{X} \le \boldsymbol{u}'\boldsymbol{x} | \boldsymbol{X} \in a(\boldsymbol{u})\right) \le \frac{1+\tau}{2}$$

and

$$\mathsf{P}\left(\boldsymbol{u}'\boldsymbol{X} \le \boldsymbol{u}'\boldsymbol{x} | \boldsymbol{X} \in a(\boldsymbol{u})\right) \ge \frac{1-\tau}{2}.$$

These intervals are then used to define the τ-quantile region $C^W_{\boldsymbol{X}}(\tau)$ of $\boldsymbol{X}$,

$$C^W_{\boldsymbol{X}}(\tau) = \bigcup_{\|\boldsymbol{u}\|=1} [l_\tau(\boldsymbol{u}), u_\tau(\boldsymbol{u})],$$

that has its coverage probability $P(\boldsymbol{X} \in C^W_{\boldsymbol{X}}(\tau))$ equal to τ. Of course, $\|\cdot\|$ symbolizes the Euclidean norm.

If $\boldsymbol{X} = (X_1, \ldots, X_p)' \in \mathbb{R}^p$ is a general random vector with location parameter $\boldsymbol{\mu}$ and scale diagonal matrix $\mathbf{S}$, then the τ-quantile region $C^W_{\boldsymbol{X}}(\tau)$ of $\boldsymbol{X}$ can be obtained as follows:

$$C^W_{\boldsymbol{X}}(\tau) = \boldsymbol{\mu} + \mathbf{S}\, C^W_{\mathbf{S}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}(\tau) = \left\{\boldsymbol{\mu} + \mathbf{S}\boldsymbol{x} : \boldsymbol{x} \in C^W_{\mathbf{S}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}(\tau)\right\}.$$

Wei (2008) uses vector $\boldsymbol{\mu} = (\mathsf{med}(X_1), \ldots, \mathsf{med}(X_p))'$ of componentwise medians and diagonal matrix $\mathbf{S} = \mathsf{diag}(\mathsf{MAD}(X_1), \ldots, \mathsf{MAD}(X_p))$ created from componentwise median absolute deviations, for example.

The article of Wei (2008) also discusses the inclusion of covariates and an application to the construction of growth charts. Unfortunately, the theory presented there lacks important details. Furthermore, the resulting population multivariate quantile regions are not fully affine equivariant due to the particular choice of $\boldsymbol{\mu}$ and $\boldsymbol{S}$ functionals. On top of that, they need not have star shapes and well-behaving boundaries, see figure B.1 of Wei (2008). What is worse, the problematic and strange contour behavior crucially depends on the point used for centering and cannot be escaped in case of asymmetric distributions for small quantile levels τ. It is because that contour behavior can be observed whenever the directional interval $[l_\tau(\boldsymbol{u}), u_\tau(\boldsymbol{u})]$ does not contain the origin for some direction $\boldsymbol{u}$. Needless to say that the recommended estimation algorithm then fails.

The multivariate quantile concept presented here is similar to that of Wei (2008), but it avoids the normalization of $\boldsymbol{X}$ and uses a different definition of the directional τ-quantile intervals (based on rays, i.e., half-lines) to construct quantile regions around any predetermined center point $\boldsymbol{\mu}$ that need not be related to any distributional symmetry or to $\boldsymbol{X}$ at all. Therefore, it might be better to speak about reference regions, but the whole article calls even such sets as quantiles for the sake of simplicity and brevity.

The difference between rays (half-lines) and lines in the definition of the directional τ-quantile intervals is subtle but very important because it ensures that they always contain the origin, which excludes the strange or problematic behavior of the resulting quantile contours for *any* continuous distribution, *any* quantile level, and *any* center point in the interior of the distributional support. And it produces star-shaped quantile regions with continuous quantile contours having many good properties, including equivariance with respect to affine and many other transformations; see Theorems 1, 3, and 4. Recall that a set is called star-shaped (also star domain, star-convex, or radially convex) if it contains a center point such that all the line segments linking it to other points of the set also lie in the set. Such sets conveniently generalize the convex ones while remaining connected.

Furthermore, the approach presented here uses a different (and simpler) estimation method and a different (and simpler) regression generalization, both based on locally polynomial quantile regression.

As for Petersen (2003), it considers the same ray-based population quantile regions as this paper, mentions their right overall coverage probability, and suggests their locally constant/linear estimation, but all that only in the bivariate location case and with $\boldsymbol{\mu}$ taken as a center of the distribution. This article thus extends the proposal of Petersen (2003) to a general *multidimensional* and *regression* setting with an *arbitrary* center point $\boldsymbol{\mu}$ and provides a *rigorous theoretical analysis* of the extension.

It is also worth mentioning that the quantile regions discussed here need not be necessarily convex, which distinguishes them from the large majority of other multivariate quantiles described in the literature. Nonconvex quantile regions have been growing popular only recently, see, for example, the proposals of Chen et al. (2009), Hlubinka et al. (2010), Agostinelli and Romanazzi (2011), Paindaveine and Van Bever (2013), Chernozhukov et al. (2016), and Carlier et al. (2016). The population concept presented here seems to outperform all those nonconvex proposals in terms of probability interpretation (having both overall and directional meaning) and equivariance properties (going far beyond full affine equivariance and certain monotone equivariance; see Theorem 3).

2 | PRELIMINARY CONSIDERATIONS

Assume a univariate continuous random variable X with distribution function F, quantile function

$$F^{-1}(\tau) = \inf\{x : \ F(x) \geq \tau\}, \quad 0 < \tau < 1,$$

and a center point μ that can then be used to define two conditional functions F_l and F_r associated with the only two unit directions available in the univariate case, namely -1 and 1 (or, left and right):

$$\begin{aligned} F_l(x) &= \mathsf{P}(X > x | X \le \mu) = (F(\mu) - F(x))/F(\mu) && \text{for } x \le \mu, \\ F_l(x) &= 0 && \text{for } x > \mu, \\ F_r(x) &= \mathsf{P}(X \le x | X > \mu) = (F(x) - F(\mu))/(1 - F(\mu)) && \text{for } x > \mu, \\ F_r(x) &= 0 \quad \text{for } x \le \mu. \end{aligned}$$

Consequently,

$$F_l^{-1}(\tau) = F^{-1}(F(\mu) - \tau F(\mu)) \quad \text{and} \quad F_r^{-1}(\tau) = F^{-1}(F(\mu) + \tau - \tau F(\mu)),$$

and the intervals $[F_l^{-1}(\tau), F_r^{-1}(\tau)]$ have their coverage probability equal to τ regardless of μ:

$$\mathsf{P}\left(F_l^{-1}(\tau) \le X \le F_r^{-1}(\tau)\right) = \tau, \quad \tau \in (0,1).$$

In contrast, Wei (2008) would result in univariate intervals

$$[F^{-1}(1/2 - \tau/2), F^{-1}(1/2 + \tau/2)]$$

(regardless of μ and also with the coverage probability τ) that coincide with $[F_l^{-1}(\tau), F_r^{-1}(\tau)]$ only if $\mu = F^{-1}(1/2)$.

The multivariate quantiles presented here attempt to generalize intervals $[F_l^{-1}(\tau), F_r^{-1}(\tau)]$ to the general multivariate case where a random vector $\boldsymbol{X} = (X_1, \dots, X_p)' \in \mathbb{R}^p$, $p \ge 2$, follows a continuous probability distribution with a center point $\boldsymbol{\mu}(\boldsymbol{X}) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_p)' \in \mathbb{R}^p$.

The only problem to be solved lies in the fact that $\mathsf{P}(\boldsymbol{X} \in \{\boldsymbol{\mu} + t\boldsymbol{u},\ t \ge 0\})$ is zero for any direction $\boldsymbol{u} \in \mathbb{R}^p$. Fortunately, this issue can be overcome in an elegant way by means of the hyperspherical coordinate system centered at $\boldsymbol{\mu}(\boldsymbol{X})$ where $\boldsymbol{X}$ can be expressed by means of the hyperspherical coordinates vector $(R, \boldsymbol{U}')'$, resp. $(R, \boldsymbol{\Phi}')' \equiv (R, (\Phi_1, \dots, \Phi_{p-1})')'$, where $R \in [0, \infty)$ denotes the Euclidean distance between $\boldsymbol{X}$ and $\boldsymbol{\mu}(\boldsymbol{X})$, $\boldsymbol{U} \in \mathcal{S}^{p-1} = \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\| = 1\}$ stands for the unit vector from $\boldsymbol{\mu}(\boldsymbol{X})$ in the direction of $\boldsymbol{X}$, and $\boldsymbol{\Phi} = (\Phi_1, \dots, \Phi_{p-1})' \in \Omega$ is the angular vector uniquely describing the unit vector $\boldsymbol{U}$. Here, $\Omega = [0, \pi]^{p-2} \times [0, 2\pi) + \boldsymbol{\omega}$ is used for parametrizing $\mathcal{S}^{p-1}$ where often $\boldsymbol{\omega} = \boldsymbol{0}$ by convention but any other $\boldsymbol{\omega} = (0, \dots, 0, \omega)'$, $\omega \in \mathbb{R}$, could be used as well.

Each of the two parametrizations, by unit directions and by angles, has its own appeal. On the one hand, the use of (manifold-valued) unit vectors is very intuitive and straightforward. On the other hand, the angles are usually easy to handle mathematically. Consequently, both parametrizations are used when appropriate.

As far as $\boldsymbol{\mu}(\boldsymbol{X})$ is concerned, its knowledge is still required. As the location functional $\boldsymbol{\mu}$ is naturally assumed to be shift equivariant, $\boldsymbol{\mu}(\boldsymbol{X})$ is considered zero without any loss of generality hereinafter. (Otherwise, one would have to use the observations centered with $\boldsymbol{\mu}$.)

Needless to say that there are plenty of shift-equivariant location functionals $\boldsymbol{\mu}$ including the mode, mean, or various medians (Small, 1990) of the multivariate distribution. For example, the mean and the medians induced by the simplicial, halfspace, and projection depths (Serfling & Zuo, 2000) are all fully affine equivariant, and their regression modifications are already available to be used in the regression extensions mentioned below; see, for example, Hallin and Šiman (2017) for a review of some available options, and also Šiman (2011) and Boček and Šiman (2017) with the references therein.

The particular choice of $\boldsymbol{\mu}$ may depend on further distributional assumptions and take robustness and computational aspects into consideration. Alternatively, the center point $\boldsymbol{\mu}$ may be known in advance or given by external circumstances, and the task may be to construct a safe/just set around it with a given probability coverage where the safety/justice is interpreted in terms of the same conditional directional probabilities in all directions from $\boldsymbol{\mu}$. It should be stressed that $\boldsymbol{\mu}$ is called a center point only because it (in a sense) lies in the center of all the quantile regions (not necessarily in the center of the distribution). Its choice influences robustness and some equivariance properties of the presented quantile regions and is itself quite a separate problem. It may also follow from the practical applications. For example, it may also be the position of the observer, the source, the reference point, or the target/ideal point, depending on the particular context.

The transformation $\psi : (r, \boldsymbol{\varphi}) \mapsto \boldsymbol{x} = (x_1, \dots, x_p)'$ from the hyperspherical coordinates to the Cartesian ones is continuous and straightforward:

$$
\begin{aligned}
x_1 &= ru_1 = r \sin \varphi_1 \sin \varphi_2 \ \dots \ \sin \varphi_{p-2} \sin \varphi_{p-1}, \\
x_2 &= ru_2 = r \sin \varphi_1 \sin \varphi_2 \ \dots \ \sin \varphi_{p-2} \cos \varphi_{p-1}, \\
&\dots \\
x_{p-1} &= ru_{p-1} = r \sin \varphi_1 \cos \varphi_2, \\
x_p &= ru_p = r \cos \varphi_1.
\end{aligned}
$$

The Jacobian determinant $J(r, \boldsymbol{\varphi})$ corresponding to ψ is

$$
J(r, \boldsymbol{\varphi}) = r^{p-1} \sin^{p-2}(\varphi_1) \ \dots \ \sin(\varphi_{p-2}).
$$

The inverse transformation $\psi^{-1} : \mathbb{R}^p \to [0, \infty] \times \Omega$ is unique except for some special border cases with zero probability to occur. In other words, it is unique for $\boldsymbol{x} = \psi(r, \boldsymbol{\phi})$ for some $(r, \boldsymbol{\phi}) \in (0, \infty) \times \Omega^\circ$ where Ω° denotes the interior of Ω.

The hyperspherical coordinates describe the distance from $\boldsymbol{\mu}(\boldsymbol{X}) = \boldsymbol{0}$ by r and the directions $\boldsymbol{u}$ by $\boldsymbol{\varphi}$. The joint density $p(r, \boldsymbol{\varphi})$ of R and $\boldsymbol{\Phi}$ as well as the marginal density $s(\boldsymbol{\varphi})$ of $\boldsymbol{\Phi}$ and the conditional density $q(r|\boldsymbol{\varphi})$ of R given $\boldsymbol{\Phi} = \boldsymbol{\varphi}$ can then be expressed as follows:

$$
\begin{aligned}
p(r, \boldsymbol{\varphi}) &= f(\psi(r, \boldsymbol{\varphi}))\, |J(r, \boldsymbol{\varphi})| = f(\psi(r, \boldsymbol{\varphi}))\, r^{p-1} \sin^{p-2}(\varphi_1) \ \dots \ \sin(\varphi_{p-2}), \\
s(\boldsymbol{\varphi}) &= \int_0^{+\infty} p(\rho, \boldsymbol{\varphi})\, \mathrm{d}\rho, \\
I(\boldsymbol{\varphi}) &= \int_0^{\infty} f(\psi(\rho, \boldsymbol{\varphi}))\, \rho^{p-1}\, \mathrm{d}\rho, \\
q(r|\boldsymbol{\varphi}) &= \begin{cases} \frac{p(r, \boldsymbol{\varphi})}{s(\boldsymbol{\varphi})} = \frac{f(\psi(r, \boldsymbol{\varphi})) r^{p-1}}{I(\boldsymbol{\varphi})} & \text{if } s(\boldsymbol{\varphi}) \in (0, \infty), \\ \frac{f(\psi(r, \boldsymbol{\varphi})) r^{p-1}}{I(\boldsymbol{\varphi})} & \text{if } s(\boldsymbol{\varphi}) = 0 \text{ and } I(\boldsymbol{\varphi}) \in (0, \infty), \\ 0 & \text{in the other cases.} \end{cases}
\end{aligned}
$$

The introduction of $I(\boldsymbol{\varphi})$ makes it possible to continuously extend the definition of $q(r|\boldsymbol{\varphi})$ to those $\boldsymbol{\varphi}$'s where $s(\boldsymbol{\varphi})$ is zero only due to $J(r, \boldsymbol{\varphi}) = 0$.

After this spadework, the following definition of multivariate directional quantiles must appear very straightforward because it only reduces to finding conditional quantiles of R given $\boldsymbol{\Phi}$.

3 | THEORY

In this section, assume a random vector $\boldsymbol{X} = (X_1, \ldots, X_p)' \in \mathbb{R}^p$, $p \geq 2$, with $\boldsymbol{\mu}(\boldsymbol{X}) = \boldsymbol{0}$ and with a continuous density $f(\boldsymbol{x})$ positive on an open set $\mathcal{M}$, and write $(R, \boldsymbol{\Phi}')'$ for the hyperspherical coordinates of the random vector $\boldsymbol{X}$.

Definition 1. Consider the conditional distribution function $Q(r|\boldsymbol{\varphi})$ and the conditional quantile function $Q^{-1}(\tau|\boldsymbol{\varphi})$ of the radius $\boldsymbol{R}$ given $\boldsymbol{\Phi} = \boldsymbol{\varphi}$ for any $\boldsymbol{\varphi} \in \Omega$ with $I(\boldsymbol{\varphi}) \in (0, \infty)$ as follows:

$$Q(r|\boldsymbol{\varphi}) = \mathsf{P}(\boldsymbol{R} \leq r|\boldsymbol{\Phi} = \boldsymbol{\varphi}) = \int_0^r q(\rho|\boldsymbol{\varphi})\, \mathrm{d}\rho,\ r \in [0, \infty),$$

and

$$Q^{-1}(\tau|\boldsymbol{\varphi}) = \inf\{r \geq 0 :\ Q(r|\boldsymbol{\varphi}) \geq \tau\},\ \tau \in (0, 1).$$

Then the directional τ-quantile $\theta_\tau(\boldsymbol{\varphi})$ (or $\theta_\tau(\boldsymbol{u})$ for the direction $\boldsymbol{u} \in \mathcal{S}^{p-1}$ described by $\boldsymbol{\varphi}$) is defined as

$$\theta_\tau(\boldsymbol{\varphi}) = \psi\left(Q^{-1}(\tau|\boldsymbol{\varphi}), \boldsymbol{\varphi}\right),\ \tau \in (0, 1).$$

In other words, $\theta_\tau(\boldsymbol{\varphi})$ is the point that lies in the distance of $Q^{-1}(\tau|\boldsymbol{\varphi})$ from the origin in the direction determined by the angle $\boldsymbol{\varphi}$.

In fact, analogous directional quantile approximations can be defined directly in the Cartesian coordinates if the rays and their beginning parts are respectively replaced with the cones $\mathcal{A}^\delta(\boldsymbol{u})$ and their bounded versions $\mathcal{A}^\delta(\boldsymbol{u}, r)$ having vertex $\boldsymbol{0}$, aperture $2\delta > 0$ and the axis of symmetry parallel with the direction $\boldsymbol{u} \in \mathcal{S}^{p-1}$:

$$\mathcal{A}^\delta(\boldsymbol{u}) = \{\boldsymbol{x} \in \mathbb{R}^p :\ \measuredangle(\boldsymbol{x}, \boldsymbol{u}) \leq \delta\} \quad \text{and}$$
$$\mathcal{A}^\delta(\boldsymbol{u}, r) = \{\boldsymbol{x} \in \mathbb{R}^p :\ \measuredangle(\boldsymbol{x}, \boldsymbol{u}) \leq \delta,\ \|\boldsymbol{x}\| \leq r\},$$

where $\measuredangle(\boldsymbol{v}_1, \boldsymbol{v}_2) \in [0, \pi]$ stands for the angle between two p-dimensional vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, $\cos(\measuredangle(\boldsymbol{v}_1, \boldsymbol{v}_2)) = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle / (\|\boldsymbol{v}_1\| \|\boldsymbol{v}_2\|)$.

Definition 2. Consider

$$F_\delta(r|\boldsymbol{u}) = \mathsf{P}\left(\boldsymbol{X} \in \mathcal{A}^\delta(\boldsymbol{u}, r) \mid \boldsymbol{X} \in \mathcal{A}^\delta(\boldsymbol{u})\right)$$

and corresponding inverse function $F_\delta^{-1}(\tau|\boldsymbol{u}),\ \tau \in (0, 1)$. Then the directional τ-quantile approximation $\Theta_\tau^\delta(\boldsymbol{u})$ can be defined for any $\tau \in (0, 1)$ and in any unit direction $\boldsymbol{u} \in \mathbb{R}^p$ as follows:

$$\Theta_\tau^\delta(\boldsymbol{u}) = F_\delta^{-1}(\tau|\boldsymbol{u}) \cdot \boldsymbol{u}.$$

The $\limsup_{\delta \to 0+} \Theta_\tau^\delta(\boldsymbol{u})$ is so important that it deserves a special symbol, say $\Theta_\tau(\boldsymbol{u})$.

For mutually corresponding $\boldsymbol{\varphi}$ and $\boldsymbol{u}$, $\limsup_{\delta \to 0+} F_\delta(r|\boldsymbol{u})$ should intuitively play the same role as $Q(r|\boldsymbol{\varphi})$ and, therefore, $\theta_\tau(\boldsymbol{\varphi})$ should have the same meaning as $\Theta_\tau(\boldsymbol{u})$. Indeed, these directional quantiles often coincide according to the next theorem.

Theorem 1. *Suppose that the random vector* $\boldsymbol{X}$ *has a probability distribution with a continuous and bounded density f satisfying*

$$\lim_{r\to+\infty}\sup_{\|\boldsymbol{u}\|=1}\int_r^{+\infty}\rho^{p-1}f(\rho\boldsymbol{u})\,\mathrm{d}\rho=0, \tag{1}$$

that is positive on a set $\mathcal{M}$, $\boldsymbol{\mu}(\boldsymbol{X})=\boldsymbol{0}\in\mathrm{int}(\mathcal{M})$. *Then always* $\theta_\tau(\boldsymbol{\varphi})=\Theta_\tau(\boldsymbol{u})$ *for corresponding* $\boldsymbol{\varphi}\in\Omega^\circ$ *and* $\boldsymbol{u}\in\mathcal{S}^{p-1}$.

Proof. Arbitrarily fix $r\in[0,\infty)$ and $\boldsymbol{\varphi}\in\Omega^\circ$ without any loss of generality. Furthermore, define $U_\delta\equiv U_\delta(\boldsymbol{\varphi})$ by $\psi(1,U_\delta)=\mathcal{A}^\delta(\boldsymbol{u})\cap\mathcal{S}^{p-1}$ and set $S_\delta:=\int_{U_\delta}1\,\mathrm{d}\boldsymbol{\eta}$.

Continuous f implies continuous p and bounded f implies bounded p on any compact set. Note also that $s(\boldsymbol{\varphi})<\infty$ thanks to (1) and that $s(\boldsymbol{\varphi})>0$ because $\boldsymbol{0}\in\mathrm{int}(\mathcal{M})$ and $\boldsymbol{\varphi}\in\Omega^\circ$. Obviously, $s(\boldsymbol{\varphi})$ is absolutely integrable (as every density function).

In fact, $s(\boldsymbol{\varphi})$ must be continuous in $\boldsymbol{\varphi}$, which easily results from Heine's definition of continuity. That is to say that, for any $\varepsilon>0$, there exists $r_0>0$ such that

$$\left|\sup_{\boldsymbol{\varphi}}\int_{r_0}^{+\infty}p(\rho,\boldsymbol{\varphi})\,\mathrm{d}\rho\right|\le\sup_{\boldsymbol{\varphi}}\int_{r_0}^{+\infty}\rho^{p-1}f(\psi(\rho,\boldsymbol{\varphi}))\,\mathrm{d}\rho<\varepsilon$$

thanks to (1). Consequently, for any sequence $\{\boldsymbol{\varphi}_n\}_n$ converging to $\boldsymbol{\varphi}$, there exists n_0 such that $n>n_0$ implies

$$\begin{aligned}|s(\boldsymbol{\varphi}_n)-s(\boldsymbol{\varphi})|\le{}&\left|\int_0^{r_0}p(\rho,\boldsymbol{\varphi}_n)\,\mathrm{d}\rho-\int_0^{r_0}p(\rho,\boldsymbol{\varphi})\,\mathrm{d}\rho\right|\\&+\left|\int_{r_0}^{+\infty}p(\rho,\boldsymbol{\varphi}_n)\,\mathrm{d}\rho\right|+\left|\int_{r_0}^{+\infty}p(\rho,\boldsymbol{\varphi})\,\mathrm{d}\rho\right|<3\varepsilon,\end{aligned}$$

and thus $s(\boldsymbol{\varphi}_n)\xrightarrow{n\to+\infty}s(\boldsymbol{\varphi})$.

If $\delta\to0$, then $\boldsymbol{\eta}\in U_\delta\to\boldsymbol{\varphi}$, $p(\rho,\boldsymbol{\eta})\to p(\rho,\boldsymbol{\varphi})$ for any $\rho\in[0,\infty)$ thanks to the continuity of p, and

$$\frac{s(\boldsymbol{\eta})}{s(\boldsymbol{\varphi})}\to1,\ \frac{1}{S_\delta}\int_{U_\delta}\frac{s(\boldsymbol{\xi})}{s(\boldsymbol{\varphi})}\,\mathrm{d}\boldsymbol{\xi}\to1,\ \int_{U_\delta}\frac{1}{\int_{U_\delta}s(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}}\,\mathrm{d}\boldsymbol{\eta}\to\frac{1}{s(\boldsymbol{\varphi})}\in(0,\infty).$$

Therefore,

$$\int_0^r\left[p(\rho,\boldsymbol{\eta})-p(\rho,\boldsymbol{\varphi})\frac{1}{S_\delta}\int_{U_\delta}\frac{s(\boldsymbol{\eta})}{s(\boldsymbol{\varphi})}\,\mathrm{d}\boldsymbol{\eta}\right]\,\mathrm{d}\rho\xrightarrow{\delta\to0}0.$$

Consequently,

$$\begin{aligned}&\mathsf{P}\left(R\le r|\boldsymbol{\Phi}\in U_\delta(\boldsymbol{\varphi})\right)-\mathsf{P}\left(R\le r|\boldsymbol{\Phi}=\boldsymbol{\varphi}\right)\\&=\frac{\int_{U_\delta}\int_0^r p(\rho,\boldsymbol{\eta})\,\mathrm{d}\rho\,\mathrm{d}\boldsymbol{\eta}}{\int_{U_\delta}s(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}}-\frac{\int_{U_\delta}\int_0^r p(\rho,\boldsymbol{\varphi})\,\mathrm{d}\rho\,\mathrm{d}\boldsymbol{\eta}}{\int_{U_\delta}s(\boldsymbol{\varphi})\,\mathrm{d}\boldsymbol{\xi}}\end{aligned}$$

$$= \int_{U_\delta} \frac{1}{\int_{U_\delta} s(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}} \left(\int_0^r \left[p(\rho, \boldsymbol{\eta}) - p(\rho, \boldsymbol{\varphi}) \frac{1}{S_\delta} \int_{U_\delta} \frac{s(\boldsymbol{\xi})}{s(\boldsymbol{\varphi})}\,\mathrm{d}\boldsymbol{\xi} \right] \mathrm{d}\rho \right) \mathrm{d}\boldsymbol{\eta}$$

$$\xrightarrow{\delta \to 0} 0$$

thanks to the preceding considerations, which completes the proof. ∎

The assumption (1) can be replaced with a stronger condition that may be easier to check:

Theorem 2. *If*

$$\exists \gamma > 0 : \quad \sup_{\|\boldsymbol{u}\|=1} \int_0^{+\infty} \rho^{p-1+\gamma} f(\rho \boldsymbol{u})\,\mathrm{d}\rho < +\infty, \tag{2}$$

then (1) holds.

Proof. If (2) holds for some $\gamma > 0$, then

$$\sup_{\|\boldsymbol{u}\|=1} \int_r^{+\infty} \rho^{p-1} f(\rho \boldsymbol{u})\,\mathrm{d}\rho \le \frac{1}{r^\gamma} \sup_{\|\boldsymbol{u}\|=1} \int_0^{+\infty} \rho^{p-1+\gamma} f(\rho \boldsymbol{u})\,\mathrm{d}\rho \xrightarrow{r \to +\infty} 0.$$

∎

Although (2) need not hold even for $\boldsymbol{X}$ with a continuous distribution having all moments finite, it is satisfied for many common distributions including all the elliptical distributions with probability density $f(\boldsymbol{x}) \propto g(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})$ such that $\boldsymbol{\Sigma}$ is a positive definite matrix, $g : [0, \infty) \to [0, \infty)$, and $\int_0^{+\infty} \rho^{p-1+\gamma} g(\rho^2)\,\mathrm{d}\rho < \infty$ for some $\gamma > 0$, for example, for the multivariate normal distribution with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$.

Theorem 1 also guarantees that $\theta_\tau(\boldsymbol{\varphi})$ is always correctly defined because its assumptions imply $I(\boldsymbol{\varphi}) \in (0, \infty)$ for any $\boldsymbol{\varphi} \in \Omega$. Unfortunately, the condition (1) cannot be ignored or omitted entirely even in a simple bivariate case:

Example 1. Assume a bivariate random vector $\boldsymbol{X}$ with independent marginals and with the (bounded and continuous) probability density $f(x_1, x_2) = h(x_1)h(x_2)$ centrally symmetric around $\boldsymbol{0}$ where

$$h(t) = \begin{cases} \frac{1}{4t^2} & \text{if } |t| > 1, \\ \frac{1}{4} & \text{if } |t| \le 1. \end{cases}$$

The center of symmetry $\boldsymbol{0}$ can be naturally considered as the central point $\boldsymbol{\mu}(\boldsymbol{X})$. Then (1) is violated, $I(\boldsymbol{u}) = \infty$ for $\boldsymbol{\varphi}$ corresponding to $\boldsymbol{u} = (\pm 1, 0)'$ or $\boldsymbol{u} = (0, \pm 1)'$, and, therefore, $\theta_\tau(\boldsymbol{u})$ cannot be correctly defined for those directions.

The directional quantiles also induce multivariate quantile regions and contours:

Definition 3. Consider random vector $\boldsymbol{X} \in \mathbb{R}^p$ with directional τ-quantiles $\theta_\tau(\boldsymbol{u})$, $\tau \in (0, 1)$, $\boldsymbol{u} \in \mathcal{S}^{p-1}$. Then the multivariate τ-quantile regions $\mathcal{C}_{\boldsymbol{X}}(\tau)$ of $\boldsymbol{X}$ can be defined in the following way:

$$\mathcal{C}_{\boldsymbol{X}}(\tau) = \{\rho \boldsymbol{u} : \; 0 \le \rho \le \|\theta_\tau(\boldsymbol{u})\|, \; \boldsymbol{u} \in \mathcal{S}^{p-1}\}.$$

Their borders $\partial \mathcal{C}_{\boldsymbol{X}}(\tau)$ are called τ-quantile contours.

These τ-quantile regions and contours have some favorable properties:

Theorem 3. *Let $\boldsymbol{X} \in \mathbb{R}^p$ be a random vector satisfying the assumptions of Theorem 1. Then*

(i) $\mathsf{P}\left(\boldsymbol{X} \in C_{\boldsymbol{X}}(\tau)\right) = \tau \quad \forall \tau \in (0, 1)$,
(ii) $C_{\boldsymbol{X}}(\tau_1) \subseteq C_{\boldsymbol{X}}(\tau_2)$ *if* $0 < \tau_1 \leq \tau_2 < 1$,
(iii) $C_{\mathbf{c}+\boldsymbol{X}}(\tau) = \mathbf{c} + C_{\boldsymbol{X}}(\tau)$ *for any* $\mathbf{c} \in \mathbb{R}^p$ *if* $\boldsymbol{\mu}(\mathbf{c} + \boldsymbol{X}) = \mathbf{c} + \boldsymbol{\mu}(\boldsymbol{X})$,
(iv) $C_{\mathbf{A}\boldsymbol{X}}(\tau) = \mathbf{A}C_{\boldsymbol{X}}(\tau)$ *for any invertible matrix* $\mathbf{A} \in \mathbb{R}_{p\times p}$ *if* $\boldsymbol{\mu}(\mathbf{A}\boldsymbol{X}) = \mathbf{A}\boldsymbol{\mu}(\boldsymbol{X})$,
(v) $C_{T_r(\boldsymbol{X})}(\tau) = T_r(C_{\boldsymbol{X}}(\tau))$ *for any map* $T_r : \boldsymbol{x} = \psi(r, \boldsymbol{\varphi}) \mapsto \psi(t_r(r, \boldsymbol{\varphi}), \boldsymbol{\varphi})$ *such that* $\boldsymbol{\mu}(T_r(\boldsymbol{X})) = T_r(\boldsymbol{\mu}(\boldsymbol{X}))$, $t_r(0, \cdot) = \mathbf{0}$, *and* $t_r(r, \boldsymbol{\varphi})$ *is continuous and strictly increasing in the first coordinate,*
(vi) $C_{T_{\boldsymbol{\varphi}}(\boldsymbol{X})}(\tau) = T_{\boldsymbol{\varphi}}(C_{\boldsymbol{X}}(\tau))$ *for any map* $T_{\boldsymbol{\varphi}} : \boldsymbol{x} = \psi(r, \boldsymbol{\varphi}) \mapsto \psi(r, t_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}))$ *such that* $\boldsymbol{\mu}(T_{\boldsymbol{\varphi}}(\boldsymbol{X})) = T_{\boldsymbol{\varphi}}(\boldsymbol{\mu}(\boldsymbol{X}))$ *and* $t_{\boldsymbol{\varphi}} : \Omega \to \Omega$ *is an arbitrary continuous bijection, and*
(vii) if $\boldsymbol{X}$ *contains* $\mu(\boldsymbol{X})$ *in the interior of its support, then any* $C_{\boldsymbol{X}}(\tau)$, $\tau \in (0, 1)$, *also contains* $\mu(\boldsymbol{X})$ *in its interior and is star-shaped about it.*

Proof. The properties (ii), (iii), (iv), (vi), and (vii) follow directly from the definition while (v) also relies on the monotone equivariance of univariate quantiles. As for (i),

$$\begin{aligned}\mathsf{P}\left(\boldsymbol{X} \in C_{\boldsymbol{X}}(\tau)\right) &= \int_0^{2\pi}\int_0^{\pi}\dots\int_0^{\pi}\int_0^{Q^{-1}(\tau|\boldsymbol{\varphi})} q(r|\boldsymbol{\varphi})s(\boldsymbol{\varphi})\,\mathrm{d}r\,\mathrm{d}\varphi_1\dots\,\mathrm{d}\varphi_{p-1}\\ &= \tau\int_0^{2\pi}\int_0^{\pi}\dots\int_0^{\pi} s(\boldsymbol{\varphi})\,\mathrm{d}\varphi_1\dots\,\mathrm{d}\varphi_{p-1} = \tau.\end{aligned}$$

■

In other words, the coverage probability of τ-quantile regions is truly equal to τ. Furthermore, the quantile regions are always nested with increasing τ and inherit the equivariance properties from the functional $\boldsymbol{\mu}$. Of course, (iii)–(vi) may be combined or applied repeatedly.

The directional τ-quantiles often continuously depend on the direction for any $\tau \in (0, 1)$:

Theorem 4. *Assume a random vector $\boldsymbol{X} \in \mathbb{R}^p$ satisfying the assumptions of Theorem 1. Then the function $\boldsymbol{u} \mapsto \theta_\tau(\boldsymbol{u})$ is continuous on $\mathcal{S}^{p-1}$ for any $\tau \in (0, 1)$.*

Proof. Fix $\tau \in (0, 1)$ and $\boldsymbol{u}_0 \in \mathcal{S}^{p-1}$ corresponding to $\boldsymbol{\varphi}_0$. As $\theta_\tau(\boldsymbol{\varphi}) = \psi\left(Q^{-1}(\tau|\boldsymbol{\varphi}), \boldsymbol{\varphi}\right)$, it is enough to prove that $Q^{-1}(\tau|\boldsymbol{\varphi}_n) \xrightarrow{n\to+\infty} Q^{-1}(\tau|\boldsymbol{\varphi}_0)$ whenever $\boldsymbol{\varphi}_n \xrightarrow{n\to+\infty} \boldsymbol{\varphi}_0$, which holds if $Q(r|\boldsymbol{\varphi}_n) \xrightarrow{n\to+\infty} Q(r|\boldsymbol{\varphi}_0)$ for any such sequence $\{\boldsymbol{\varphi}_n\}$ according to lemma 8.3.1 of Resnick (1999). But

$$Q(r|\boldsymbol{\varphi}_n) = \int_0^r q(\rho|\boldsymbol{\varphi}_n)\,\mathrm{d}\rho \xrightarrow{n\to+\infty} \int_0^r q(\rho|\boldsymbol{\varphi}_0)\,\mathrm{d}\rho = Q(r|\boldsymbol{\varphi}_0)$$

because $q(\rho|\boldsymbol{\varphi})$ is continuous in $\boldsymbol{\varphi}$ and bounded on compact sets. ■

4 | ESTIMATION

Consider a random sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ of size n that comes from a continuous distribution with a center point $\mathbf{0}$ (without any loss of generality). For example, the distribution may have zero expectation or be centrally or angularly symmetric around the origin. Alternatively, the origin may be the point around which the desired quantile or tolerance region should be constructed. Furthermore, R_i and $\boldsymbol{\Phi}_i$ denote the corresponding hyperspherical coordinates of $\boldsymbol{X}_i$, $i = 1, \dots, n$.

Definition 3 suggests that the sample multivariate τ-quantile region $\hat{C}_{\boldsymbol{X}}(\tau)$ can be obtained as

$$\hat{C}_{\boldsymbol{X}}(\tau) = \{\rho\boldsymbol{u} : \ 0 \le \rho \le \|\hat{\theta}_\tau(\boldsymbol{u})\|, \ \boldsymbol{u} \in \mathcal{S}^{p-1}\},$$

where $\|\hat{\theta}_\tau(\boldsymbol{u})\| = \|\hat{\theta}_\tau(\boldsymbol{\varphi})\| = \hat{Q}^{-1}(\tau|\boldsymbol{\varphi})$ where $\hat{Q}^{-1}(\tau|\boldsymbol{\varphi})$ is a positive estimator of the positive conditional τ-quantile of R given $\boldsymbol{\Phi} = \boldsymbol{\varphi}$ (where $\boldsymbol{\varphi}$ corresponds to $\boldsymbol{u}$). If, as usual, there is no apriori knowledge about the dependence of R on $\boldsymbol{\Phi}$, then the conditional quantile should be estimated nonparametrically. And the standard way to do that is to use a nonparametric quantile regression technique in hyperspherical coordinates with angular regressors.

In principle, any such method could be used in a straightforward way, and each of them has its own advantages and disadvantages. In fact, Wei (2008) has already used the estimation based on splines in a very similar context. Nevertheless, something different is suggested and explored here, namely the locally polynomial quantile regression of (Yu & Jones, 1997; Yu & Jones, 1998). It is mainly because of its simplicity, easy incorporation of trend regressors (in Section 6), and elimination of the boundary effect thanks to the centering trick for the bivariate case described below. It is also the choice of Petersen (2003). The literature generally recommends to employ only locally constant or linear trends. The method works with special n-dependent kernel weights $w_i = w_{i,n,\tau,\boldsymbol{\varphi}_0,K,\mathbb{H}}$ defined for any angle $\boldsymbol{\varphi}_0$ as

$$w_i = w_i(\boldsymbol{\Phi}_i) = \det(\mathbb{H}_n)^{-1} K\left(\mathbb{H}_n^{-1}(\boldsymbol{\Phi}_i - \boldsymbol{\varphi}_0)\right),$$

where $\mathbb{H}_n$ is a symmetric positive definite $(p-1) \times (p-1)$ n-dependent bandwidth matrix, possibly also dependent on τ, $\boldsymbol{\varphi}_0$, and K, and K stands for a $(p-1)$-variate kernel density function, for example, for the Gaussian kernel K_G or the Epanechnikov kernel K_E:

$$K_G(\boldsymbol{\psi}) \propto \exp\left(-\boldsymbol{\psi}'\boldsymbol{\psi}/2\right) \quad \text{and} \quad K_E(\boldsymbol{\psi}) \propto (1 - \boldsymbol{\psi}'\boldsymbol{\psi})\mathrm{I}_{[\boldsymbol{\psi}'\boldsymbol{\psi} \le 1]}.$$

Usually, $\mathbb{H}_n$ is considered to be a diagonal matrix composed of marginal bandwidths h_n^i, $i = 1, \dots, p-1$.

Assume any $\tau \in (0,1)$ and consider the τ-quantile check function $\varrho_\tau(t) = t(\tau - \mathrm{I}(t < 0)) = \max\{(\tau-1)t, \tau t\}$. Then the locally constant estimator $\hat{Q}_0^{-1}(\tau|\boldsymbol{\varphi}_0) = \hat{a}_\tau$ of $Q^{-1}(\tau|\boldsymbol{\varphi}_0)$ minimizes, by definition,

$$\Psi^0_{\tau,w}(a) = \frac{1}{n}\sum_{i=1}^{n} w_i \varrho_\tau(R_i - a),$$

and the locally linear estimator of the same quantity is defined as $\hat{Q}_1^{-1}(\tau|\boldsymbol{\varphi}_0) = \hat{a}_\tau$ where $(\hat{a}_\tau, \hat{\boldsymbol{b}}_\tau')'$ minimizes

$$\Psi^1_{\tau,w}(a, \boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^{n} w_i \varrho_\tau\left(R_i - a - \boldsymbol{b}'(\boldsymbol{\Phi}_i - \boldsymbol{\varphi}_0)\right).$$

Then $\hat{\boldsymbol{b}}_\tau$ often contains the information about the first derivatives of $Q^{-1}(\tau|\boldsymbol{\varphi})$ (as a function of $\boldsymbol{\varphi}$) at $\boldsymbol{\varphi}_0$.

The only problem may be with the estimation for $\boldsymbol{\varphi}_0$ close to the border of Ω. Then one may employ special tools such as asymmetric kernels. Fortunately, there also exists an elegant solution to this problem in the bivariate case with scalar φ_0 that uses $\Omega = [\varphi_0 - \pi, \varphi_0 + \pi)$ for the parametrization in the polar coordinates. This possibility is mentioned in Section 2, uses the information from neighboring data points in full, and improves the estimation results in finite samples considerably. If the bandwidth is not too large (and ideally also if the kernel function has a bounded support), that is, if the weights w_i are close to zero at the boundary of Ω, then the boundary effect is not an issue at all. The effect is often cited as the main reason why the locally linear quantile regression variant should be favored over the locally constant one; see Yu and Jones (1997). The latter option, therefore, seems as a viable and recommended choice, especially if R does not change with the direction too wildly.

This trick can be partly applied even in $\mathbb{R}^p$, $p > 2$, with $\boldsymbol{\varphi}_0 = (\varphi_{1,0}, \dots, \varphi_{p-1,0})'$ when $\Omega = [0, \pi]^{p-2} \times [\varphi_{p-1,0} - \pi, \varphi_{p-1,0} + \pi)$ may be used for the parametrization which, however, does not eliminate problems with all border points. The rest might be dealt with by renaming (i.e., changing the order of) the coordinates before the parametrization but this is, admittedly, neither elegant nor practical.

The asymptotic theory for the locally constant and linear estimators described above is already available, see, for example, Bhattacharya and Gangopadhyay (1990), Chaudhuri (1991a), Koenker and Zhao (1996), Yu and Jones (1998), Honda (2000), Gannoun et al. (2003), Ioannides (2004), Yu and Lu (2004), Zhou and Wu (2009), Kong et al. (2010), Guerre and Sabbah (2012) or Qu and Yoon (2015). The same holds for the rules how to choose the bandwidth matrix; see, for example, Yu and Jones (1998), Yu and Lu (2004), and Gannoun et al. (2003). Of course, one could also choose it via cross-validation or select it subjectively. All the results can be applied directly, with response variable R and vector regressor $\boldsymbol{\Phi}$.

Recall that quantile crossing is absent in the locally constant case (Koenker, 2005, section 2.5). Therefore, the estimated locally constant τ-quantile regions are still nested with decreasing $\tau \in (0, 1)$. They are consistent estimators of their population counterparts under mild conditions permitting even weakly dependent data, and their asymptotic overall and directional coverage probabilities are then equal to τ.

Unfortunately, real observations need not always be centered around the origin (or another known point $\boldsymbol{\mu}$). Nevertheless, the whole concept could be used even if $\boldsymbol{\mu}$ were replaced with a consistent estimator. Note also that the bandwidth may be chosen $\boldsymbol{u}$-dependent without any problem, which may be convenient, for example, when $\boldsymbol{\mu}$ does not lie in the central area of the underlying distribution.

5 | CONVERGENCE OF THE ESTIMATED CONTOURS

The consistency of the aforementioned locally constant and locally linear estimators follows directly from the theory already available in the literature on uniform convergence of the locally polynomial regression quantile estimators, without the need to prove anything new. In particular, the results presented here are special cases of corollary 1 of Guerre and Sabbah (2012). They use the notation introduced so far and the following assumptions.

Assumption 1. The probability density function $s(\boldsymbol{\varphi})$ of $\boldsymbol{\Phi}$ exists and is strictly positive and continuously differentiable over its compact support, say Ω_1, which contains in its interior a compact subset Ω_0.

Assumption 1 is quite standard for the uniform consistency (in $\boldsymbol{\varphi}$) of the locally polynomial estimators. Local uniform consistency would require only local strict positivity and local continuous differentiability.

Assumption 2. The probability density function $q(r|\boldsymbol{\varphi})$ of R given $\boldsymbol{\Phi} = \boldsymbol{\varphi}$ exists and is continuous and strictly positive for any $r \geq 0$ and any $\boldsymbol{\varphi} \in \Omega_1$. The partial derivative $\partial Q(r|\boldsymbol{\varphi})/\partial \boldsymbol{\varphi}$ of the conditional cumulative distribution function $Q(r|\boldsymbol{\varphi})$ is continuous over $[0, \infty) \times \Omega_1$. Furthermore, there exists some $L_0 > 0$ such that the conditional density $q(r|\boldsymbol{\varphi})$ is Lipschitz continuous in both arguments:

$$|q(r_1|\boldsymbol{\varphi}_1) - q(r_2|\boldsymbol{\varphi}_2)| \leq L_0 \|(r_1, \boldsymbol{\varphi}_1')' - (r_2, \boldsymbol{\varphi}_2')'\|,$$

for all $r_1, r_2 \in [0, \infty)$, and all $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2 \in \Omega_1$.

Assumption 2 ensures that $Q^{-1}(\tau|\boldsymbol{\varphi})$ and $\theta_\tau(\boldsymbol{\varphi})$ are uniquely defined for any $\tau \in (0, 1)$. The strict positivity of $q(r|\boldsymbol{\varphi})$ may be relaxed to $q(r|\boldsymbol{\varphi}) > 0$ for $0 \leq r < b(\boldsymbol{\varphi})$ and $q(r|\boldsymbol{\varphi}) = 0$ for $r \geq b(\boldsymbol{\varphi})$ where $b(\boldsymbol{\varphi})$ is a smooth function. The Lipschitz continuity of the conditional density $q(r|\boldsymbol{\varphi})$ holds, for example, if both the joint density function $p(r, \boldsymbol{\varphi})$ and the (marginal) density function $s(\boldsymbol{\varphi})$ are bounded and Lipschitz continuous and simultaneously $s(\boldsymbol{\varphi})$ is uniformly bounded away from zero.

Assumption 3. The non-negative kernel function K is Lipschitz continuous over $\mathbb{R}^{p-1}$, has a compact support $\mathcal{K}$, integrates exactly to one: $\int K(\boldsymbol{\varphi})\, d\boldsymbol{\varphi} = 1$, and satisfies $K \geq k_0 > 0$ for some $k_0 > 0$ on an open unit ball centered at $\mathbf{0}$.

Assumption 3 is also standard and quite general.

For notational simplicity, consider $\boldsymbol{\nu} = (\nu_1, \dots, \nu_{p-1})' \in \mathbb{N}^{p-1}$, write $|\boldsymbol{\nu}|$ for $\sum_{i=1}^{p-1} \nu_i$, and define

$$b_{\boldsymbol{\nu}}(\tau|\boldsymbol{\varphi}) = \frac{\partial^{|\boldsymbol{\nu}|} Q^{-1}(\tau|\boldsymbol{\varphi})}{\partial \varphi_1^{\nu_1} \times \cdots \times \partial \varphi_{p-1}^{\nu_{p-1}}}.$$

Furthermore, write $\lfloor t \rfloor$ for the integer part of $t > 0$.

Definition 4. Consider $\tau \in (0, 1)$. It is said that $Q^{-1}(\tau|\boldsymbol{\varphi})$ is in class $C(t)$ iff

1. the mapping $\boldsymbol{\varphi} \mapsto Q^{-1}(\tau|\boldsymbol{\varphi})$ is continuously differentiable of order $\lfloor t \rfloor$ over Ω_1, that is, $b_{\boldsymbol{\nu}}(\tau|\boldsymbol{\varphi})$ exists and is continuous there for all $\boldsymbol{\nu} \in \mathbb{N}^{p-1}$ such that $|\boldsymbol{\nu}| \leq \lfloor t \rfloor$. Quite naturally, the partial derivatives are considered one-sided if $\boldsymbol{\varphi}$ lies in the border of Ω_1.
2. there exists some $L > 0$ such that it holds for all $\boldsymbol{\nu} \in \mathbb{N}^{p-1}$ with $|\boldsymbol{\nu}| = \lfloor t \rfloor$ and for all $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2 \in \Omega_1$ that

$$|b_{\boldsymbol{\nu}}(\tau|\boldsymbol{\varphi}_1) - b_{\boldsymbol{\nu}}(\tau|\boldsymbol{\varphi}_2)| \leq L \|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|^{t - \lfloor t \rfloor}.$$

Let $\widehat{Q}_c^{-1}(\tau|\boldsymbol{\varphi})$ be the local polynomial estimator of $Q^{-1}(\tau|\boldsymbol{\varphi})$ by the polynomial function of order $c \in \{0, 1\}$. In other words, $c = 0$ and $c = 1$ lead to the locally constant and locally linear estimators introduced in the previous section. Corollary 1 of Guerre and Sabbah (2012) results in

Proposition 1. *Suppose that Assumptions 1, 2 and 3 hold, $c \in \{0,1\}$, the assumptions of Theorem 1 are satisfied and $Q^{-1}(\tau|\boldsymbol{\varphi})$ is in the class $C(t)$ for some $t > 0$ with $\lfloor t \rfloor \le c$. Then*

$$sup_{\boldsymbol{\varphi}\in\Omega_0} \left|\widehat{Q}_c^{-1}(\tau|\boldsymbol{\varphi}) - Q^{-1}(\tau|\boldsymbol{\varphi})\right| = O_P\left(\frac{\log n}{n}\right)^{t/(2t+p-1)}$$

if h_n is asymptotically proportional to $\left(\log(n)/n\right)^{1/(2t+p-1)}$ and if the bandwidth matrix $\mathbb{H}_n = h_n\mathbb{I}$ is the $(p-1)\times(p-1)$ diagonal matrix with h_n on the diagonal. Then also

$$\sup_{\boldsymbol{\varphi}\in\Omega_0} \left|\widehat{\theta}_\tau^c(\boldsymbol{\varphi}) - \theta_\tau(\boldsymbol{\varphi})\right| \xrightarrow{P} 0,$$

for $\widehat{\theta}_\tau^c(\boldsymbol{\varphi}) := \psi(\widehat{Q}_c^{-1}(\tau|\boldsymbol{\varphi}), \boldsymbol{\varphi})$ thanks to the continuity of ψ.

Consider a distribution spherically symmetric around the origin, with probability density function $f(\boldsymbol{x}) = g(\boldsymbol{x}'\boldsymbol{x})$ where g is Lipschitz continuous, continuously differentiable and strictly positive on $[0,a)$ for some $a \in \mathbb{R}\cup\{\infty\}$ and zero elsewhere. Then $I(\boldsymbol{\varphi}) \in (0,\infty)$ is a finite positive constant thanks to Theorem 1, $q(r|\boldsymbol{\varphi})$ is continuous, and the Lipschitz continuity of $q(r|\boldsymbol{\varphi})$ is equivalent to the Lipschitz continuity of $g(r^2)r^{p-1}$. The spherical shape of the population quantile contours also implies that $Q^{-1}(\tau|\boldsymbol{\varphi})$ is constant in $\boldsymbol{\varphi}$. Note that it is possible to extend such reasoning to nondegenerate elliptical distributions because of the continuous linear transformation behind such a distributional change. Unfortunately, sometimes $s(\boldsymbol{\varphi}) = 0$ if $p > 2$, and then Proposition 1 technically does not directly guarantee the uniform convergence of $\widehat{Q}_c^{-1}(\tau|\boldsymbol{\varphi})$, $c = 0, 1$, on the whole unit sphere, that is, the convergence of the locally constant or linear quantile contour estimate $\widehat{C}_{\boldsymbol{X}}(\tau)$ to its population counterpart in the Hausdorff distance. Nevertheless, $s(\boldsymbol{\varphi}) = 0$ is then only an unfortunate by-product of the angular parametrization of the unit sphere that can be eliminated for any particular $\boldsymbol{\varphi}$ by parametrizing the unit sphere differently. Therefore, one can then still combine the consistency results obtained for a finite number of such reparametrizations to obtain the desired global convergence.

6 | REGRESSION EXTENSIONS

Consider n regression observations $(\boldsymbol{X}_1', \boldsymbol{Z}_1')', \dots, (\boldsymbol{X}_n', \boldsymbol{Z}_n')'$ with the corresponding population conditional distribution $\mathcal{L}(\boldsymbol{X}|\boldsymbol{Z})$ and a center point $\boldsymbol{\mu}(\boldsymbol{X}|\boldsymbol{Z})$. Then one can use the approach described in the previous section to estimate the quantile regions of $\mathcal{L}(\boldsymbol{X}|\boldsymbol{Z}=\boldsymbol{z})$ simply by extending the regressor space and considering regressors $(\boldsymbol{\Phi}_i, \boldsymbol{Z}_i)$, which, however, decreases the consistency rate of the resulting kernel estimators (and thus permits only low-dimensional regressor vectors in practical applications). See Figure 3 for such an application.

There is sometimes a natural and known candidate for $\boldsymbol{\mu}(\boldsymbol{X}|\boldsymbol{Z})$, such as in the very important case of financial return time series that are often assumed to be martingale difference sequences with zero mean conditional distributions due to the efficient market hypothesis; see, for example, section III of LeRoy (1989). In the other cases, the concept can still be used with a consistent estimator of $\boldsymbol{\mu}(\boldsymbol{X}|\boldsymbol{Z})$, for example, for an advanced residual analysis in general regression or time series models.

Of course, it is also possible to combine the locally constant regression in $\boldsymbol{\Phi}$ with the locally linear regression in $\boldsymbol{Z}$ or vice versa. And if the dependence of R on $\boldsymbol{\Phi}$ or $\boldsymbol{Z}$ is known, then the

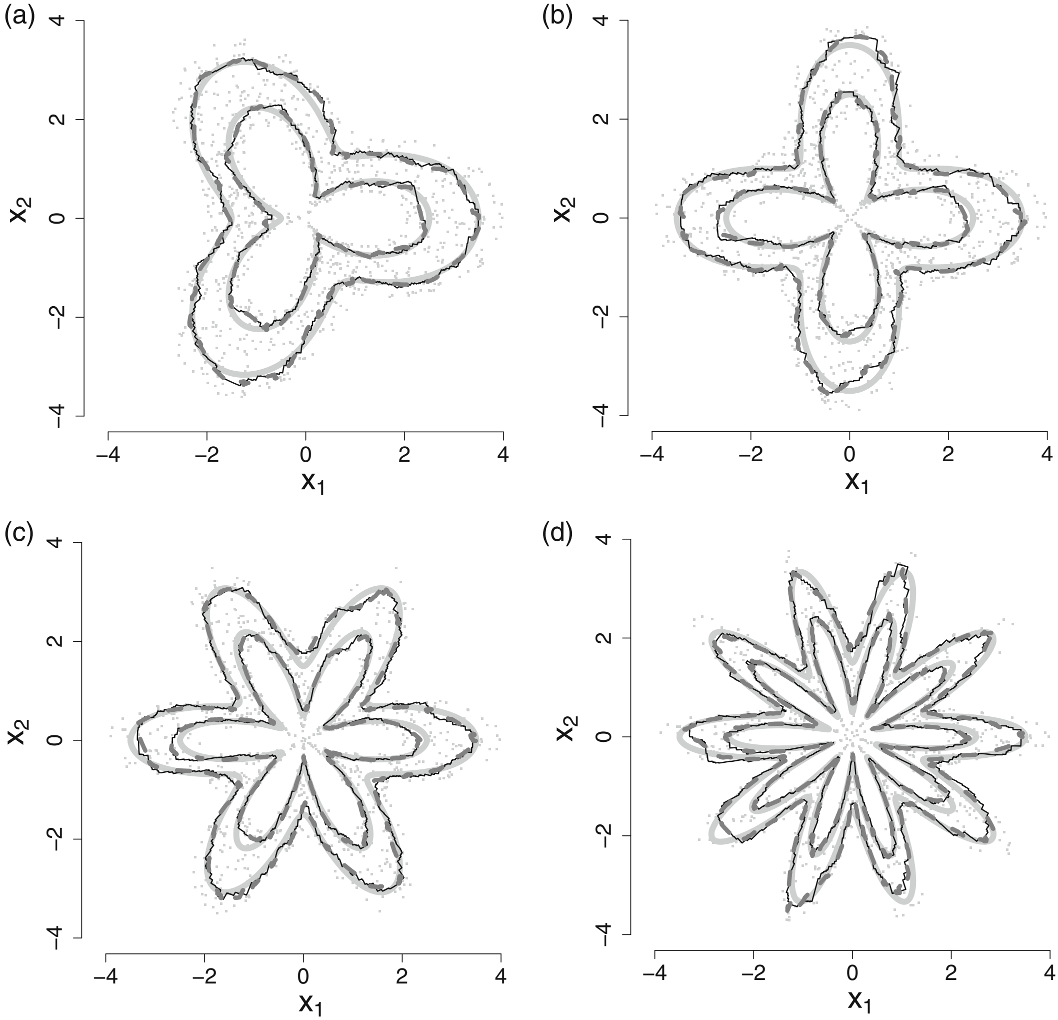

FIGURE 1 Modeling complicated distributions. The figure illustrates the population (thick solid light gray) quantiles $C_{\boldsymbol{X}}(\tau)$, $\tau = 0.25$ and $\tau = 0.75$, of zero-centered $\boldsymbol{X} = (R\cos(\Phi), R\sin(\Phi))'$ where $\mathcal{L}(\Phi) = \mathcal{U}(0, 2\pi)$, $\mathcal{L}(R|\Phi = \varphi) = 1 + 2\cos^2(k\varphi) + \mathcal{U}(-1, 1)$ is independent of $\mathcal{L}(\Phi)$, and (a) $k = 3/2$, (b) $k = 2$, (c) $k = 3$, and (d) $k = 5$. The population τ-quantiles are closely mimicked by their locally constant (thin solid black) and locally linear (dashed dark gray) estimates obtained from $n = 999$ observations with the Gaussian kernel and bandwidth (a) $h_\varphi = 0.12$, (b) $h_\varphi = 0.11$, (c) $h_\varphi = 0.10$, and (d) $h_\varphi = 0.08$

variables can enter the regression parametrically (without influencing the weights) in the correct trend parametrization like in the partially linear quantile regression models (see, e.g., Lee, 2003), which may break their curse of dimensionality.

Although the estimated conditional quantile regions are hard to display visually for $p > 2$, especially for several τ's at once, their approximate volumes, surface areas, and inner points can still be determined, which is what usually matters most. Furthermore, the evolution of the sample conditional quantiles of $\mathcal{L}(\boldsymbol{X}|\boldsymbol{Z} = \boldsymbol{z})$ with $\boldsymbol{z}$ may provide much valuable information about the tails or heteroscedasticity.

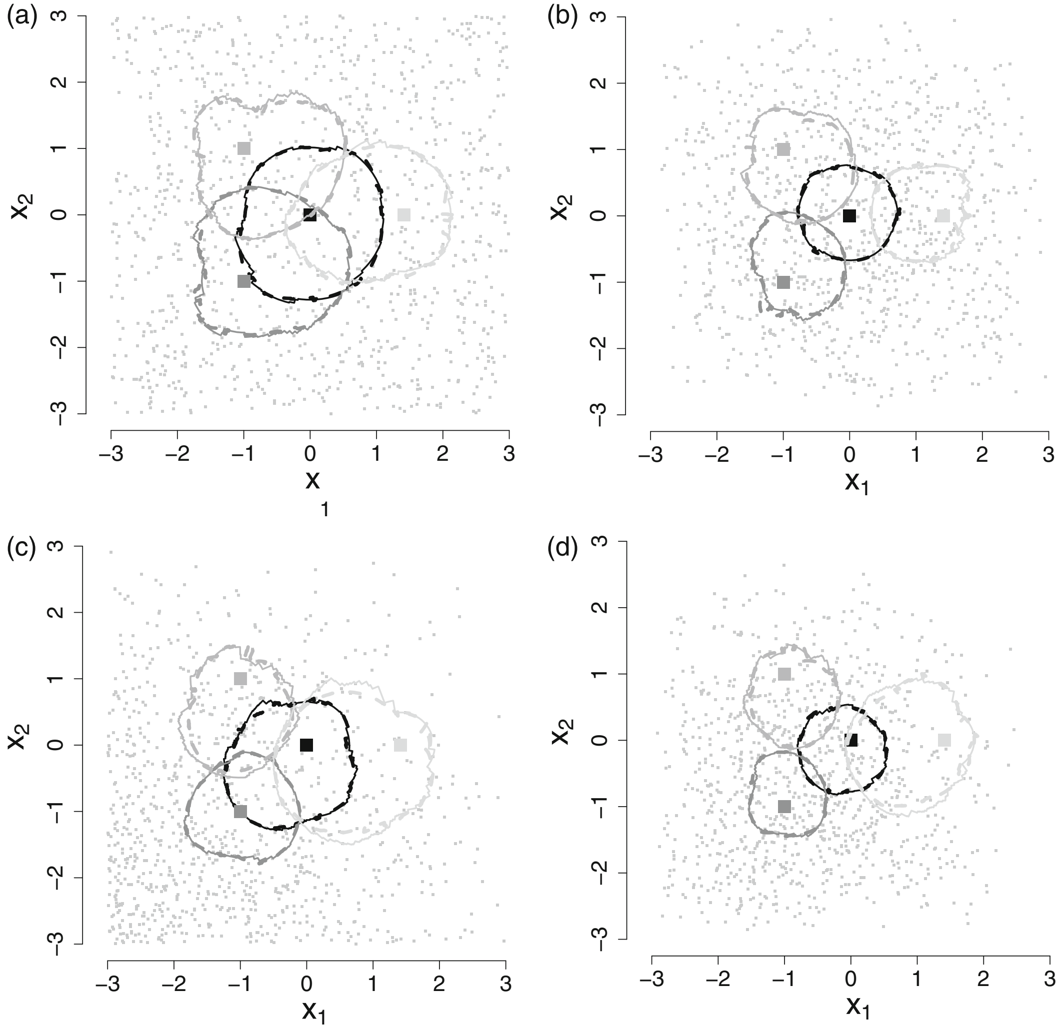

FIGURE 2 Dependence on the center point. The figure shows estimated quantiles $\widehat{C}_{\boldsymbol{X}}(0.1)$ of four bivariate distributions of $\boldsymbol{X}$ with the same independent marginals: (a) 6Beta(1, 1) − 3, (b) 6Beta(2, 2) − 3, (c) 6Beta(1, 2) − 3, (d) 6Beta(2, 3) − 3, obtained from $n = 999$ observations by means of the locally constant (solid) and locally linear (dashed) regression with the Epanechnikov kernel and bandwidth $h_\varphi = \pi/4$ for four different center points: $\boldsymbol{\mu} = (0, 0)'$ (black), $\boldsymbol{\mu} = (-1, -1)'$ (dark gray), $\boldsymbol{\mu} = (-1, 1)$ (gray) and $\boldsymbol{\mu} = (\sqrt{2}, 0)'$ (light gray)

7 | ILLUSTRATIONS

Figure 1 demonstrates that the new multivariate quantiles can naturally describe even various population distributions with nonconvex support. It also shows that both their locally constant and locally linear estimates may perform very well, at least in case of $n = 999$ independent bivariate zero-centered observations $\boldsymbol{X}_i \sim (R_i \cos(\Phi_i), R_i \sin(\Phi_i))'$, $i = 1, \ldots, n$, where $\mathcal{L}(\Phi_i) = \mathcal{U}(0, 2\pi)$ and $\mathcal{L}(R_i|\Phi_i) = 1 + 2\cos^2(k\Phi_i) + \mathcal{U}(-1, 1)$ are independent, and $k = 3/2, 2, 3$, or 5. The estimation was done with the Gaussian kernel for several bandwidths from a dense grid, and only the

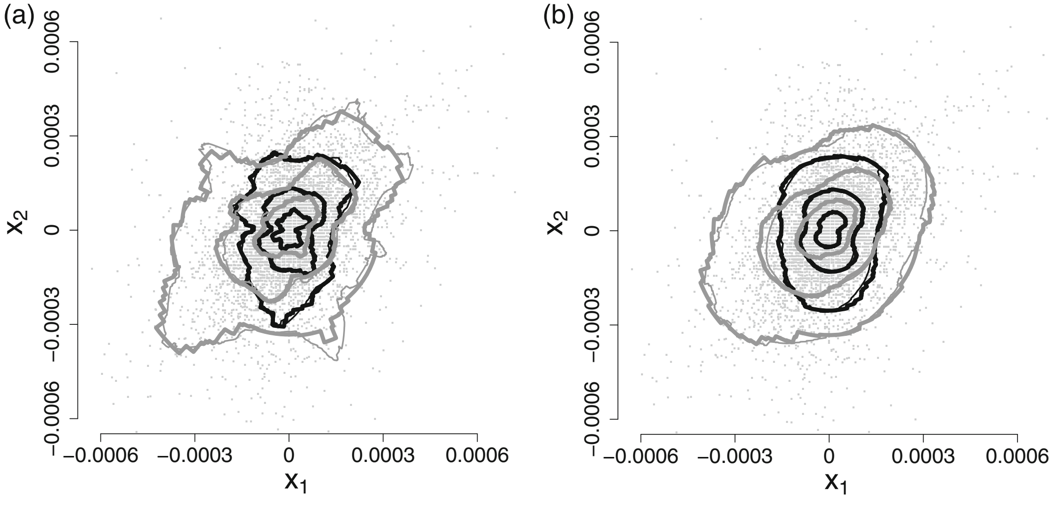

FIGURE 3 Joint modeling of Forex exchange rates. The figure presents $n = 6{,}145$ one-minute log-returns of FX rates EUR/CAD (x_1) and AUD/CAD (x_2) from 02/11/2014 17:00 to 06/11/2014 23:25 where the time is shifted and rescaled to the [0,1] interval for the sake of simplicity. The pictures show the locally constant (thick) and locally linear (thin) τ-quantiles of the log-returns, $\tau \doteq 0.203, 0.560$, and 0.853, obtained for times $t_1 = 0.50$ (black) and $t_2 = 0.85$ (dark gray) with the Gaussian kernel and the diagonal bandwidth matrix described by marginal bandwidths $h_t = 0.04$ and (a) $h_\varphi = 0.25$ or (b) $h_\varphi = 0.50$. Obviously, h_φ controls the smoothness of the resulting quantile contours. The conditional distributions of the two log-returns at t_1 and t_2 apparently differ not only in the scale, but also in the correlation structure

visually most attractive results are reported, for $h_\varphi = 0.12$, (b) $h_\varphi = 0.11$, (c) $h_\varphi = 0.10$, and (d) $h_\varphi = 0.08$ decreasing with k. The same approach to the choice of bandwidth was adopted even in the following examples.

Figure 2 shows that the quantile shapes remain reasonable regardless the choice of the quantile center point $\boldsymbol{\mu}$. It employs four bivariate distributions on $[-3, 3]^2$ with the same independent marginals generated by means of the beta distribution: $6\text{Beta}(1, 1) - 3$, $6\text{Beta}(2, 2) - 3$, $6\text{Beta}(1, 2) - 3$, and $6\text{Beta}(2, 3) - 3$, and four different center points: $\boldsymbol{\mu} = (0, 0)'$, $\boldsymbol{\mu} = (-1, -1)'$, $\boldsymbol{\mu} = (-1, 1)$, and $\boldsymbol{\mu} = (\sqrt{2}, 0)'$ to estimate the quantiles from $n = 999$ observations by means of the locally constant and locally linear regression with the Epanechnikov kernel and ad hoc bandwidth $h_\varphi = \pi/4$.

Figure 3 illustrates the new multivariate *regression* quantiles and their application to modeling conditional heteroscedasticity of financial time series, namely of $n = 6{,}145$ one-minute log-returns of FX rates EUR/CAD and AUD/CAD from 02/11/2014 17:00 to 06/11/2014 23:25 where the single regressor, time, is considered shifted and rescaled to the [0,1] interval for the sake of simplicity. The locally constant and locally linear conditional zero-centered τ-quantile estimates, $\tau \doteq 0.203, 0.560$, and 0.853, behave similarly and jointly confirm that the conditional bivariate return distributions at $t_1 = 0.50$ and $t_2 = 0.85$ are dramatically different regarding the scale and correlation structure. The peculiar quantile levels are used for easy comparison with the empirical conditional halfspace depth contours of the same data in Boček and Šiman (2017). The figure also illuminates how the choice of bandwidth can influence the smoothness and precision of the resulting quantile contour estimates.

8 | CONCLUDING REMARKS

This article generalizes the bivariate location quantile concept of Petersen (2003) to general multidimensional regression quantiles (or reference regions) with an arbitrary center point and theoretically investigates the new proposal (and thus also the original one as a special case).

The extended quantile concept seems very successful in modeling certain simulated and real data, and it has quite a few favorable properties: equivariance, simplicity, shape flexibility, excellent probability interpretation, star-shaped quantile regions with smooth contours under mild conditions, and their simple estimation based on the well-established theory and practice of (single-response) quantile regression in the most important bivariate case. The proposed estimation procedure becomes complicated in spaces beyond dimension two for some points with angular coordinates close to the border of the set Ω used for parametrization. Working directly with unit vectors as regressors might circumvent the problem at the cost of introducing manifold-valued regressors and related complications.

Unfortunately, the nonparametric estimation methods deteriorate with the dimension of (regressor) observations in terms of both the computational time and the consistency rate of the resulting estimators. Furthermore, the new method works with respect to a center point, which may be viewed as its advantage or drawback, depending on the context.

The presented quantile concept can be recommended to practitioners if they have enough observations of two-to-three dimensional responses and a clear idea of how to choose or estimate the center point of their interest. If the regression dependence of responses is assumed significant and nonparametric, then the number of regressors should also be very small such as one or two. The scope of possible applications is thus somewhat limited, as is the case of nonparametric regression methods in general. Nevertheless, it includes three important cases, namely residual analysis and modeling conditional heteroscedasticity in financial return time series (where the natural center point is zero) and the theory of multivariate process capability indices (where the natural center point is the target); see Kotz and Lovelace (1998) or Pearn and Kotz (2006). There they could serve, for example, for defining new process capability indices as in Šiman (2014a,b).

Fully or partially parametric specifications for the quantile shapes and center points may break the curse of dimensionality, allow for the estimation of the center point, and make the method generally applicable in multidimensional spaces. Such extensions will hopefully be addressed elsewhere because they require quite a different approach and come up with different problems to be solved.

ACKNOWLEDGMENTS

All the authors thank anonymous referees for their helpful comments. The research of Daniel Hlubinka was supported by PRIMUS/17/SCI/03 Project of Charles University. The research of Miroslav Šiman was supported by the Czech Science Foundation projects GA17-07384S and GA21-05325S.

ORCID

Daniel Hlubinka https://orcid.org/0000-0003-4016-3632

REFERENCES

Agostinelli, C., & Romanazzi, M. (2011). Local depth. *Journal of Statistical Planning and Inference*, *141*, 817–830.

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

Boček, P., & Šiman, M. (2017). On weighted and locally polynomial directional quantile regression. *Computational Statistics*, *32*, 929–946.

Carlier, G., Chernozhukov, V., & Galichon, A. (2016). Vector quantile regression: An optimal transport approach. *Annals of Statistics*, *44*, 1165–1192.

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics*, *53*, 380–403.

Chaudhuri, P. (1991a). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics*, *19*, 760–777.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, *91*, 862–872.

Chen, Y., Dang, X., Peng, H., & Bart, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 288–305.

Cheng, Y., & De Gooijer, J. G. (2007). On the uth geometric conditional quantile. *Journal of Statistical Planning and Inference*, *137*, 1914–1930.

Chernozhukov, V., Galichon, A., Hallin, M., & Henry, M. (2016). Monge-Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, *45*, 223–256.

Gannoun, A., Saracco, J., & Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *Journal of Statistical Planning and Inference*, *117*, 207–223.

Guerre, E., & Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, *28*, 87–129.

Hallin, M., Lu, Z., Paindaveine, D., & Šiman, M. (2015). Local bilinear multiple-output quantile/depth regression. *Bernoulli*, *21*, 1435–1466.

Hallin, M., Paindaveine, D., & Šiman, M. (2010a). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Annals of Statistics*, *38*, 635–669.

Hallin, M., Paindaveine, D., & Šiman, M. (2010b). Rejoinder. *Annals of Statistics*, *38*, 694–703.

Hallin, M., & Šiman, M. (2017). *Multiple-output quantile regression*. In R. Koenker, V. Chernozhukov, X. He, & L. Peng (Eds.), *Handbook of quantile regression*. Chapman & Hall/CRC Press.

Hlubinka, D., Kotík, L., & Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika*, *46*, 125–148.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Annals of the Institute of Statistical Mathematics*, *52*, 459–470.

Ioannides, D. A. (2004). Fixed design regression quantiles for time series. *Statistics & Probability Letters*, *68*, 235–245.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R., & Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, *12*, 793–813.

Koltchinskii, V. (1997). M-estimation, convexity and quantiles. *Annals of Statistics*, *25*, 435–477.

Kong, E., Linton, O., & Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, *26*, 1529–1564.

Kong, L., & Mizera, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statistica Sinica*, *22*, 1589–1610.

Kotz, S., & Lovelace, C. R. (1998). *Process capability indices in theory and practice*. Oxford University Press.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, *19*, 1–31.

LeRoy, S. F. (1989). Efficient capital markets and martingales. *Journal of Economic Literature*, *27*, 1583–1621.

Paindaveine, D., & Šiman, M. (2011). On directional multiple-output quantile regression. *Journal of Multivariate Analysis*, *102*, 193–392.

Paindaveine, D., & Van Bever, G. (2013). From depth to local depth: A focus on centrality. *Journal of the American Statistical Association*, *108*, 1105–1119.

Pearn, W. L., & Kotz, S. (2006). *Encyclopedia and handbook of process capability indices*. World Scientific Publishing.

Petersen, J. H. (2003). Two bivariate geometrically defined reference regions with applications to male reproductive hormones and human growth. *Statistics in Medicine*, *22*, 2603–2618.

Qu, Z., & Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, *185*, 1–19.

Resnick, S. I. (1999). *A probability path*. Birkhäuser.

Serfling, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Statistica Neerlandica*, *56*, 214–232.

Serfling, R., & Zuo, Y. (2000). General notions of statistical depth function. *Annals of Statistics*, *28*, 461–482.

Šiman, M. (2011). On exact computation of some statistics based on projection pursuit in a general regression context. *Communications in Statistics - Simulation and Computation*, *40*, 948–956.

Šiman, M. (2014a). Precision index in the multivariate context. *Communications in Statistics - Theory and Methods*, *43*, 377–387.

Šiman, M. (2014b). Multivariate process capability indices: A directional approach. *Communications in Statistics - Theory and Methods*, *43*, 1949–1955.

Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review*, *58*, 263–277.

Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, *103*, 397–409.

Yu, K., & Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics & Data Analysis*, *25*, 159–166.

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

Yu, K., & Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, *31*, 333–346.

Zhou, Z., & Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Annals of Statistics*, *37*, 2696–2729.

How to cite this article: Hlubinka, D., Kotík, L., & Šiman, M. (2022). Multivariate quantiles with both overall and directional probability interpretation. *Scandinavian Journal of Statistics*, 1–19. https://doi.org/10.1111/sjos.12603