



Directional bivariate quantiles: a robust approach based on the cumulative distribution function

Nadja Klein¹ · Thomas Kneib²

Received: 1 August 2018 / Accepted: 17 August 2019 / Published online: 31 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The definition of multivariate quantiles has gained considerable attention in previous years as a tool for understanding the structure of a multivariate data cloud. Due to the lack of a natural ordering for multivariate data, many approaches have either considered geometric generalisations of univariate quantiles or data depths that measure centrality of data points. Both approaches provide a centre-outward ordering of data points but do no longer possess a relation to the cumulative distribution function of the data generating process and corresponding tail probabilities. We propose a new notion of bivariate quantiles that is based on inverting the bivariate cumulative distribution function and therefore provides a directional measure of extremeness as defined by the contour lines of the cumulative distribution function which define the quantile curves of interest. To determine unique solutions, we transform the bivariate data to the unit square. This allows us to introduce directions along which the quantiles are unique. Choosing a suitable transformation also ensures that the resulting quantiles are equivariant under monotonically increasing transformations. We study the resulting notion of bivariate quantiles in detail, with respect to computation based on linear programming and theoretical properties including asymptotic behaviour and robustness. It turns out that our approach is especially useful for data situations that deviate from the elliptical shape typical for ‘normal-like’ bivariate distributions. Moreover, the bivariate quantiles inherit the robustness of univariate quantiles even in case of extreme outliers.

Thomas Kneib received financial support from the German Research Foundation (DFG) within the research Project KN 922/4-2. Both authors are grateful for the comments provided by two anonymous referees which, in particular, prompted us to incorporate changes in the theoretical results when transforming with the empirical instead of the true marginal cumulative distribution functions.

✉ Nadja Klein
nadja.klein@hu-berlin.de

¹ Humboldt Universität zu Berlin, Unter den Linden 6, 10999 Berlin, Germany

² Georg-August-Universität Göttingen, Göttingen, Germany

Keywords Direction-specific quantiles · Inverse cumulative distribution function · Linear programming · Quantile curves · Robustness

1 Introduction

Quantiles are an important concept in many areas of statistics. For example, they provide a notion of extremeness for the data points of interest, are utilised to determine measures of variability and can be used to derive distribution-free tests in nonparametric statistics. In case of univariate data, quantiles can be characterised in various different ways, e.g. based on inverting the cumulative distribution function (CDF), as minimisers of a weighted absolute deviation loss, based on the centrality of data points as quantified by appropriate depth functions, etc. When going beyond the univariate case, however, these different possibilities yield rather different generalisations of quantiles due to the lack of a natural ordering of observations (see Serfling 2002, for a systematic review of multivariate quantile definitions).

One common way of generalising univariate quantiles is to replace the linear ordering on the line by a centre-outward ordering which requires the appropriate definition of a most central point that replaces the median. An early example is the multivariate simplex median of Oja (1983) obtained by minimising the sum of volumes of simplices with vertices defined by observed data points. More generally, approaches based on data depths utilise nonnegative real-valued depth functions $D(\mathbf{y}, F)$ (with \mathbf{y} being a multivariate data point and F the CDF the data depth is evaluated on) to assign a depth to each observed data point. The deepest point (in the sense of the data depth) then corresponds to the median while the remaining points can be ordered centre-outward based on their depth values (see, among others, Liu et al. 1999; Zuo and Serfling 2000)

Another approach for deriving multivariate quantiles is based on the fact that for a univariate random variable Y with $\mathbb{E}(|Y|) < \infty$ the τ th quantile, $\tau \in (0, 1)$ can be characterised as the minimiser of $\mathbb{E}[|Y - q| - (2\tau - 1)(Y - q)]$, see, e.g., Ferguson (1967). Chaudhuri (1996) suggests the generalisation $\mathbb{E}[\|\mathbf{Y} - \mathbf{q}\| + \langle \mathbf{u}, \mathbf{Y} - \mathbf{q} \rangle]$ to define geometric multivariate quantiles where $\|\cdot\|$ is the Euclidean norm and $\langle \cdot, \cdot \rangle$ the Euclidean inner product while $\mathbf{u} \in B^D(0) = \{\mathbf{u} | \mathbf{u} \in \mathbb{R}^D, \|\mathbf{u}\| < 1\}$ determines the direction as well as the level of the multivariate quantile. More specifically, the quantile \mathbf{q} is ‘central’ (i.e. close to the median), for $\|\mathbf{u}\|$ close to zero and ‘extreme’ for $\|\mathbf{u}\|$ close to one. Earlier attempts in this direction have, for example, been presented by Small (1990) for the special case of the median or Abdous and Theodorescu (1992). Chakraborty (2001) criticises the approach of Chaudhuri (1996) since geometric quantiles are not equivariant under affine transformations such as rotations. Consequently, no sensible estimates are available whenever the different components of the data vectors are measured in different units or when they have different degrees of variability. To overcome this problem, Chakraborty (2001) proposes a transformation–re-transformation approach based on a data-driven coordinate system.

Another class of multivariate quantile definitions relates to an extension of the probability integral transform to a multivariate analogue which maps the multivariate quantile function to a reference distribution (Hallin et al. 2010; Carlier et al. 2016). This concept has been related to depth-based multivariate quantile definitions in Cher-

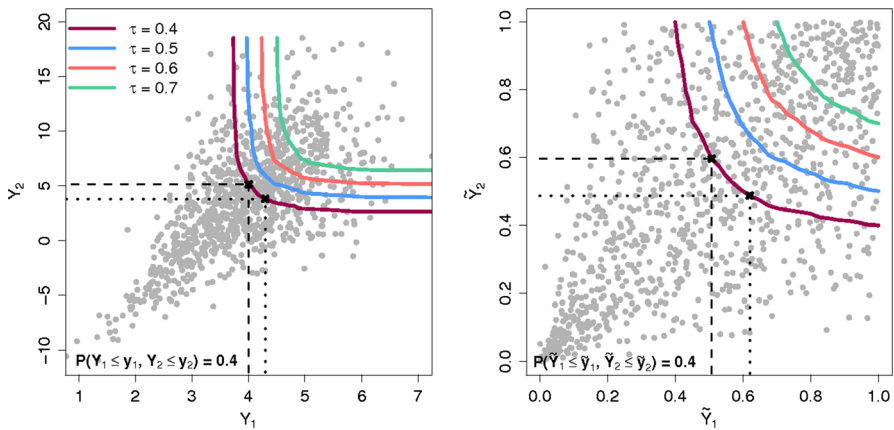


Fig. 1 Illustration of bivariate quantile curves before and after the transformation on the unit square. On the left, the proposed bivariate quantile curves on the original scale of Y are shown for selected quantile levels $\tau \in \{0.4, 0.5, 0.6, 0.7\}$. The grey dots represent a random sample from the joint CDF while the two crosses are exemplary data points on the curve at level $\tau = 0.4$. On the right, we depict the corresponding curves on the unit square (color figure online)

nozhuikov et al. (2017), a Bayesian formulation has been proposed by Guggisberg (2016) and the formulation has also been studied under misspecification (Carlier et al. 2016).

We will follow a different route that considers quantiles as resulting from the inversion of the joint CDF. This approach is related to both the abstract formulation of real-valued quantile processes proposed in Einmahl and Mason (1992) who define the multivariate τ th quantile as the smallest Borel set having a probability of at least τ and the bivariate quantile approach of Chen and Welsh (2002). Here, the authors consider the inversion of the bivariate CDF but decompose it into the inversion of one marginal and one conditional distribution to achieve a unique solution. In this paper, we also focus on the bivariate case but, in contrast to Chen and Welsh (2002), define the bivariate CDF-based quantile curve \mathcal{Q}_τ , $\tau \in (0, 1)$ as the set of points $(q_1, q_2)'$ in \mathbb{R}^2 for which the bivariate CDF is equal to the desired quantile level τ (see the left plot of Fig. 1 for an empirical illustration showing quantile curves for several quantile levels). The main differences between our approach and the proposal by Chen and Welsh (2002) are (1) the transformation on the unit square (where Chen and Welsh 2002 rely on a standardisation step instead) that makes our CDF-based quantiles equivariant under componentwise monotonically increasing transformations, (2) the robustness that we achieve by using this transformation as it avoids the determination of multivariate measures for location and dispersion and (3) the geometric interpretation offered by the polar coordinate-type characterisation of bivariate quantiles based on the angle and the distance with respect to a reference point. Belzunce et al. (2007) and Fernandez-Ponce and Suarez-Llorens (2007) also derive bivariate quantiles from inverting the CDF but they rather focus on central quantile region again that are obtained by combining regions implied by the four directions in \mathbb{R}^2 corresponding to the extreme points of the unit sphere induced by the product topology.

In contrast to most previous multivariate quantile definitions, our approach does not provide a centre-outward but rather a directional notion of extremeness as defined by the contour lines of the joint CDF. While Hallin (2017) argues that multivariate quantiles should be related to ranks in order to allow for distribution-free, rank-based inference, we are developing our novel proposal for applications where the focus is on studying the directional tail behaviour of a distribution. For example, we will investigate the joint distribution of two measures for the nutritional status of children in developing countries later on in this paper, where one is particularly interested in children that are in the joint lower tail of the bivariate distribution since these are the ones most urgently in need of nutritional improvement. This change in focus has to be taken into account when interpreting CDF-based quantile curves. As illustrated in Fig. 1, it is not the area below the quantile curve that has probability mass τ but rather the rectangles defined by the points on the quantile curve. Note also that due to the different notion of extremeness, we are interested in different theoretical properties of bivariate quantiles. For example, invariance under rotations is not of interest since it alters the orientation of the joint CDF. In contrast, we are interested in invariance under componentwise monotonic transformation such that the same children are identified as extremely malnourished, regardless of the precise measurement instrument utilised for the two malnutrition dimensions.

Since the CDF-based quantile curve is no longer a single point as in one dimension and thus cannot be uniquely determined numerically, we introduce an appropriate reference point such that bivariate points can be described in a polar coordinate-type fashion. Similar as in copula-based models, we therefore transform the data to the unit square (see the right plot of Fig. 1) such that each point in the unit square can be characterised by an angle and the distance along that angle from the upper right corner of the unit square. For a given angle, we show that elements of the quantile curve are obtained by minimising an appropriate loss criterion along the angles. Similar to the univariate case, we demonstrate that this can be done efficiently via linear programming techniques.

In summary, the main advantages of our new definition of bivariate quantiles are as follows:

- The bivariate CDF-based quantiles are related to tail probabilities as determined by the bivariate CDF.
- They provide a direct relation to the quantile level τ in terms of a joint probability, whereas the quantile level is only determined implicitly in most alternatives that provide a centre-outward ordering.
- The robustness properties of empirical bivariate CDF-based quantiles are inherited from univariate quantiles leading to the same breakdown point.
- Empirical bivariate CDF-based quantiles can easily be determined by linear programming.
- The univariate quantiles of the marginal distributions are obtained as special cases.

The rest of this paper is structured as follows: Sect. 2 introduces CDF-based bivariate quantiles in more detail and studies their theoretical properties. Section 3 discusses numerical optimisation based on linear programming. Section 4 considers asymptotic results including consistency and asymptotic normality while Sect. 5 provides empir-

ical evaluations of the novel approach both in simulations and an application. Finally, Sect. 6 summarises our findings and discusses avenues for future research such as the generalisation beyond the bivariate case.

2 Bivariate quantiles derived from the cumulative distribution function

2.1 Prerequisites

For the definition of bivariate quantiles, we consider either a bivariate, continuous, real-valued random variable $Y = (Y_1, Y_2)'$ or, for the empirical counterpart, an i.i.d. sample y_1, \dots, y_n of size n from the distribution of Y . Throughout the rest of the paper, we will always assume that the domain \mathcal{D} of Y is rectangular with a positivity constraint on the density $f(y_1, y_2)$, i.e. $f(y_1, y_2) > 0$ for all pairs $(y_1, y_2)' \in \mathcal{D}$. Without loss of generality, we will consider $\mathcal{D} = \mathbb{R}^2$ in the following. The joint density is furthermore assumed to be continuously differentiable such that the joint CDF

$$F(y_1, y_2) = \mathbb{P}(Y_1 \leq y_2, Y_2 \leq y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(y_1, y_2) dy_2 dy_1,$$

is strictly monotonically increasing in both arguments and continuously differentiable. From the joint CDF, the marginal CDFs can be obtained via

$$F_1(y_1) = \lim_{y_2 \rightarrow \infty} F(y_1, y_2), \quad F_2(y_2) = \lim_{y_1 \rightarrow \infty} F(y_1, y_2).$$

For the characterisation of bivariate quantiles based on loss functions, we will furthermore assume that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y_1| f(y_1, y_2) dy_1 dy_2 < \infty$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y_2| f(y_1, y_2) dy_2 dy_1 < \infty$, i.e. existence of first moments for the two components.

2.2 The general set-up

To obtain a bivariate generalisation of univariate quantiles that can be obtained by inverting the univariate CDF, we define the bivariate quantile curve for fixed quantile level τ as follows:

Definition 1 (*Bivariate quantile curve*) For $\tau \in (0, 1)$, the bivariate quantile curve \mathcal{Q}_τ is defined as

$$\mathcal{Q}_\tau = \left\{ \mathbf{q} = (q_1, q_2)' \in \mathbb{R}^2 \mid F(q_1, q_2) = \tau \right\}. \tag{1}$$

If the strict monotonicity assumption for the bivariate CDF is not fulfilled, e.g. for (partially) discrete random vectors, the definition should be relaxed to

$$\mathcal{Q}_\tau = \{ \mathbf{q} \in \mathbb{R}^2 \mid \mathbb{P}(Y_1 \leq q_1, Y_2 \leq q_2) \geq \tau \text{ and } \mathbb{P}(Y_1 \geq q_1, Y_2 < q_2) + \mathbb{P}(Y_1 < q_1, Y_2 \geq q_2) + \mathbb{P}(Y_1 \geq q_1, Y_2 \geq q_2) \geq 1 - \tau \}. \tag{2}$$

To get a better understanding of the concept of bivariate quantile curves, Fig. 2 shows quantile curves for the levels $\tau = 0.1, \dots, 0.9$ for bivariate distributions with standard normal marginals and the Gaussian, the Clayton, the Gumbel and the Frank copula with dependence parameters chosen such that Kendall's tau τ^K equals 0.2, 0.4, 0.6 and 0.8. From these curves, we can draw some conclusions on stylised features represented in bivariate quantile curves:

- For the Gaussian copula with elliptical contour lines, the bivariate quantile curves for different quantile levels but fixed dependence shift in an almost parallel way towards the upper right corner of the domain. Increasing dependence, on the other hand, leads to a sharper kink close to the diagonal line. For all dependence parameters, the quantile curves approach the same value as one of the coordinates approaches ∞ which reflects the fact that marginal quantiles are obtained as special cases in this situation (see below for details).
- For the case of lower tail dependence (illustrated along the Clayton copula), we see strong changes in the shape of the quantile curve over the quantile levels. For small values of τ , we observe the sharp kink on the diagonal line that was associated with strong dependence in case of the normal copula while the quantile curve is more circular for larger values of τ . This exactly fits with the notion of lower tail dependence where there is strong association in the lower tail of the distribution but weaker dependence in the upper tail.
- For the Gumbel copula as a representative of copulas with upper tail dependence, the behaviour observed for Clayton copula reverses, i.e. one observes circular quantile curves for the lower tails and sharp kinks in the quantile curve for the upper tail.
- The Frank copula behaves similarly to the Gaussian copula since it is also invariant under rotations of 180.

To characterise the bivariate quantile curve, we now define the loss function

$$\rho_\tau(\mathbf{y}, \mathbf{q}) = \max(y_1 - q_1, y_2 - q_2) \left(\tau - \mathbb{1}_{\{\max(y_1 - q_1, y_2 - q_2) < 0\}} \right) \quad (3)$$

with $\mathbf{y} = (y_1, y_2)'$, $\mathbf{q} = (q_1, q_2)' \in \mathbb{R}^2$ which provides the bivariate analogue to the check function known from univariate quantile optimisation (Koenker 2005).

Theorem 2 (Loss function for bivariate quantiles) *Under the general assumptions from Sect. 2.1, the bivariate quantile curve \mathcal{Q}_τ is equal to the set of minimisers of the expected loss under ρ_τ , i.e.*

$$\mathcal{Q}_\tau = \left\{ \mathbf{q} \in \mathbb{R}^2 \mid \mathbb{E}(\rho_\tau(\mathbf{Y}, \mathbf{q})) = \min_{\mathbf{v} \in \mathbb{R}^2} \mathbb{E}(\rho_\tau(\mathbf{Y}, \mathbf{v})) \right\}.$$

The proof of Theorem 2 is given in ‘‘Appendix A.1’’. The basic idea is to separate the domain of \mathbf{Y} into appropriate rectangles and to apply the Leibniz rule for parameter integrals twice.

An empirical version of the bivariate quantile curve can easily be defined based on the bivariate empirical CDF. Let therefore $\mathbf{y}_1, \dots, \mathbf{y}_n$ be an i.i.d. sample from the

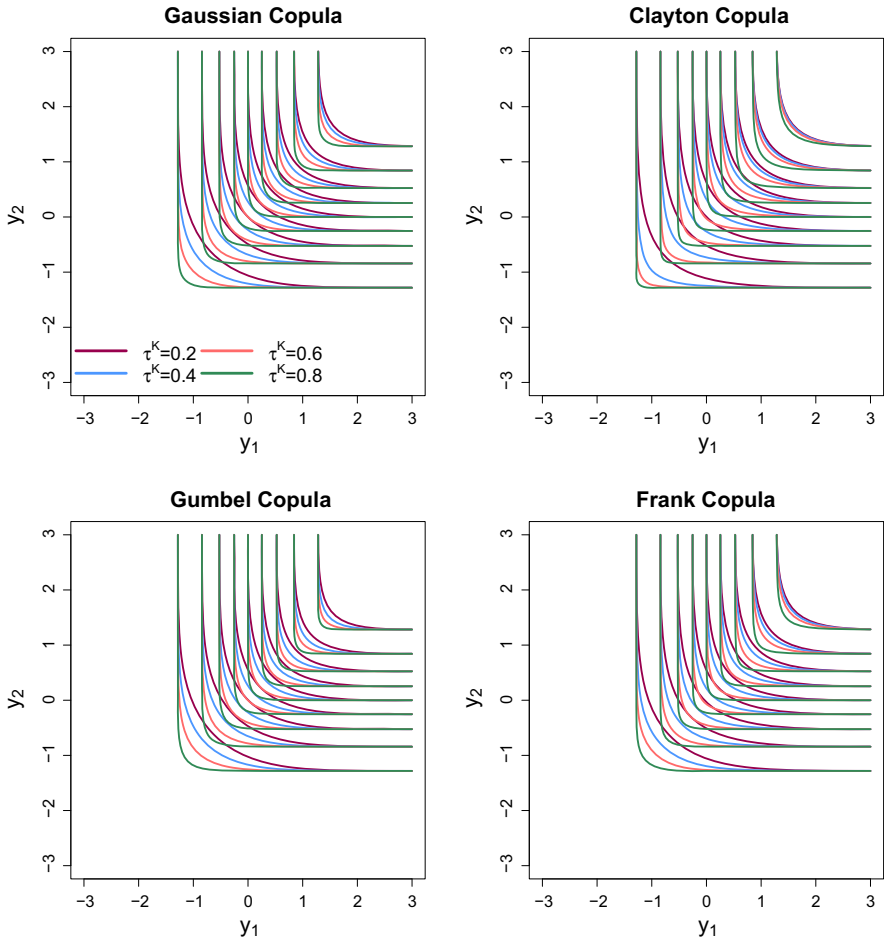


Fig. 2 Illustration of bivariate quantile curves for four parametric copulas C with increasing Kendall's tau $\tau^K = 0.2, 0.4, 0.6, 0.8$. Shown are quantiles $\tau = 0.1, \dots, 0.9$ of the theoretical CDF $F(y_1, y_2) = C(F_1(y_1), F_2(y_2))$. The margins are assumed to be $N(0, 1)$ distributions (color figure online)

bivariate CDF $F: \mathbb{R}^2 \rightarrow [0, 1]$ and define the bivariate empirical CDF as

$$F_n(y_1, y_2) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \{y_{i1} \leq y_1, y_{i2} \leq y_2\}.$$

The empirical bivariate quantile curve $Q_{n,\tau}$ can be either obtained by plugging the empirical CDF into (2) or by using the empirical loss $\sum_{i=1}^n \rho_\tau(y_i, \mathbf{q})$. In fact, we have the following corollary to Theorem 2.

Corollary 3 *The empirical bivariate quantile curve $Q_{n,\tau}$ is equal to the set of minimisers of the empirical loss under ρ_τ , i.e.*

$$Q_{n,\tau} = \left\{ \mathbf{q} \in \mathbb{R}^2 \mid \sum_{i=1}^n \rho_\tau(\mathbf{y}_i, \mathbf{q}) = \min_{\mathbf{v} \in \mathbb{R}^2} \sum_{i=1}^n \rho_\tau(\mathbf{y}_i, \mathbf{v}) \right\}.$$

Even for strictly monotonic bivariate CDFs, the bivariate quantile curve does not reduce to a single scalar value but rather yields a set along the contour lines of the CDF as visualised in Fig. 1. We will therefore consider a transformation of \mathbb{R}^2 to the unit square which allows us to characterise points in a polar coordinate-type parameterisation. This enables the introduction of unique direction-specific quantiles that allow us to estimate Q_τ and to study its theoretical properties.

2.3 Transformation to the unit square

Let $t: \mathbb{R}^2 \rightarrow [0, 1]^2, \mathbf{y} \mapsto t(\mathbf{y}) = (t_1(y_1), t_2(y_2))'$ denote a componentwise, monotonically increasing transformation from the original domain to the unit square. Applying this transformation to the observed data yields

$$\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2)' = (t_1(y_1), t_2(y_2))' = t(\mathbf{y}),$$

and similarly random vectors can be transformed to obtain $\tilde{\mathbf{Y}} = t(\mathbf{Y})$. While for theoretical considerations on the underlying random vectors, we require strict monotonicity (which ensures that t is invertible), any order-preserving univariate monotonically increasing function is a candidate for the components of the transformation t for observed data (which includes piecewise constant functions that do not induce ties in the observations).

To establish a link to copulae (Joe 2014), transformations based on the marginal distributions of the components of \mathbf{Y} are particularly interesting. Such transformations also ensure that the transformed data marginally spread over the complete domain $[0, 1]$ which would not necessarily be the case when relying on some pre-specified CDF that might place most of its probability mass where no data have been observed. For most of the theoretical results that we derive in Sect. 4, we will consider the transformation based on the true marginal CDFs, leading to transformed observations

$$\tilde{y}_{i1} = F_1(y_{i1}), \quad \tilde{y}_{i2} = F_2(y_{i2}), \quad i = 1, \dots, n.$$

In practice however, we will rely on the marginal empirical CDFs

$$F_{j,n}(y_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(y_{ij} \leq y_j), \quad j = 1, 2$$

leading to

$$\tilde{y}_{i1n} = F_{1,n}(y_{i1}), \quad \tilde{y}_{i2n} = F_{2,n}(y_{i2}), \quad i = 1, \dots, n.$$

The latter resembles the definition of empirical copulae (Joe 2014) such that the transformed data contain the information about the dependence structure only but are also inherently dependent due to the transformation with the empirical CDF. This dependence between the transformed observations has to be taken into account when studying statistical properties of the estimates. In Sect. 4, we will therefore also comment on the changes for the theoretical results when transforming with the empirical CDFs.

2.4 Direction-specific CDF-based bivariate quantiles

After having transformed to the unit square, we represent points $\tilde{q} \in [0, 1]^2$ in a polar coordinate-type fashion, where $\alpha \in D_\alpha = [0, \pi/2]$ is the corresponding angle of the distance \tilde{r} to the upper right corner of the unit square. This is illustrated in Fig. 3 for

$$\tilde{q} = (1 - \tilde{r} \cos(\alpha), 1 - \tilde{r} \sin(\alpha))'$$

As a consequence,

$$\begin{aligned} \mathbb{P}(\tilde{Y}_1 \leq \tilde{q}_1, \tilde{Y}_2 \leq \tilde{q}_2) &= \mathbb{P}\left(\tilde{Y}_1 \leq 1 - \tilde{r} \cos(\alpha), \tilde{Y}_2 \leq 1 - \tilde{r} \sin(\alpha)\right) \\ &= \mathbb{P}\left(\frac{1 - \tilde{Y}_1}{\cos(\alpha)} > \tilde{r}, \frac{1 - \tilde{Y}_2}{\sin(\alpha)} > \tilde{r}\right) \\ &= \mathbb{P}\left(\min\left(\frac{1 - \tilde{Y}_1}{\cos(\alpha)}, \frac{1 - \tilde{Y}_2}{\sin(\alpha)}\right) > \tilde{r}\right). \end{aligned}$$

This leads to the following definition from which we derive direction-specific quantiles along a pre-specified angle α :

Definition 4 (*Distance survivor function along α*) For a random vector \tilde{Y} with domain $[0, 1]^2$, we define the distance to the upper right corner of the unit square as

$$\tilde{R} = \min\left(\frac{1 - \tilde{Y}_1}{\cos(\alpha)}, \frac{1 - \tilde{Y}_2}{\sin(\alpha)}\right).$$

The distance survivor function $\tilde{S}_\alpha(\tilde{r}) = \mathbb{P}(\tilde{R} > \tilde{r})$ then coincides with the bivariate CDF of \tilde{Y} evaluated at $\tilde{q} = (1 - \tilde{r} \cos(\alpha), 1 - \tilde{r} \sin(\alpha))'$, i.e.

$$\tilde{S}_\alpha(\tilde{r}) = \mathbb{P}\left(\tilde{Y}_1 \leq 1 - \tilde{r} \cos(\alpha), \tilde{Y}_2 \leq 1 - \tilde{r} \sin(\alpha)\right).$$

The empirical analogue yields

$$\tilde{S}_{n,\alpha}(\tilde{r}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\tilde{y}_{i1} \leq 1 - \tilde{r} \cos(\alpha), \tilde{y}_{i2} \leq 1 - \tilde{r} \sin(\alpha))$$

i.e. the empirical survivor function of the distances

$$r_i = \min \left(\frac{1 - \tilde{y}_{i1}}{\cos(\alpha)}, \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} \right)$$

coincides with the bivariate empirical CDF of $\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_n$ evaluated at $\tilde{\mathbf{q}} = (1 - \tilde{r} \cos(\alpha), 1 - \tilde{r} \sin(\alpha))'$.

We proceed to define direction-specific bivariate quantiles by inverting the survivor function of \tilde{R} for a given value of the angle α .

Definition 5 (*Bivariate quantiles along α*) For $\tau \in (0, 1)$ and $\alpha \in D_\alpha$, the τ th quantile distance along α is defined as

$$\tilde{r}_{\alpha,\tau} = \inf \{ \tilde{r} \in D_{\tilde{r}}(\alpha) \mid \tilde{S}_\alpha(\tilde{r}) = \tau \} \tag{4}$$

where $D_{\tilde{r}}(\alpha)$ denotes the (angle-dependent) domain of \tilde{r} . From this, the corresponding direction-specific bivariate quantile along α can be deduced as

$$\tilde{\mathbf{q}}_{\alpha,\tau} = (1 - \tilde{r}_{\alpha,\tau} \cos(\alpha), 1 - \tilde{r}_{\alpha,\tau} \sin(\alpha))'$$

Similarly, the empirical τ th quantile distance along α is

$$\tilde{r}_{n,\alpha,\tau} = \inf \{ \tilde{r} \in D_{\tilde{r}}(\alpha) \mid \tilde{S}_{n,\alpha}(\tilde{r}) \geq \tau \},$$

and the empirical direction-specific bivariate quantile along α is given by

$$\tilde{\mathbf{q}}_{n,\alpha,\tau} = (1 - \tilde{r}_{n,\alpha,\tau} \cos(\alpha), 1 - \tilde{r}_{n,\alpha,\tau} \sin(\alpha))'. \tag{5}$$

In order to obtain the quantile distances as minimisers of the expected loss function in (3), the latter has to be adapted to account for the transformation on $[0, 1]^2$. To be more precise, since the introduction of α allows us to reduce the original bivariate problem to a set of univariate ones, the distance to be considered can be determined along α leading to the loss function

$$\rho_{\alpha,\tau}(\tilde{\mathbf{y}}, \tilde{r}) = \min(\tilde{v}_1, \tilde{v}_2) (\tau - \mathbb{1}_{\{\min(\tilde{v}_1, \tilde{v}_2) \leq 0\}}) \tag{6}$$

with $\tilde{v}_1 = (1 - \tilde{y}_1) / \cos(\alpha) - \tilde{r}$ and $\tilde{v}_2 = (1 - \tilde{y}_2) / \sin(\alpha) - \tilde{r}$.

Theorem 6 *The theoretical quantile distances $\tilde{r}_{\alpha,\tau}$ and the empirical quantile distance $\tilde{r}_{n,\alpha,\tau}$ can be obtained by minimising*

$$\mathcal{L}_{\alpha,\tau}(\tilde{r}) = \mathbb{E} \left(\rho_{\alpha,\tau}(\tilde{\mathbf{Y}}, \tilde{r}) \right) \tag{7}$$

and

$$\mathcal{L}_{n,\alpha,\tau}(\tilde{r}) = \frac{1}{n} \sum_{i=1}^n \rho_{\alpha,\tau}(\tilde{\mathbf{y}}_i, \tilde{r}). \tag{8}$$

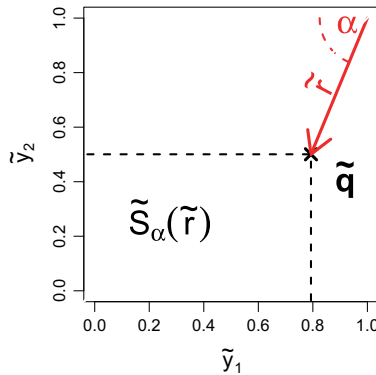


Fig. 3 Transforming an observed data point \tilde{q} into an angle α and a distance \tilde{r} based on the upper right corner of the unit square

A proof of Theorem 6 is given in “Appendix A.2”. In particular, Eq. (8) will form the basis for numerically determining bivariate quantiles based on linear programming (LP) as derived and discussed later in Sect. 3.1.

We finally define bivariate quantiles along α on the original scale by retransforming with the marginal inverse CDFs.

Definition 7 (*Bivariate CDF-based quantiles along α on the original scale*) For a bivariate real-valued random vector Y , the CDF-based bivariate quantile along α on the original scale is defined as

$$q_{\alpha,\tau} = \left(F_1^{-1}(\tilde{q}_{1,\alpha,\tau}), F_2^{-1}(\tilde{q}_{2,\alpha,\tau}) \right)' \tag{9}$$

where $\tilde{q}_{1,\alpha,\tau} = 1 - \tilde{r}_{\alpha,\tau} \cos(\alpha)$ and $\tilde{q}_{2,\alpha,\tau} = 1 - \tilde{r}_{\alpha,\tau} \sin(\alpha)$. The bivariate quantile curve (1) is then given by $Q_\tau = \{q_{\alpha,\tau}, \alpha \in D_\alpha\}$. Similarly, the empirical bivariate quantile along α on the original scale is

$$q_{n,\alpha,\tau} = (F_1^{-1}(\tilde{q}_{1,n,\alpha,\tau}), F_2^{-1}(\tilde{q}_{2,n,\alpha,\tau}))'. \tag{10}$$

Of course, in practice the true marginal CDFs in (10) will typically be replaced by empirical CDFs, as discussed in the following.

General strategy for obtaining CDF-based bivariate quantiles The following recipe summarises how empirical bivariate quantiles for a fixed quantile level $\tau \in (0, 1)$ and a random sample y_1, \dots, y_n from a continuous bivariate distribution F can be calculated:

1. Transform the observed data to the unit square utilising the univariate empirical CDFs, i.e. determine $\tilde{y}_i = (F_{1,n}(y_{i1}), F_{2,n}(y_{i2}))'$ for $i = 1, \dots, n$.
2. Estimate quantile distances $\tilde{r}_{n,\alpha,\tau}$ for a fine grid of directions represented by a sequence of angles $\alpha \in D_\alpha$ to approximate the quantile curve.

3. Transform the estimators back onto the original scale by applying the inverse univariate empirical quantile functions to determine $q_{n,\alpha,\tau}$ based on (10).

Some properties of the proposed CDF-based bivariate quantiles are summarised in the following theorem (where all properties hold both for the theoretical as well as the empirical versions):

Theorem 8 (Properties of CDF-based bivariate quantiles)

1. *CDF-based quantiles are equivariant under componentwise strictly monotonically increasing transformations $h: \mathbb{R}^2 \rightarrow \mathbb{R}^2, h(\mathbf{y}) = (h_1(y_1), h_2(y_2))'$.*

Proof Let $\mathbf{W} = h(\mathbf{Y}) = (h_1(Y_1), h_2(Y_2))'$ denote the transformed observations, and let G_1, G_2 be the marginal CDFs of $\mathbf{W}_1, \mathbf{W}_2$, respectively. We then have

$$\arg \min_{\tilde{r} \in D_{\tilde{r}}(\alpha)} \mathcal{L}_{\alpha,\tau}(\tilde{r}|\mathbf{Y}) = \arg \min_{\tilde{r} \in D_{\tilde{r}}(\alpha)} \mathcal{L}_{\alpha,\tau}(\tilde{r}|\mathbf{W})$$

where $D_{\tilde{r}}(\alpha)$ denotes the domain of \tilde{r} . This follows from $G_j(w_j) = F_j(h_j^{-1}(w_j)) = F_j(y_j)$ on the one hand and $h_j(F_j^{-1}(y_j)) = F_j^{-1}(h_j(y_j)), j = 1, 2$, on the other hand. The proof works in complete analogy in the empirical case when using the empirical CDFs and in fact equivariance would still hold with arbitrary order-preserving transformation t used to transform to the unit square. □

2. *CDF-based quantiles are in general not equivariant under affine transformations.* While equivariance for componentwise affine transformations (with positive slope parameter) follows from Property 1., general affine transformations alter the orientation of the data/distribution such that the fraction of data points/probability mass to the lower left of a given point in \mathbb{R}^2 is changed. For definitions of bivariate quantiles that provide a centre-outward ordering of the data, equivariance of the bivariate quantile under affine transformations seems a plausible prerequisite since the affine transformation merely rotates and scales the data cloud and therefore any meaningful notion of ‘centrality’ should be preserved. However, when considering the bivariate CDF to define bivariate quantiles and therefore a direction-based approach, the affine transformation alters the coordinate system in such a way that equivariance is no longer achievable (and not desirable as well).
3. *The expected loss has a unique, global minimum.* For the empirical loss, any local minimum is also a global minimum and the minimum is unique up to the inherent nonidentifiability resulting from the discreteness of the data.

Proof For the expected loss, the statement is a direct consequence from the proof of Theorem 6 since $\mathbb{E}(\rho_{\alpha,\tau}(\tilde{\mathbf{y}}, \tilde{\mathbf{r}}))$ is a strictly convex function on the compact set $D_{\tilde{r}}(\alpha)$. For the empirical loss, we have that the individual contributions to the loss function are piecewise linear, convex functions (see Sect. 3.1) such that the complete loss is also convex. As a consequence, any local minimum is also a global minimum. The nonuniqueness arises in a similar way as for univariate quantiles where, if $n\tau$ is a natural number, the empirical quantile is only determined up to an interval formed by two adjacent observations. A similar statement holds for the CDF-based quantiles

based on the transformed data. This also implies that the empirical quantiles will converge to a unique solution as the sample size grows to infinity (see Sect. 4.2 for a more precise result). \square

4. The boundary cases $\alpha = 0$ and $\alpha = \pi/2$ yield the marginal distributions for Y_1 and Y_2 , respectively.

Proof For $\alpha = 0$ we obtain $\tilde{S}_\alpha(\tilde{r}) = \mathbb{P}(\tilde{Y}_1 \leq 1 - \tilde{r}, \tilde{Y}_2 \leq 1)$ which is in fact the marginal CDF of \tilde{Y}_1 and hence we retrieve quantiles of the marginal distribution for Y_1 . The analogue for $\alpha = \pi/2$ results in the marginal distribution of Y_2 . Similarly, when considering the loss function in (6), we find that $\tilde{v}_1 \rightarrow 1 - \tilde{y}_1 - \tilde{r}$ and $\tilde{v}_2 \rightarrow \infty$ for $\alpha \rightarrow 0$ such that the loss function reduces to the loss in case of a univariate quantile for \tilde{y}_1 . On the other hand, we have $\tilde{v}_1 \rightarrow \infty$ and $\tilde{v}_2 \rightarrow 1 - \tilde{y}_2 - \tilde{r}$ for $\alpha \rightarrow \pi/2$ such that the loss for the univariate quantile of \tilde{y}_2 results. \square

5. The breakdown point of the bivariate quantiles is $\min(\lfloor n\tau \rfloor, \lfloor n(1 - \tau) \rfloor)$.

Proof This follows from the construction of the loss function which, after reducing information to the distances to the upper right corner of the unit square, has the same structure as the loss function for univariate quantiles. \square

2.5 Using the origin as the reference point

For our considerations, we have chosen the upper right corner of the unit square as the reference point for the polar coordinate-type characterisation of points on purpose. To motivate this choice, we will now discuss results based on the origin as an alternative reference point. In this case, we obtain the representation

$$\tilde{q} = (\tilde{r} \cos(\alpha), \tilde{r} \sin(\alpha))'$$

and based on the distances $\tilde{R} = \max(\tilde{Y}_1/\cos(\alpha), \tilde{Y}_2/\sin(\alpha))$ we find that $\tilde{F}_\alpha(\tilde{r}) = \mathbb{P}(\tilde{R} \leq \tilde{r})$ coincides with the bivariate CDF at $\tilde{q} = (\tilde{r} \cos(\alpha), \tilde{r} \sin(\alpha))'$, i.e.

$$\tilde{F}_\alpha(\tilde{r}) = \mathbb{P}\left(\tilde{Y}_1 \leq \tilde{r} \cos(\alpha), \tilde{Y}_2 \leq \tilde{r} \sin(\alpha)\right), \quad \tilde{r} \in D_{\tilde{r}}(\alpha),$$

for a given angle $\alpha \in D_\alpha = (0, \pi/2)$ and $\tilde{r} \in D_{\tilde{r}}(\alpha) = [0, \min(1/\cos(\alpha), 1/\sin(\alpha))]$ (and a similar result holds for the empirical versions). One important difference to our standard definition is that $\tilde{F}_\alpha(\tilde{r})$ is in fact not a proper CDF since (for $\alpha < \pi/4$)

$$\lim_{\tilde{r} \rightarrow 1/\cos(\alpha)} \tilde{F}_\alpha(\tilde{r}) = \mathbb{P}\left(\tilde{Y}_1 \leq 1, \tilde{Y}_2 \leq \sin(\alpha)/\cos(\alpha)\right) = \mathbb{P}\left(\tilde{Y}_2 \leq \sin(\alpha)/\cos(\alpha)\right)$$

and therefore $\tilde{F}_\alpha(\tilde{r})$ does not approach 1 as \tilde{r} increases towards its upper limit. This also implies that with the origin as a reference point, not all quantile levels can actually be achieved for a given angle. Similar statements hold for $\alpha > \pi/4$ and the empirical direction-specific CDF.

Nevertheless, we could still continue to proceed as in the previous section if we restrict ourselves to cases where, for a given angle α , the quantile of interest is indeed existing. Then, the loss criterion (8) would have to be replaced by

$$\begin{aligned} \mathcal{L}_{n,\alpha,\tau}(\tilde{r}) &= \frac{1}{n} \sum_{i=1}^n \max \left(\frac{\tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r}, \frac{\tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} \right) \left(\mathbb{1}_{\left\{ \max \left(\frac{\tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r}, \frac{\tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} \right) \geq 0 \right\}} - \tau \right). \end{aligned}$$

In generalisation of Theorem 8, we then obtain the following properties:

Corollary 9 1. *Properties 1–3 of Theorem 8 hold when using the origin as the reference point.*

2. *The boundary cases $\alpha = 0$ and $\alpha = \pi/2$ have to be excluded (and in particular they do not correspond to determining marginal quantiles) with $(0, 0)'$ as reference point.*

Proof For $\alpha = 0$, we obtain $\tilde{F}_\alpha(\tilde{r}) = \mathbb{P}(\tilde{Y}_1 \leq \tilde{r}, \tilde{Y}_2 \leq 0)$ which is zero due to the transformation to the unit square. Similarly, with $\alpha = \pi/2$ we obtain $\tilde{F}_\alpha(\tilde{r}) = \mathbb{P}(\tilde{Y}_1 \leq 0, \tilde{Y}_2 \leq \tilde{r}) \equiv 0$. This is a consequence of the difficulty discussed above, i.e. the fact that the α -specific CDFs are degenerate. □

3 Estimation of direction-specific quantile curves

Due to the convex, piecewise linear structure of the loss function in (6), empirical bivariate quantiles can be determined as the solution of a linear program. Compared to the check function in the univariate case, however, the individual contributions $\mathcal{L}_{i,\alpha,\tau}(\tilde{r})$ do not only depend on the quantile level τ but also on the angle α . While it is also possible to estimate the quantiles by direct inversion of an empirical CDF on the transformed data, casting the estimation problem as a minimisation problem solved via linear programming has two advantages: first, we gain a better understanding of the geometry underlying our definition of bivariate quantiles and, second, it can be expected to prove useful when considering extended bivariate quantile settings, for example with a regression specification on the quantile distances. In the following, we investigate the geometry of the loss function and show how the minimization problem can be cased into a linear program.

3.1 Geometric perspectives on the loss function

Define the two index sets $\mathcal{I}_1, \mathcal{I}_2$ to divide the set of observation indices $\mathcal{I} = \{1, \dots, n\}$ into two disjunct subsets $\mathcal{I} = \mathcal{I}_1 \dot{\cup} \mathcal{I}_2$ as

$$\mathcal{I}_1 = \left\{ i \in \mathcal{I} \mid \frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} \leq \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} \right\}, \quad \mathcal{I}_2 = \mathcal{I} \setminus \mathcal{I}_1 = \left\{ i \in \mathcal{I} \mid \frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} > \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} \right\}.$$

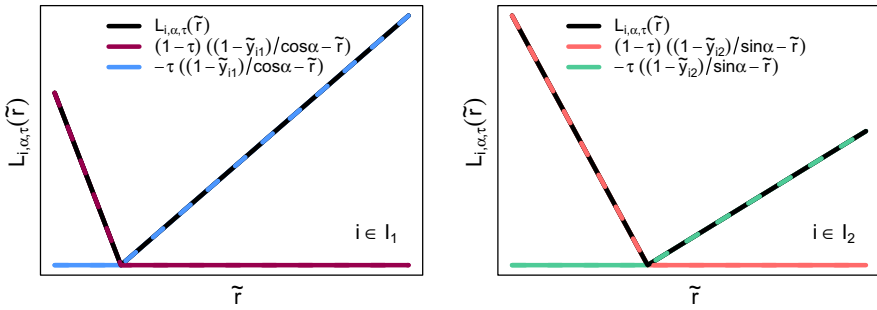


Fig. 4 Contributions to the loss function with rotated origin as reference point. Subfigures visualise the piecewise linear contributions $\mathcal{L}_{i, \alpha, \tau}(\tilde{r})$ for $i \in \mathcal{I}_1$ (left) and $i \in \mathcal{I}_2$ (right)

Then, depending on i , the piecewise linear contributions $\mathcal{L}_{i, \alpha, \tau}(\tilde{r})$ in each subset are given by

$$\begin{aligned}
 i \in \mathcal{I}_1: \mathcal{L}_{i, \alpha, \tau}(\tilde{r}) &= \begin{cases} (1 - \tau) \left(\frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r} < 0 \\ -\tau \left(\frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} - \tilde{r} \geq 0 \end{cases} \\
 i \in \mathcal{I}_2: \mathcal{L}_{i, \alpha, \tau}(\tilde{r}) &= \begin{cases} (1 - \tau) \left(\frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} < 0 \\ -\tau \left(\frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} - \tilde{r} \geq 0 \end{cases}
 \end{aligned}$$

(see Fig. 4 for a graphical illustration).

3.2 Linear programming

To take advantage of the piecewise linear structure of the convex loss function, we formulate a linear program for the estimation of bivariate quantiles as follows. For $i \in \mathcal{I}_1$, we let u_i^+, u_i^- be auxiliary variables with

$$\begin{aligned}
 u_i^+ &= \max(0, (1 - \tilde{y}_{i1}) / \cos(\alpha) - \tilde{r}) \\
 u_i^- &= -\min(0, (1 - \tilde{y}_{i1}) / \cos(\alpha) - \tilde{r}).
 \end{aligned}$$

Similarly, we specify v_i^+, v_i^- for $i \in \mathcal{I}_2$ as

$$\begin{aligned}
 v_i^+ &= \max(0, (1 - \tilde{y}_{i2}) / \sin(\alpha) - \tilde{r}) \\
 v_i^- &= -\min(0, (1 - \tilde{y}_{i2}) / \sin(\alpha) - \tilde{r}).
 \end{aligned}$$

By definition, the $2n$ additional variables $u_i^+, u_i^-, v_i^+, v_i^-$ are functions of the unknown quantile distance \tilde{r} but allow us to rewrite the problem $\arg \min_{\tilde{r} \in D_{\tilde{r}}(\alpha)} \mathcal{L}_{n, \alpha, \tau}(\tilde{r})$ into a linear program. Let therefore $\tilde{y}_\alpha = (\tilde{y}_{1, \alpha}, \dots, \tilde{y}_{n, \alpha})'$ be the vector of transformed observations with

$$\tilde{y}_{i,\alpha} = \frac{1 - \tilde{y}_{i1}}{\cos(\alpha)} \mathbb{1}_{\{i \in \mathcal{I}_1\}} + \frac{1 - \tilde{y}_{i2}}{\sin(\alpha)} \mathbb{1}_{\{i \in \mathcal{I}_2\}}$$

and define an additional set of auxiliary variables $\mathbf{w}_\alpha^+ = (w_{1,\alpha}^+, \dots, w_{n,\alpha}^+)'$ and $\mathbf{w}_\alpha^- = (w_{1,\alpha}^-, \dots, w_{n,\alpha}^-)'$ with

$$w_{i,\alpha}^+ = u_i^+ \mathbb{1}_{\{i \in \mathcal{I}_1\}} + v_i^+ \mathbb{1}_{\{i \in \mathcal{I}_2\}} \quad w_{i,\alpha}^- = u_i^- \mathbb{1}_{\{i \in \mathcal{I}_1\}} + v_i^- \mathbb{1}_{\{i \in \mathcal{I}_2\}}$$

such that the complete set of unknown parameters is given by

$$\mathbf{w}_\alpha = (\tilde{r}, (\mathbf{w}_\alpha^+)', (\mathbf{w}_\alpha^-)') \in D_{\tilde{r}}(\alpha) \times [0, 1]^{2n}.$$

With the constraint matrix $\mathbf{A}_\alpha = (\mathbf{1}_n, \text{diag}(w_{i,\alpha}^+), -\text{diag}(w_{i,\alpha}^-)) \in \mathbb{R}^n \times \mathbb{R}^{2n+1}$ and coefficient vector $\mathbf{c} = (0, (1 - \tau)\mathbf{1}'_n, \tau\mathbf{1}'_n)' \in \mathbb{R}^{2n+1}$, this finally yields the linear program representation

$$\min_{\tilde{r}, \mathbf{w}_\alpha^+, \mathbf{w}_\alpha^-} \{ (1 - \tau)\mathbf{1}'_n \mathbf{w}_\alpha^+ + \tau\mathbf{1}'_n \mathbf{w}_\alpha^- | \tilde{\mathbf{y}}_\alpha = \tilde{r}\mathbf{1}_n + \mathbf{w}_\alpha^+ - \mathbf{w}_\alpha^- \} = \min_{\mathbf{w}_\alpha} \{ \mathbf{c}' \mathbf{w}_\alpha | \tilde{\mathbf{y}}_\alpha = \mathbf{A}_\alpha \mathbf{w}_\alpha \}$$

of our optimisation problem. The optimal quantile distance $\tilde{r}_{n,\alpha,\tau}$ is thus simply the first entry of \mathbf{w}_α which can be used to compute an estimate $\mathbf{q}_{n,\alpha,\tau}$ for the direction-specific bivariate quantile along α on the original scale using Eq. (10).

3.3 Implementation

An implementation of our novel directional bivariate quantiles is provided in the R-package `bivquant` (Klein 2019). Results in the package are based on solutions of linear programming systems available in the R-package `lpSolve` (Berkelaar et al. 2015). We also provide an implementation for estimating the geometric quantiles of Chakraborty (2001) that we will use for comparison purposes in our simulations. Estimates for depth-based bivariate quantiles will be derived using the R-package `ddalpha` (Pokotylo et al. 2015).

4 Asymptotic properties

In this section, we investigate the asymptotic properties of the proposed empirical bivariate quantiles. We start by establishing some preliminary results before proving consistency of the bivariate empirical quantile $\mathbf{q}_{n,\alpha,\tau}$ and asymptotic normality of the quantile distance $\tilde{r}_{n,\alpha,\tau}$. We will always assume that the upper right corner of the unit square is used as the reference point such that \tilde{S}_α (the survivor function along α from Definition 4) is a proper survivor function in the usual sense.

4.1 Preliminary results for direction-specific quantiles

Lemma 10 (Properties of $\tilde{S}_{n,\alpha}(\tilde{r})$) *Let $(\tilde{R}_1, \dots, \tilde{R}_n)$ with $\tilde{R}_i = \min\left(\frac{1-\tilde{Y}_{i1}}{\cos(\alpha)}, \frac{1-\tilde{Y}_{i2}}{\sin(\alpha)}\right)$ be the sample of distances with survivor function \tilde{S}_α and assume that the general assumptions from Sect. 2.1 are fulfilled. Then,*

1. $\tilde{R}_1, \dots, \tilde{R}_n$ are i.i.d.
2. For all $\tilde{r} \in D_\alpha(\tilde{r})$ the empirical survivor function $\tilde{S}_{n,\alpha}(\tilde{r})$ converges almost surely to $\tilde{S}_\alpha(\tilde{r})$, i.e.

$$\tilde{S}_{n,\alpha}(\tilde{r}) \xrightarrow{a.s.} \tilde{S}_\alpha(\tilde{r}).$$

3. For all $\tilde{r} \in D_\alpha(\tilde{r})$, the empirical survivor function $\tilde{S}_{n,\alpha}(\tilde{r})$ is asymptotically normal, i.e.

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}) - \tilde{S}_\alpha(\tilde{r})) \xrightarrow{d} N\left(0, \tilde{S}_\alpha(\tilde{r})(1 - \tilde{S}_\alpha(\tilde{r}))\right).$$

4. The empirical survivor function $\tilde{S}_{n,\alpha}$ converges uniformly to \tilde{S}_α , i.e.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \tilde{S}_{n,\alpha}(\tilde{r}) = \tilde{S}_\alpha(\tilde{r}), \tilde{r} \in D_{\tilde{r}}(\alpha)\right) = 1.$$

The proof of the lemma can be found in ‘‘Appendix A.3’’.

While Lemma 10 provides us with nice theoretical results on the distances and their survivor function, the results crucially rely on the assumption that the transformation to the unit square was conducted based on the true marginal CDFs while in practice one will rely on the marginal empirical CDFs. The latter have the consequence that the distances $\tilde{R}_1, \dots, \tilde{R}_n$ are no longer i.i.d., which forms the basis for all further results of Lemma 10. This is similar to the consideration of inference for copulas which also involve a transformation based on the marginal CDFs.

Genest and Segers (2010) studied the asymptotic properties of the empirical copula process and the impact of using either the true or the empirical marginal CDFs. Interestingly, they found that the asymptotic variance of the empirical copula process is uniformly smaller when relying on the empirical marginal CDFs. Their results also allow us to derive the asymptotic variance of $\tilde{S}_{n,\alpha}(\tilde{r})$ when using the empirical CDFs for transforming to the unit square, as we detail in the following remark.

Remark 1 We start by considering convergence of \tilde{S}_n (obtained by transforming with the true marginal CDFs) as a function on $[0, 1]^2$ indexed by both the angle α and the distance \tilde{r} . Note that

$$\tilde{S}_{n,\alpha}(\tilde{r}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\tilde{Y}_{i1} \leq 1 - \tilde{r} \cos(\alpha), \tilde{Y}_{i2} \leq 1 - \tilde{r} \sin(\alpha)) = C_n(u_1, u_2)$$

where C_n denotes the empirical copula obtained by transforming with the known marginals and $u_1 = 1 - \tilde{r} \cos(\alpha)$, $u_2 = 1 - \tilde{r} \sin(\alpha)$. This equivalence can, of course,

also be established for the true distance survivor function such that $\tilde{S}_\alpha(\tilde{r}) = C(u_1, u_2)$ where C is the true copula of the data generating process.

From Genest and Segers (2010), it follows that the normalised survivor function $\tilde{S}_n = \sqrt{n}(\hat{S}_n - \tilde{S})$ converges weakly to a zero-mean Gaussian process \tilde{S} with covariance function

$$\begin{aligned} \text{Cov}(\tilde{S}_{\alpha_1}(\tilde{r}_1), \tilde{S}_{\alpha_2}(\tilde{r}_2)) &= \mathbb{P}(\tilde{Y}_1 \leq \min(u_1, v_1), \tilde{Y}_2 \leq \min(u_2, v_2)) \\ &\quad - \mathbb{P}(\tilde{Y}_1 \leq u_1, \tilde{Y}_2 \leq u_2)\mathbb{P}(\tilde{Y}_1 \leq v_1, \tilde{Y}_2 \leq v_2) \end{aligned}$$

where $u_1 = 1 - \tilde{r}_1 \cos(\alpha_1)$, $u_2 = 1 - \tilde{r}_1 \sin(\alpha_1)$, $v_1 = 1 - \tilde{r}_2 \cos(\alpha_2)$, $v_2 = 1 - \tilde{r}_2 \sin(\alpha_2)$. Note that our result on the asymptotic variance of $\tilde{S}_{n,\alpha}(\tilde{r})$ appears as a special case when $\alpha_1 = \alpha_2$ and $\tilde{r}_1 = \tilde{r}_2$ and the covariance reduces to the variance. Furthermore, the result can be equivalently expressed in terms of the normalised empirical copula process $\mathbb{C}_n = \sqrt{n}(C_n - C)$ which converges to a zero-mean Gaussian process \mathbb{C} with covariance function

$$\text{Cov}(\mathbb{C}(u_1, u_2), \mathbb{C}(v_1, v_2)) = C(\min(u_1, v_1), \min(u_2, v_2)) - C(u_1, u_2)C(v_1, v_2) \tag{11}$$

Turning to the situation where the transformation is achieved by the empirical CDFs, let \hat{S}_n denote the corresponding estimate of the directional survivor function. From Genest and Segers (2010), we then have that the normalised version $\hat{S}_n = \sqrt{n}(\hat{S}_n - \tilde{S})$ converges weakly to a zero-mean Gaussian process \hat{S} with pointwise evaluations given by

$$\hat{S}_\alpha(\tilde{r}) = \mathbb{C}(u_1, u_2) - \mathbb{C}(u_1, 1) \frac{\partial}{\partial u_1} C(u_1, u_2) - \mathbb{C}(1, u_2) \frac{\partial}{\partial u_2} C(u_1, u_2).$$

Note that this is a linear combination of three evaluations of the Gaussian process $\mathbb{C}(u_1, u_2)$ such that the asymptotic variance of $\hat{S}_\alpha(\tilde{r})$ can easily be calculated (for given true copula C) by constructing the corresponding trivariate zero-mean normal distribution where the elements of the covariance matrix are obtained from (11).

Lemma 11 (Almost sure convergence of the quantile distance) *Under the same assumptions as in Lemma 10, the empirical τ th quantile distance along α converges almost surely to the true τ th quantile distance along α for $\tau \in (0, 1)$ and $\alpha \in D_\alpha$:*

$$\tilde{r}_{n,\alpha,\tau} \xrightarrow{a.s.} \tilde{r}_{\alpha,\tau}.$$

Consequently,

$$\tilde{q}_{n,\alpha,\tau} \xrightarrow{a.s.} \tilde{q}_{\alpha,\tau}.$$

The lemma is proved in ‘‘Appendix A.4’’.

Remark 2 Similar as with Lemma 10, Lemma 11 relies on the fact that the transformation to the unit square was accomplished based on the true marginal CDFs. Following

the considerations in Genest et al. (1995) on the convergence of the maximum likelihood estimate for copula parameters obtained from the empirical copula, one can expect that a similar generalisation will be possible also for directional bivariate quantiles. However, the results of Genest et al. (1995) are not directly applicable here since our optimisation criterion is not continuously differentiable with respect to the quantile distance.

4.2 Consistency of bivariate quantiles on the original scale

As a direct consequence from Lemma 11 and Slutsky’s theorem, consistency of bivariate quantiles on the original scale is obtained as summarised in the following theorem.

Theorem 12 (Almost sure convergence of the bivariate quantile on the original scale) *Under the general assumptions from Sect. 2.1 and Lemma 10, the empirical τ th quantile along α converges almost surely to the τ th quantile along α for $\tau \in (0, 1)$*

$$q_{n,\alpha,\tau} \xrightarrow{a.s.} q_{\alpha,\tau}.$$

4.3 Asymptotic normality of the quantile distance

The results from Lemma 10 allow us to show the asymptotic normality of the quantile distance:

Theorem 13 (Asymptotic normality of the quantile distance) *In addition to the usual assumptions, assume that $\tilde{S}_\alpha(\tilde{r})$ is continuously differentiable and let $\tilde{f}_\alpha(\tilde{r}) = -\frac{\partial}{\partial \tilde{r}} \tilde{S}_\alpha(\tilde{r})$ be the density corresponding to $\tilde{S}_\alpha(\tilde{r})$. For $\tau \in (0, 1)$ and $\alpha \in D_\alpha$ the empirical τ th quantile distance along α ($\tilde{r}_{n,\alpha,\tau} = \tilde{S}_{n,\alpha}^{-1}(\tau)$) is asymptotically normally distributed with*

$$\sqrt{n} \left(\tilde{S}_{n,\alpha}^{-1}(\tau) - \tilde{S}_\alpha^{-1}(\tau) \right) \xrightarrow{d} N \left(0, \frac{\tau(1-\tau)}{(\tilde{f}_\alpha(\tilde{r}_{\alpha,\tau}))^2} \right).$$

A detailed proof of Theorem 13 can be found in “Appendix A.5”.

Remark 3 Note that by the assumptions on $f(y_1, y_2)$ in Sect. 2.1, $\tilde{f}_\alpha(\tilde{r})$ exists and is strictly positive on the unit square. From Definition 5, it follows directly that $\tilde{q}_{n,\alpha,\tau,1}$ and $\tilde{q}_{n,\alpha,\tau,2}$ are also individually asymptotically normally distributed. However, this does not lead to a joint asymptotic normality result for $\tilde{q}_{n,\alpha,\tau}$ since both components of $\tilde{q}_{n,\alpha,\tau}$ are determined from $\tilde{r}_{n,\alpha,\tau}$ by transformation.

Remark 4 Similar as with Lemma 10, also Theorem 13 can be extended to the case of transforming with the empirical CDFs rather than the true CDFs by relying the asymptotic variance of $\hat{\tilde{S}}_\alpha(\tilde{r})$ derived in Remark 1. In the proof of Theorem 13, we need Lemma 10.2, which in case of having transformed with the empirical CDFs has to be replaced by its weak consistency analogue. The latter follows from the results of Genest and Segers (2010).

Table 1 Data generating processes for the simulation

Name	Data generating process
Gaussian	$y \sim N(\mu, \Sigma), \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$
Linear shift	$y \sim N(\tilde{\mu}, \Sigma), \tilde{\mu} = \begin{pmatrix} -12 \\ 50 \end{pmatrix}$
Scaling	$y \sim N(\mu, \tilde{\Sigma}), \tilde{\Sigma} = \begin{pmatrix} -2 & 0 \\ 0 & 20 \end{pmatrix} \Sigma \begin{pmatrix} -2 & 0 \\ 0 & 20 \end{pmatrix}$
Rotation	$y \sim N(\mu, \tilde{\Sigma}), \tilde{\Sigma} = \begin{pmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{pmatrix}' \Sigma \begin{pmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{pmatrix}$
Shearing	$y \sim N(\mu, \tilde{\Sigma}), \tilde{\Sigma} = \begin{pmatrix} 1 & -0.6 \\ 0 & 1 \end{pmatrix}' \Sigma \begin{pmatrix} 1 & -0.6 \\ 0 & 1 \end{pmatrix}$
Lower tail	Clayton copula, $\rho = 2, y_j \sim N(0, 1), j = 1, 2$
Upper tail	Gumbel copula, $\rho = 2, y_j \sim N(0, 1), j = 1, 2$
Asymmetry	Clayton copula, $\rho = 2, y_1 \sim N(0, 1), y_2 \sim \text{Ga}(3, 1.5)$
Single outlier	$y_1 \sim N(\tilde{\mu}, \tilde{\Sigma}), \tilde{\mu} = \begin{pmatrix} -3 \\ 3 \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}, y_i \sim N(\mu, \Sigma), i = 2, \dots, n$
Outlier	$y_k \sim N(\tilde{\mu}, \tilde{\Sigma}), \tilde{\mu} = \begin{pmatrix} -5.5 \\ 5.5 \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}, k = 1, \dots, 5,$ $y_i \sim N(\mu, \Sigma), i = 6, \dots, n$

5 Empirical evaluation

5.1 Simulations

We supplement the discussion of theoretical properties of CDF-based quantiles with simulation-based empirical evidence. We compare the performance of our bivariate quantiles with some of the previously suggested approaches and in particular the geometric quantiles of Chakraborty (2001) and several depth-based proposals. The main difficulties for such a comparison are on the one hand the lack of implementations for most of the competing approaches (and in particular for those that are rather based on abstract, theoretical concepts) and on the other hand the very different population quantities they are targeting at. It therefore does not make sense to try to relate, for example, geometric quantiles and CDF-based quantiles directly to each other. Hence, we rather construct exemplary data generating mechanisms that allow us to study specific properties of the approaches separately such as deviations from elliptical contour lines, sensitivity with respect to outliers.

Simulation set-up More precisely, we consider the following data generating processes (see Table 1 for details):

- *Gaussian*: Data generated from a bivariate normal distribution with standard normal marginals.
- *Linear shift*: A linear shift applied to *Gaussian* to modify the expectation vector.
- *Scaling*: Separate scaling factors are applied to the components of *Gaussian*.

- *Rotation*: Applying an affine transformation that implies a rotation of *Gaussian*.
- *Shearing*: Applying an affine transformation that implies shearing of *Gaussian*.
- *Lower tail*: Data generated from a Clayton copula (implying lower tail dependence) with standard normal marginals.
- *Upper tail*: Data generated from a Gumbel copula (implying upper tail dependence) with standard normal marginals.
- *Asymmetry*: One standard normal of *Lower tail* is replaced by a gamma marginal.
- *Single outlier*: One observation of *Gaussian* replaced by an outlier.
- *Outlier*: Five observations of *Gaussian* replaced by outliers.

The first five data generating processes all imply elliptical shapes of the true data generating density. The next three data generating processes result in asymmetric distributions with symmetric lower tail dependence (*Lower tail*), symmetric upper tail dependence (*Upper tail*) and asymmetric lower tail dependence (*Asymmetry*). The final two data generating processes introduce one single and five more extreme outliers, respectively.

We compare two different sample sizes ($n = 200$ and $n = 1000$) and generate $R = 6$ replicates for each of the data generating processes and each sample size to avoid conclusions being drawn from one single, potentially misleading data set. All scenarios are then estimated with the following approaches:

1. *CDF-based quantiles* based on a sequence of 32 equidistant values for the angle $\alpha \in [0, 2\pi)$ and for quantile levels $\tau = \{0.1, 0.2, \dots, 0.8, 0.9\}$.
2. *Geometric quantiles* for a grid of 128 (i.e. four times the number of angles α) directions $\mathbf{u} \in B^2$ in the unit open ball with directions \mathbf{u} chosen based on $\|\mathbf{u}\| = 2\tau - 1$.
3. *Depth-based quantiles* based on the *halfspace depth* (Tukey 1975), the *simplicial depth* (Liu 1990), the *simplicial volume depth*, the *spatial depth* (Koltchinskii 1997) and the *zonoid depth* (Koshevoy and Mosler 1997; Mosler 2002). For the determination of depth-based quantiles, we rely on the R-package *ddalpha* (Pokotylo et al. 2015) and compute the depths on a 100×100 equidistant grid within the range of observed values \mathbf{y} . Since results obtained with the functions `depth.simplicial/depth.simplicialVolume` turned out to be very wiggly with the default settings of randomly choosing 5% of the simplices, we changed the percentage to 99%. The exact approach was not computationally feasible, in particular for the larger sample size where estimates for only one single data set could not be completed within 3 days of computing time.

Summarised results Since the first five data generating processes are all affine transformations of normally distributed data, we did not find any structural differences between the results such that we only present results for the scaled scenario. A similar statement holds for the nonelliptical data generating processes where we focus on the asymmetric case in the following. All methods turned out to be rather robust against one single outlier, such that we focus on the second, more extreme scenario with five outliers.

Concerning the depth-based quantiles, we restrict the presentation to *halfspace* and *simplicial* depth since results with the *simplicial volume* depth and the *spatial* depth

are resembling those of geometric quantiles, i.e. they turn out to be reliable only for an elliptical structure of estimated curves. The results obtained with the *zonoid* depth are comparable to those from the *halfspace* depth.

Finally, among the six replicates, no eye-catching structural differences could be identified such that we only present results from replication $r = 1$. Figure 5 shows estimates for the three chosen scenarios and four approaches. Note that in the upper three rows the colours indicate the centrality of the points while for the fourth row they differentiate the regions between the quantile curves of level τ .

From our simulation results, we can draw the following conclusions:

Scaling Elliptical structures can well be captured with all approaches and for both small and large sample sizes.

Asymmetry Estimates of geometric quantiles remain mostly elliptical and are not even able to clearly identify lower or upper tail dependence for the large sample size ($n = 1000$). Quantiles based on the *halfspace* depth somewhat deviate from elliptical shapes but still do not cover the strong lower tail dependence, especially for smaller sample sizes. In contrast, the quantiles based on the *simplicial* depth reflect the shape of the true data generating density quite well, in particular for the large sample size. CDF-based quantiles identify the true quantile levels fairly well already for small sample sizes and are rather close to the true values for large samples. In comparison with elliptical data, the lower tail dependence results in more peaked quantiles (both empirically and theoretically) for this method while the density reveals an asymmetric structure in its contour lines.

Outlier The CDF-based quantiles are very robust against outliers even for small samples. Robustness can also be found for depth-based and geometric quantiles for the case of large samples while for small samples both approaches react more sensitively. In particular, extreme quantiles obtained from the *simplicial* depth are strongly dominated by the outliers.

5.2 Childhood undernutrition scores in India

To illustrate one potential area of application for bivariate quantiles, we consider a data set on childhood malnutrition in India obtained from the 1998/1999 demographic and health survey (DHS, www.dhsprogram.com). The data set contains information on the nutritional status of $n = 24,316$ children assessed via different Z -scores comparing the nutritional status of children in the population of interest with the nutritional status in a reference population. More precisely, the Z -score is defined as

$$Z = \frac{AC - \mu_{AC}}{\sigma_{AC}}$$

where AC denotes an anthropometric characteristic for the child, while μ_{AC} and σ_{AC} correspond to median and standard deviation in the reference population. We consider two indicators jointly in the following: insufficient weight for height capturing acute undernutrition (Z_1 , *wasting*) and insufficient weight for age reflecting both, chronic and acute undernutrition (Z_2 , *underweight*).

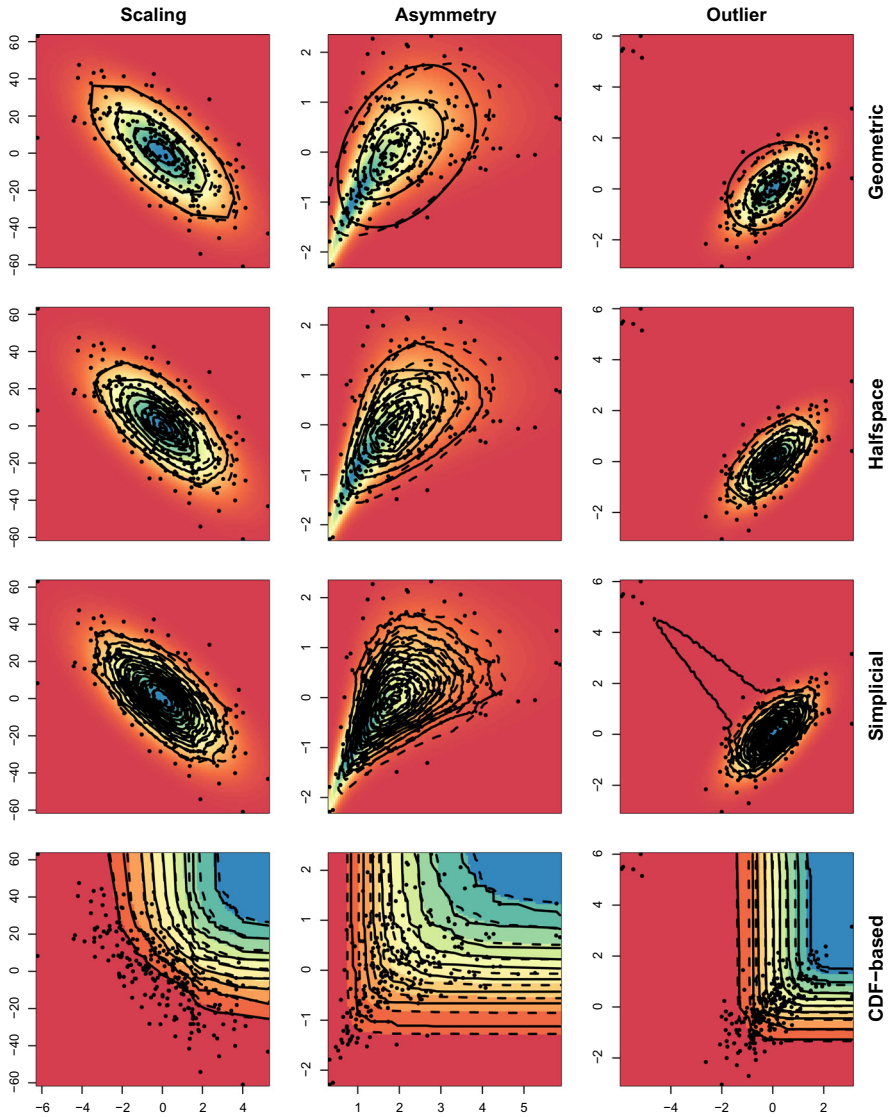


Fig. 5 Simulation results. The first three rows show estimated quantiles for geometric quantiles as well as quantiles based on the *halfspace* and *simplicial* depth. The colour scheme represents the true underlying densities. The fourth row shows estimated quantile curves of CDF-based quantiles where the colours now represent the theoretical quantile levels. The solid (dashed) lines are obtained from samples of size $n = 200$ ($n = 1000$) for the first replication; the dots represent the sampled data points for $n = 200$. Columnwise, the data generating processes *Scaling*, *Asymmetry* and *Outlier* of Table 1 are shown. Note that in the upper three rows, the colours indicate the centrality of the points while for the fourth row they belong to the region starting at the τ -quantile curves (color figure online)

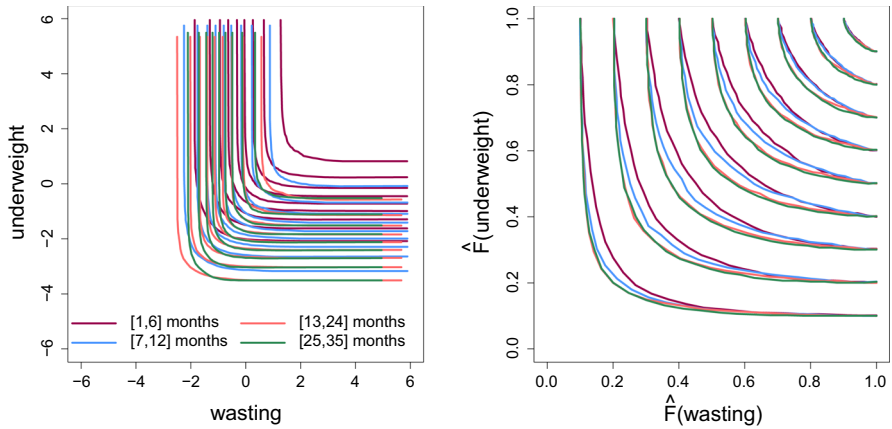


Fig. 6 Childhood Undernutrition. CDF-based quantiles for quantile levels $\tau \in \{0.1, 0.2, \dots, 0.8, 0.9\}$ on the original scale (left) and on the unit square (right) (color figure online)

Since the distribution of the scores is strongly dependent on the ages of children, we divide the whole data set into four age-stratified subsets and estimate CDF-based bivariate quantiles for each subset separately. Note that still the four resulting data sets comprise a couple of thousand observations, each. Due to the linear programming formulation, CDF-based bivariate quantiles are computationally feasible also for such large data sets.

Figure 6 depicts the resulting estimated bivariate quantiles at quantile levels $\tau \in \{0.1, 0.2, \dots, 0.8, 0.9\}$ and for a dense grid of α values since this is the major interest in real data analysis and not for fixed α . In addition to the quantiles on the scale of the original observations (left), this figure also shows the quantiles for the data transformed to the unit square to emphasise changes in the dependence structure with respect to the age of the child (right). For better visibility, Fig. 7 depicts separate graphs for quantile curves on the original scale. One observation that can be made from the bivariate quantiles is that the variability of wasting is reducing considerably with increasing age. Concerning the dependence between the two measures, it is helpful to compare the shapes we estimate with those from the exemplary data generating processes from the previous simulation (Sect. 5.1). For children of a very young age, the contours are close to those of elliptical contour lines while the ones for older ages are showing stronger signs of lower tail dependence (compare the discussion in Sect. 2.2 on the shape of quantile curves for different types of stylised dependence patterns). This is in line with previous findings presented in Klein and Kneib (2016) based on parametric copula regression and can also be seen from the stronger kink of the quantile contours close to the bisecting line (in particular for the lower quantile levels). The change in the dependence structure is more clearly detectable when considering the data transformed to the unit square where all quantile curves show a trend in their shift towards the origin for increasing age of the child.

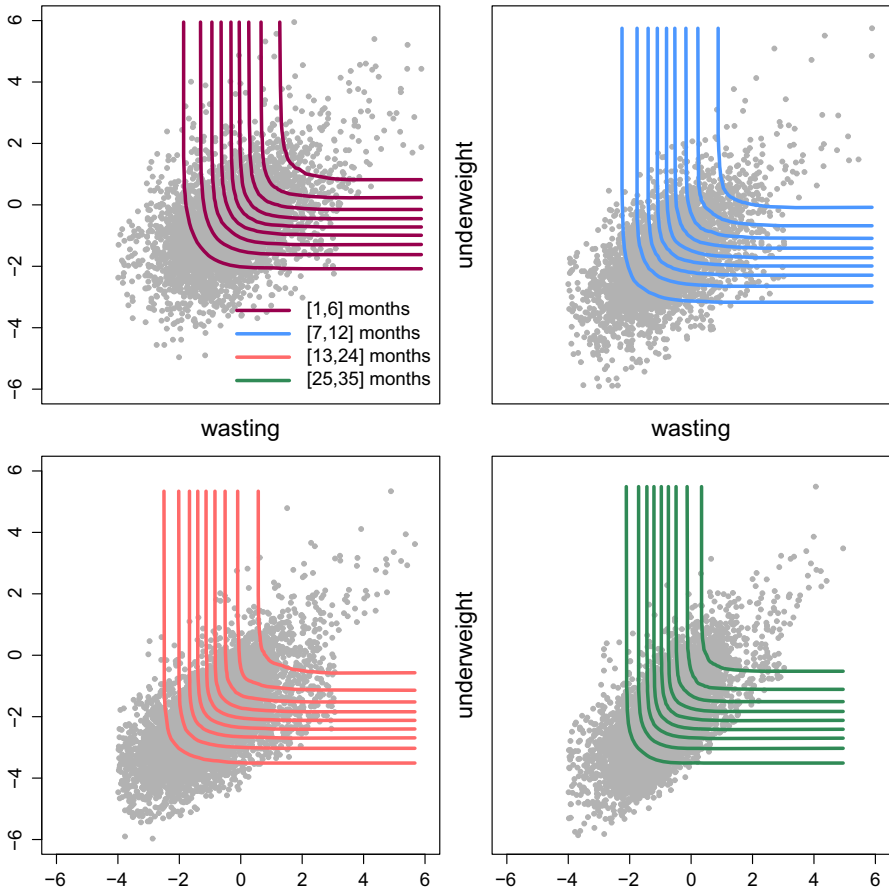


Fig. 7 Childhood Undernutrition. CDF-based quantiles for quantile levels $\tau \in \{0.1, 0.2, \dots, 0.8, 0.9\}$ on the original scale. Graphs show results of Fig. 6 (left) but depict the quantile curves in each subset separately for visibility reasons. Here, grey points are the observed pairs of wasting (x axis) and underweight (y axis) in each subset (color figure online)

6 Summary and conclusions

We proposed a novel notion of bivariate quantiles based on inverting the bivariate CDF. More precisely, we transformed observed bivariate data clouds to the unit square via the empirical CDF and introduced direction-specific quantiles based on the upper right corner of the unit square as the reference point. This construction enables the determination of bivariate CDF-based quantiles via linear programming and yields desirable properties such as invariance under monotonically increasing transformations and robustness. The resulting quantiles do not provide a centre-outward ordering as most of the previous definitions of multivariate quantiles but rather measure extremeness of observations in terms of the bivariate CDF such that the quantile level exhibits a clear interpretation. As a major advantage, CDF-based quantiles can be applied without any

prior assumptions on the shape of the underlying distribution while several competing approaches only work well for distributions that are close to an elliptical shape.

The linear program formulation of bivariate CDF-based quantiles does not only make the determination of bivariate quantiles computationally feasible even for large data sets but will also be useful for developing bivariate quantile regression models where the quantile distance will be specified depending on covariates. As a second area of future research, we will consider generalisations beyond bivariate data. While the transformation to the unit cube would still work, an immediate generalisation would require the specification of $D - 1$ angles in the D -dimensional case. It will therefore be important to study alternative ways of parametrising the D -dimensional unit cube.

A Further proofs

A.1 Proof of Theorem 2

Proof For fixed $b \in \mathbb{R}$, we define the expected loss as a function of $v \in \mathbb{R}$ by

$$\mathbb{E}(\rho_{b,\tau}(\mathbf{Y}, \mathbf{q})) = \mathbb{E}(\rho_{\tau}(\mathbf{Y}, (v, v + b)')). \tag{12}$$

Clearly, the bivariate quantile curves from (1) can be obtained as

$$\mathcal{Q}_{\tau} = \bigcup_{b \in \mathbb{R}} \{v \in \mathbb{R} | F(v, v + b) = \tau\}$$

which intuitively means that we describe \mathbb{R}^2 by straight lines with slope one and intercepts b . With the definition

$$u(\mathbf{y}, v) = \max(y_1 - v, y_2 - v - b),$$

the expected loss for fixed $b \in \mathbb{R}$ is given by

$$\begin{aligned} \mathbb{E}(\rho_{b,\tau}(\mathbf{Y}, v)) &= (\tau - 1) \int_{-\infty}^v \int_{-\infty}^{v+b} u(\mathbf{y}, v) f(y_1, y_2) dy_2 dy_1 \\ &+ \tau \int_v^{\infty} \int_{v+b}^{\infty} u(\mathbf{y}, v) f(y_1, y_2) dy_2 dy_1 \\ &+ \tau \int_v^{\infty} \int_{-\infty}^{v+b} u(\mathbf{y}, v) f(y_1, y_2) dy_2 dy_1 \\ &+ \tau \int_{-\infty}^v \int_{v+b}^{\infty} u(\mathbf{y}, v) f(y_1, y_2) dy_2 dy_1 \end{aligned}$$

$$\begin{aligned}
 &= (\tau - 1) \int_{-\infty}^v \int_{-\infty}^{y_1+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + (\tau - 1) \int_{-\infty}^v \int_{y_1+b}^{v+b} (y_2 - v - b) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_v^{\infty} \int_{v+b}^{y_1+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_v^{\infty} \int_{y_1+b}^{\infty} (y_2 - v - b) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_v^{\infty} \int_{-\infty}^{v+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_{-\infty}^v \int_{v+b}^{\infty} (y_2 - v - b) f(y_1, y_2) dy_2 dy_1.
 \end{aligned}$$

Our strategy is now to show that for all $b \in \mathbb{R}$ the expected loss $\mathbb{E}(\rho_{b,\tau}(\mathbf{Y}, v))$ is uniquely minimised at $q \in \mathbb{R}$ and fulfils the condition $\mathbb{P}(Y_1 \leq q, Y_2 \leq q + b) = \tau$. We therefore investigate the first derivative of $\mathbb{E}(\rho_{b,\tau}(\mathbf{Y}, v))$ with respect to v . The derivative is obtained by applying the Leibniz rule for parameter integrals twice.

$$\begin{aligned}
 \frac{\partial}{\partial v} \mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v)) &= (\tau - 1) \int_{-\infty}^v \frac{\partial}{\partial v} \int_{-\infty}^{y_1+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + (\tau - 1) \int_{-\infty}^v \frac{\partial}{\partial v} \int_{y_1+b}^{v+b} (y_2 - v - b) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_v^{\infty} \frac{\partial}{\partial v} \int_{v+b}^{y_1+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
 &\quad + \tau \int_v^{\infty} \frac{\partial}{\partial v} \int_{y_1+b}^{\infty} (y_2 - v - b) f(y_1, y_2) dy_2 dy_1 \\
 &\quad - \tau \int_{v+b}^{\infty} (y_2 - v - b) f(v, y_2) dy_2
 \end{aligned}$$

$$\begin{aligned}
& + \tau \int_v^\infty \frac{\partial}{\partial v} \int_{-\infty}^{v+b} (y_1 - v) f(y_1, y_2) dy_2 dy_1 \\
& + \tau \int_{-\infty}^v \frac{\partial}{\partial v} \int_{v+b}^\infty (y_2 - v - b) f(y_1, y_2) dy_2 dy_1 \\
& + \tau \int_{v+b}^\infty (y_2 - v - b) f(v, y_2) dy_2 \\
= & -(\tau - 1) \int_{-\infty}^v \int_{-\infty}^{y_1+b} f(y_1, y_2) dy_2 dy_1 \\
& - (\tau - 1) \int_{-\infty}^v \int_{y_1+b}^{v+b} f(y_1, y_2) dy_2 dy_1 \\
& - \tau \int_v^\infty \int_{v+b}^{y_1+b} f(y_1, y_2) dy_2 dy_1 - \tau \int_v^\infty (y_1 - v) f(y_1, v + b) dy_1 \\
& - \tau \int_v^\infty \int_{y_1+b}^\infty f(y_1, y_2) dy_2 dy_1 - \tau \int_{v+b}^\infty (y_2 - v - b) f(v, y_2) dy_2 \\
& - \tau \int_v^\infty \int_{-\infty}^{v+b} f(y_1, y_2) dy_2 dy_1 + \tau \int_v^\infty (y_1 - v) f(y_1, v + b) dy_1 \\
& - \tau \int_{-\infty}^v \int_{v+b}^\infty f(y_1, y_2) dy_2 dy_1 + \tau \int_{v+b}^\infty (y_2 - v - b) f(v, y_2) dy_2 \\
= & -(\tau - 1) \int_{-\infty}^v \int_{-\infty}^{v+b} f(y_1, y_2) dy_2 dy_1 - \tau \int_v^\infty \int_{v+b}^\infty f(y_1, y_2) dy_2 dy_1 \\
& - \tau \int_v^\infty \int_{-\infty}^{v+b} f(y_1, y_2) dy_2 dy_1 - \tau \int_{-\infty}^v \int_{v+b}^\infty f(y_1, y_2) dy_2 dy_1 \\
= & \int_{-\infty}^v \int_{-\infty}^{v+b} f(y_1, y_2) dy_2 dy_1 - \tau.
\end{aligned}$$

In summary, we have

$$\frac{\partial}{\partial v} \mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v)) = \int_{-\infty}^v \int_{-\infty}^{v+b} f(y_1, y_2) dy_2 dy_1 - \tau. \tag{13}$$

Let us first assume that $\mathbb{P}(Y_1 \leq q, Y_2 \leq q + b) = \tau$ holds. It then follows from Eq. (13) that

$$\left. \frac{\partial}{\partial v} \mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v)) \right|_{v=q} = \tau - \tau = 0.$$

In addition,

$$\left. \frac{\partial^2}{\partial v^2} \mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v)) \right|_{v=q} = \int_{-\infty}^q f(y_1, q + b) dy_1 + \int_{-\infty}^{q+b} f(q, y_2) dy_2 > 0$$

holds since we assumed $f(y_1, y_2) > 0$. Consequently, $(q, q + b)'$ is a minimiser of $\mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v))$ and in particular of $\mathbb{E}(\rho_{\tau}(\mathbf{y}, \mathbf{q}))$.

Reversely, if $(q, q + b)'$ is a minimiser of $\mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v))$, a zero first derivative

$$\left. \frac{\partial}{\partial v} \mathbb{E}(\rho_{b,\tau}(\mathbf{y}, v)) \right|_{v=q} = \int_{-\infty}^q \int_{-\infty}^{q+b} f(y_1, y_2) dy_2 dy_1 - \tau = 0$$

is required which is equivalent to

$$\mathbb{P}(Y_1 \leq q, Y_2 \leq q + b) = \tau.$$

□

A.2 Proof of Theorem 6

Proof Recall first that $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \tilde{Y}_2)' = (F_1(Y_1), F_2(Y_2))'$, and let $\tilde{f}(\tilde{y}_1, \tilde{y}_2) > 0$ be the density of $\tilde{\mathbf{Y}}$. From Sect. 3.1, we furthermore have that $\rho_{\alpha,\tau}(\tilde{\mathbf{y}}, \tilde{r})$ can be decomposed into

$$\begin{cases} (1 - \tau) \left(\frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r} \leq \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} < 0 \\ (1 - \tau) \left(\frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} \right) & \text{if } \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} < \frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r} < 0 \\ -\tau \left(\frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r} \right) & \text{if } \min \left(\frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r}, \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} \right) \geq 0 \text{ and } \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} \geq \frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r} \\ -\tau \left(\frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} \right) & \text{if } \min \left(\frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r}, \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} \right) \geq 0 \text{ and } \frac{1 - \tilde{y}_2}{\sin(\alpha)} - \tilde{r} < \frac{1 - \tilde{y}_1}{\cos(\alpha)} - \tilde{r}. \end{cases}$$

Accordingly, the expected loss is

$$\begin{aligned} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{y}, \tilde{r})) &= (1 - \tau) \int_0^{1-\tilde{r} \cos(\alpha)} \int_0^{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}} \frac{1 - \tilde{r} \cos(\alpha) - \tilde{y}_1}{\cos(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ (1 - \tau) \int_0^{1-\tilde{r} \cos(\alpha)} \int_{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}}^{1-\tilde{r} \sin(\alpha)} \frac{1 - \tilde{r} \sin(\alpha) - \tilde{y}_2}{\sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_0^{1-\tilde{r} \cos(\alpha)} \int_{1-\tilde{r} \sin(\alpha)}^1 \frac{\tilde{y}_2 - 1 + \tilde{r} \sin(\alpha)}{\sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}}^1 \frac{\tilde{y}_2 - 1 + \tilde{r} \sin(\alpha)}{\sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_{1-\tilde{r} \sin(\alpha)}^{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}} \frac{\tilde{y}_1 - 1 + \tilde{r} \cos(\alpha)}{\cos(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_0^{1-\tilde{r} \sin(\alpha)} \frac{\tilde{y}_1 - 1 + \tilde{r} \cos(\alpha)}{\cos(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \end{aligned}$$

Now, in analogy to the strategy of the proof of Theorem 2, we apply the Leibniz rule for integrals twice, add or subtract terms with identical limits of integration afterwards and after some further basic calculations obtain

$$\begin{aligned} \frac{\partial}{\partial \tilde{r}} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{y}, \tilde{r})) &= (\tau - 1) \int_0^{1-\tilde{r} \cos(\alpha)} \int_0^{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ (\tau - 1) \int_0^{1-\tilde{r} \cos(\alpha)} \int_{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}}^{1-\tilde{r} \sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_0^{1-\tilde{r} \cos(\alpha)} \int_{1-\tilde{r} \sin(\alpha)}^1 \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}}^1 \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_{1-\tilde{r} \sin(\alpha)}^{1-(1-\tilde{y}_1) \frac{\sin(\alpha)}{\cos(\alpha)}} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \\ &+ \tau \int_{1-\tilde{r} \cos(\alpha)}^1 \int_0^{1-\tilde{r} \sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \end{aligned}$$

Adding together the different integrals yields

$$\frac{\partial}{\partial \tilde{r}} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{y}, \tilde{r})) = \tau - \int_0^{1-\tilde{r} \cos(\alpha)} \int_0^{1-\tilde{r} \sin(\alpha)} \tilde{f}(\tilde{y}_1, \tilde{y}_2) d\tilde{y}_2 d\tilde{y}_1 \tag{14}$$

A necessary condition for \tilde{r} giving a minimum of the expected loss is that (14) is zero, i.e. that $\frac{\partial}{\partial \tilde{r}} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{y}, \tilde{r}))|_{\tilde{r}=\tilde{r}_{\alpha, \tau}} = 0$. This implies that the quantile condition

$\mathbb{P}(\tilde{Y}_1 \leq \tilde{q}_1, \tilde{Y}_2 \leq \tilde{q}_2) = \tau$ is fulfilled if and only if the first derivative of the expected loss at \tilde{r} is zero. What remains to show is that $\frac{\partial^2}{\partial \tilde{r}^2} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{\mathbf{y}}, \tilde{\mathbf{q}}))|_{\tilde{r}=\tilde{r}_{\alpha, \tau}} > 0$ holds.

- For $\alpha \in (0, \pi/2)$, this follows from

$$\begin{aligned} \frac{\partial^2}{\partial \tilde{r}^2} \mathbb{E}(\rho_{\alpha, \tau}(\tilde{\mathbf{y}}, \tilde{\mathbf{q}})) &= \int_0^{1-\tilde{r} \cos(\alpha)} \sin(\alpha) \tilde{f}(\tilde{y}_1, 1 - \tilde{r} \sin(\alpha)) d\tilde{y}_1 \\ &\quad + \int_0^{1-\tilde{r} \sin(\alpha)} \cos(\alpha) \tilde{f}(1 - \tilde{r} \cos(\alpha), \tilde{y}_2) d\tilde{y}_2 \quad (15) \end{aligned}$$

since $\tilde{f}(\cdot, \cdot) > 0$ and $\cos(\alpha) > 0, \sin(\alpha) > 0$.

- In case of $\alpha = 0$, we have $\sin(\alpha) = 0, \cos(\alpha) = 1$ such that the second integral in (15) is zero while the first one is $\int_0^1 \tilde{f}(1 - \tilde{r}, \tilde{y}_2) d\tilde{y}_2 = \tilde{f}_1(1 - \tilde{r}) > 0$.
- In case of $\alpha = \pi/2$, we have $\cos(\alpha) = 0, \sin(\alpha) = 1$ such that the first integral in (15) is zero while the second one is $\int_0^1 \tilde{f}(\tilde{y}_1, 1 - \tilde{r}) d\tilde{y}_1 = \tilde{f}_2(1 - \tilde{r}) > 0$. \square

A.3 Proof of Lemma 10

In the following and in order to prove asymptotic results of Sect. 4, we treat the observed data as i.i.d. replications of \mathbf{Y} defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), F)$. Consequently, the transformed data are i.i.d. replicates of $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \tilde{Y}_2)' = (F_1(Y_1), F_2(Y_2))$ supplemented with the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) = ([0, 1]^2, \mathcal{B}([0, 1]^2), \tilde{F})$ and CDF $\tilde{F}(\tilde{y}_1, \tilde{y}_2) = \mathbb{P}(Y_1 \leq F_1^{-1}(\tilde{y}_1), Y_2 \leq F_2^{-1}(\tilde{y}_2))$. In addition, we introduce

$$\tilde{R} = \min \left(\frac{1-\tilde{Y}_1}{\cos(\alpha)}, \frac{1-\tilde{Y}_2}{\sin(\alpha)} \right), \quad \alpha \in D(\alpha)$$

as a random variable on the probability space $(\tilde{\Omega}_\alpha, \tilde{\mathcal{F}}_\alpha, \tilde{\mathbb{P}}_\alpha) = (D_{\tilde{r}}(\alpha), \mathcal{B}(D_{\tilde{r}}(\alpha)), \tilde{F}_\alpha)$.

Proof *On 1.* The claim follows directly due to the i.i.d. property of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$.

On 2. We introduce the random variables $Z_i = \mathbb{1}_{\{(\tilde{Y}_{i1} \leq 1 - \tilde{r} \cos(\alpha), \tilde{Y}_{i2} \leq 1 - \tilde{r} \sin(\alpha))\}}$ which are i.i.d. since $\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \dots$ are assumed to be i.i.d. We then have

$$\begin{aligned} \mathbb{P}(Z_i = 1) &= \mathbb{P}(\tilde{Y}_{i1} \leq 1 - \tilde{r} \cos(\alpha), \tilde{Y}_{i2} \leq 1 - \tilde{r} \sin(\alpha)) = \tilde{S}_\alpha(\tilde{r}) \\ \mathbb{P}(Z_i = 0) &= 1 - \tilde{S}_\alpha(\tilde{r}) \end{aligned}$$

and hence $\mathbb{E}(Z_i) = \tilde{S}_\alpha(\tilde{r})$. With the strong law of large numbers, we immediately find

$$\tilde{S}_{n, \alpha}(\tilde{r}) = \frac{1}{n} \sum_{i=1}^n Z_i \xrightarrow{a.s.} \mathbb{E}(Z_i) = \mathbb{E}(Z_1) = \tilde{S}_\alpha(\tilde{r}).$$

On 3. From 2., we have that $\mathbb{E}(Z_i) = \tilde{S}_\alpha(\tilde{r})$ and $\text{Var}(Z_i) = \tilde{S}_\alpha(\tilde{r})(1 - \tilde{S}_\alpha(\tilde{r}))$. Applying the central limit theorem implies

$$\sqrt{n} \frac{\tilde{S}_{n,\alpha}(\tilde{r}) - \tilde{S}_\alpha(\tilde{r})}{\sqrt{\tilde{S}_\alpha(\tilde{r})(1 - \tilde{S}_\alpha(\tilde{r}))}} = \frac{\frac{1}{n} \sum_{i=1}^n Z_i - \mathbb{E}(Z_1)}{\sqrt{\text{Var}(Z_1)}} = \frac{\sum_{i=1}^n Z_i - n\mathbb{E}(Z_1)}{\sqrt{n \text{Var}(Z_1)}} \xrightarrow{d} N(0, 1).$$

On 4. Define

$$\tilde{D}_n := \sup_{\tilde{r} \in D_{\tilde{r}}(\alpha)} |\tilde{S}_{n,\alpha}(\tilde{r}) - \tilde{S}_\alpha(\tilde{r})|.$$

- (i) \tilde{S}_α is continuous and monotonically decreasing in \tilde{r} . Hence, we can find a decomposition $\tilde{r}_{\min} = z_0 < z_1 < z_2 < \dots < z_{m-1} < z_m = \tilde{r}_{\max}$ such that $\tilde{S}_\alpha(z_0) = 1, \tilde{S}_\alpha(z_1) = \frac{m-1}{m}, \tilde{S}_\alpha(z_2) = \frac{m-2}{m}, \dots, \tilde{S}_\alpha(z_{m-1}) = \frac{1}{m}, \tilde{S}_\alpha(z_m) = 0$ and where \tilde{r}_{\min} is the smallest $r \in D_{\tilde{r}}(\alpha)$ and similar \tilde{r}_{\max} the largest $r \in D_{\tilde{r}}(\alpha)$.
- (ii) We use this decomposition to obtain approximations of $\tilde{S}_{n,\alpha}(z) - \tilde{S}_\alpha(z)$ for arbitrary $z \in D_{\tilde{r}}(\alpha)$. Let k be such that $z \in [z_k, z_{k+1})$. Then,

$$\begin{aligned} \tilde{S}_{n,\alpha}(z) - \tilde{S}_\alpha(z) &\leq \tilde{S}_{n,\alpha}(z_k) - \tilde{S}_\alpha(z_{k+1}) = \tilde{S}_{n,\alpha}(z_k) - \left(\tilde{S}_{n,\alpha}(z_k) - \frac{1}{m}\right) \\ \tilde{S}_{n,\alpha}(z) - \tilde{S}_\alpha(z) &\geq \tilde{S}_{n,\alpha}(z_{k+1}) - \tilde{S}_\alpha(z_k) = \tilde{S}_{n,\alpha}(z_{k+1}) - \left(\tilde{S}_{n,\alpha}(z_{k+1}) + \frac{1}{m}\right) \end{aligned}$$

due to the monotonicity of \tilde{S}_α .

- (iii) For $m \in \mathbb{N}, k = 0, 1, \dots, m$, define

$$A_{m,k} := \left\{ \tilde{\omega}_\alpha \in \tilde{\Omega}_\alpha : \lim_{n \rightarrow \infty} \tilde{S}_{n,\alpha}(z_k; \tilde{\omega}_\alpha) = \tilde{S}_\alpha(z_k) \right\}.$$

Due to the almost sure convergence of $\tilde{S}_{n,\alpha}$ from 2., we have

$$\mathbb{P}[A_{m,k}] = 1 \quad \forall m \in \mathbb{N}, \quad k = 0, 1, \dots, m.$$

- (iv) Define $A_m = \bigcap_{k=0}^m A_{m,k}$. This is a finite intersection of sets such that $\mathbb{P}[A_m] = 1$ for all $m \in \mathbb{N}$. Define $A = \bigcap_{m \in \mathbb{N}} A_m$. This is a countable intersection of sets such that $\mathbb{P}[A] = 1$.
- (v) Consider now $\tilde{\omega}_\alpha \in A_m$. By definition of $A_{m,k}$, there exists an $n(\tilde{\omega}_\alpha, m) \in \mathbb{N}$ such that

$$|\tilde{S}_{n,\alpha}(z_k; \tilde{\omega}_\alpha) - \tilde{S}_\alpha(z_k)| < \frac{1}{m} \quad \forall n > n(\tilde{\omega}_\alpha, m), \quad k = 1, \dots, m. \text{ Hence,}$$

$$|\tilde{S}_{n,\alpha}(z) - \tilde{S}_\alpha(z)| < \frac{1}{m} \quad \forall \tilde{\omega}_\alpha \in A_m, \quad n > n(\tilde{\omega}_\alpha, m), \quad z \in D_{\tilde{r}}(\alpha).$$

From (ii), it follows

$$\tilde{D}_n(\tilde{\omega}_\alpha) := \sup_{\tilde{r} \in D_{\tilde{r}(\alpha)}} |\tilde{S}_{n,\alpha}(\tilde{r}; \tilde{\omega}_\alpha) - \tilde{S}_\alpha(\tilde{r})| < \frac{2}{m}.$$

Furthermore, due to the definition of A , $\tilde{\omega}_\alpha \in A$ is element of all $A_m, m \in \mathbb{N}$. Hence, $\forall m \in \mathbb{N}$ there exists an $n(\tilde{\omega}_\alpha, m) \in \mathbb{N}$ such that $\forall n > n(\tilde{\omega}_\alpha, m)$

$$0 \leq \tilde{D}_n(\tilde{\omega}_\alpha) < \frac{2}{m} \text{ and in consequence } \lim_{n \rightarrow \infty} \tilde{D}_n(\tilde{\omega}_\alpha) = 0 \quad \forall \tilde{\omega}_\alpha \in A.$$

Finally, we have $\{\tilde{\omega}_\alpha \in \tilde{\Omega}_\alpha: \lim_{n \rightarrow \infty} \tilde{D}_n(\tilde{\omega}_\alpha) = 0\} \supseteq A$ and from 4. that $\mathbb{P}[A] = 1$ holds such that

$$\mathbb{P}\left\{\tilde{\omega}_\alpha \in \tilde{\Omega}_\alpha: \lim_{n \rightarrow \infty} \tilde{D}_n(\tilde{\omega}_\alpha) = 0\right\} \geq \mathbb{P}[A] = 1.$$

□

A.4 Proof of Lemma 11

Proof The uniqueness of $\tilde{r}_{\alpha,\tau}$ yields $\tilde{S}_\alpha(r_{\alpha,\tau} + \varepsilon) < \tau < \tilde{S}_\alpha(r_{\alpha,\tau} - \varepsilon)$ for any $\varepsilon > 0$. The strong consistency of $\tilde{S}_{n,\alpha}(\tilde{r})$ furthermore ensures

$$\begin{aligned} \tilde{S}_{n,\alpha}(\tilde{r}_{\alpha,\tau} - \varepsilon) &\xrightarrow{a.s.} \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau} - \varepsilon) \\ \tilde{S}_{n,\alpha}(\tilde{r}_{\alpha,\tau} + \varepsilon) &\xrightarrow{a.s.} \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau} + \varepsilon) \end{aligned}$$

which is equivalent to

$$\begin{aligned} \mathbb{P}\left(\lim_{n \rightarrow \infty} \tilde{S}_{n,\alpha}(\tilde{r}_{\alpha,\tau} - \varepsilon) = \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau} - \varepsilon) > \tau\right) &= 1 \\ \mathbb{P}\left(\lim_{n \rightarrow \infty} \tilde{S}_{n,\alpha}(\tilde{r}_{\alpha,\tau} + \varepsilon) = \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau} + \varepsilon) < \tau\right) &= 1. \end{aligned}$$

Using that almost sure convergence $\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$ is equivalent to $\lim_{n \rightarrow \infty} \mathbb{P}(|X_m - X| < \varepsilon \quad \forall m \geq n)$ in combination with $\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$ implies

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\tilde{S}_{m,\alpha}(\tilde{r}_{\alpha,\tau} + \varepsilon) < \tau < \tilde{S}_{m,\alpha}(\tilde{r}_{\alpha,\tau} - \varepsilon) \quad \forall m \geq n\right) = 1.$$

Due to the monotonicity of \tilde{S}_α , we have $\tilde{S}_\alpha(\tilde{r}) \leq \tau \Leftrightarrow \tilde{r} \leq \tilde{S}_\alpha^{-1}(\tau)$ and therefore

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\left(\tilde{r}_{\alpha,\tau} + \varepsilon < \tilde{S}_{m,\alpha}^{-1}(\tau) = \tilde{r}_{m,\alpha,\tau} < \tilde{r}_{\alpha,\tau} - \varepsilon \quad \forall m \geq n\right) &= 1 \\ \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}\left(|\tilde{r}_{m,\alpha,\tau} - \tilde{r}_{\alpha,\tau}| < \varepsilon \quad \forall m \geq n\right) &= 1. \end{aligned}$$

Finally, $\tilde{q}_{j,n,\alpha,\tau} \xrightarrow{a.s.} \tilde{q}_{j,\alpha,\tau}$, $j = 1, 2$, is a direct consequence of the continuous mapping theorem which in turn implies $\tilde{\mathbf{q}}_{n,\alpha,\tau} = (\tilde{q}_{1,n,\alpha,\tau}, \tilde{q}_{2,n,\alpha,\tau})' \xrightarrow{a.s.} \tilde{\mathbf{q}}_{\alpha,\tau} = (\tilde{q}_{1,\alpha,\tau}, \tilde{q}_{2,\alpha,\tau})'$, compare Serfling (1980, 1.P, 2.b on page 52). \square

A.5 Proof of Theorem 13

Proof For the proof of Theorem 13, we will use Lemma 10 together with the following Lemma 14.

Lemma 14 (Jump heights of $\tilde{S}_{n,\alpha}$) *Given the general assumptions from Sect. 2.1, the ordered sample $\tilde{R}_{(1)} < \tilde{R}_{(2)} < \dots < \tilde{R}_{(n-1)} < \tilde{R}_{(n)}$ of distances $\tilde{R}_i = \min(\frac{1-\tilde{Y}_{i1}}{\cos(\alpha)}, \frac{1-\tilde{Y}_{i2}}{\sin(\alpha)})$ will almost surely have no ties and therefore*

$$|\tilde{S}_{n,\alpha}(\tilde{S}_{n,\alpha}^{-1}(\tau)) - \tau| \leq \frac{1}{n} \text{ a.s.}$$

From Lemma 10.3, we have that for any $\tilde{r} \in D_\alpha(\tilde{r})$ with survivor function \tilde{S}_α

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}) - \tilde{S}_\alpha(\tilde{r})) \xrightarrow{d} N(0, \tilde{S}_\alpha(\tilde{r})(1 - \tilde{S}_\alpha(\tilde{r})))$$

holds. Let $\tilde{r} = \tilde{r}_{\alpha,\tau} = \tilde{S}_\alpha^{-1}(\tau)$. Then, we know that

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}_{\alpha,\tau}) - \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau})) \xrightarrow{d} N(0, \tau(1 - \tau)).$$

Using the property of stochastic equicontinuity for $\tilde{S}_{n,\alpha}$ interpreted as an empirical process (for an introduction and definition of stochastic equicontinuity, see Andrews 1994), we can replace $\tilde{r}_{\alpha,\tau}$ by a consistent estimator $\tilde{r}_{n,\alpha,\tau}$ such that

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}_{n,\alpha,\tau}) - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})) \xrightarrow{d} N(0, \tau(1 - \tau))$$

holds. From Lemma (ii) in Serfling (1980, Sec. 1.1.4, p. 3) it now follows that

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}_{n,\alpha,\tau}) - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})) \geq \sqrt{n}(\tau - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})).$$

Since \tilde{f}_α is continuous, the probability of observing duplicates of \tilde{R}_i is zero. Hence, using Lemma 14

$$\sqrt{n}(\tilde{S}_{n,\alpha}(\tilde{r}_{n,\alpha,\tau}) - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})) = \sqrt{n}(\tau - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})) + \mathcal{O}_p(1/\sqrt{n})$$

holds with probability one which (using Lemma 10.2) implies

$$\sqrt{n}(\tau - \tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau})) \xrightarrow{d} N(\tau(1 - \tau)).$$

Applying the Delta-method, i.e. Taylor expansion around of $\tilde{S}_\alpha, \tilde{r}_{\alpha,\tau}$ yields

$$\tilde{S}_\alpha(\tilde{r}_{n,\alpha,\tau}) \approx \tilde{S}_\alpha(\tilde{r}_{\alpha,\tau}) - \tilde{f}_\alpha(\tilde{r}_{\alpha,\tau})(\tilde{r}_{n,\alpha,\tau} - \tilde{r}_{\alpha,\tau}),$$

for $\tilde{r}_{\alpha,\tau}$ on the line segment between $\tilde{r}_{n,\alpha,\tau}$ and $\tilde{r}_{\alpha,\tau}$. The last step is to apply Slutsky's theorem and the fact that $\tilde{r}_{\alpha,\tau} \rightarrow \tilde{r}_{\alpha,\tau}$ since $\tilde{r}_{n,\alpha,\tau} \rightarrow \tilde{r}_{\alpha,\tau}$, such that we obtain

$$\sqrt{n}(\tilde{r}_{n,\alpha,\tau} - \tilde{r}_{\alpha,\tau}) = \sqrt{n} \frac{\tau - S_\alpha(\tilde{r}_{n,\alpha,\tau})}{\tilde{f}_\alpha(\tilde{r}_{\alpha,\tau})} \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{(\tilde{f}_\alpha(\tilde{r}_{\alpha,\tau}))^2}\right).$$

□

References

- Abdous, B., Theodorescu, R.: Note on the spatial quantile of a random vector. *Stat. Probab. Lett.* **13**, 333–336 (1992)
- Andrews, D.W.K.: Empirical process methods in econometrics. In: Engle, R.F., McFadden, D.L. (eds.) *Handbook of Econometrics*, vol. 4, pp. 2247–2294. Elsevier Science B.V., North-Holland, New York (1994)
- Belzunce, F., Castaño, A., Olvera-Cervantes, A., Suárez-Llorens, A.: Quantile curves and dependence structure for bivariate distributions. *Comput. Stat. Data Anal.* **51**, 5112–5129 (2007)
- Berkelaar, M., et al.: IpSolve: Interface to “Lp_solve” v. 5.5 to Solve Linear/Integer Program. R package version 5.6.13 (2015)
- Carlier, G., Chernozhukov, V., Galichon, A.: Vector quantile regression: an optimal transport approach. *Ann. Stat.* **44**(3), 1165–1192 (2016)
- Chakraborty, B.: On affine equivariant multivariate quantiles. *Ann. Inst. Stat. Math.* **53**, 380–403 (2001)
- Chaudhuri, P.: On a geometric notion of quantiles for multivariate data. *J. Am. Stat. Assoc.* **91**, 862–872 (1996)
- Chen, L.-A., Welsh, A.H.: Distribution-function-based bivariate quantiles. *J. Multivar. Anal.* **83**, 208–231 (2002)
- Chernozhukov, V., Galichon, A., Hallin, M., Henry, M.: Monge–Kantorovich depth, quantiles, ranks, and signs. *Ann. Stat.* **45**(1), 223–256 (2017)
- Einmahl, J.H.J., Mason, D.M.: Generalized quantile processes. *Ann. Stat.* **20**, 1062–1078 (1992)
- Ferguson, T.S.: *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York (1967)
- Fernandez-Ponce, J.M., Suarez-Llorens, A.: Quantile curves and dependence structure for bivariate distributions. *Comput. Stat. Data Anal.* **17**, 236–256 (2007)
- Genest, C., Segers, J.: On the covariance of the asymptotic empirical copula process. *J. Multivar. Anal.* **101**, 1837–1845 (2010)
- Genest, C., Ghoudi, K., Rivest, L.-P.: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552 (1995)
- Guggisberg, M.: *A Bayesian Approach to Multiple-Output Quantile Regression*, Technical report (2016)
- Hallin, M.: On distribution and quantile functions, ranks and signs in r^d , ECARES working paper 2017-34 (2017)
- Hallin, M., Paindaveine, D., Šíman, M.: Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Stat.* **38**, 635–669 (2010)
- Joe, H.: *Dependence Modeling with Copulas*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, London (2014)
- Klein, N.: bivquant: Estimation of Bivariate Quantiles. R package version 0.1 (2019)
- Klein, N., Kneib, T.: Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. *Stat. Comput.* **26**, 841–860 (2016)
- Koenker, R.: *Quantile Regression*. Economic Society Monographs. Cambridge University Press, New York (2005)
- Koltchinskii, V.I.: M -estimation, convexity and quantiles. *Ann. Stat.* **25**, 435–477 (1997)

- Koshevoy, G., Mosler, K.: Zonoid trimming for multivariate distributions. *Ann. Stat.* **25**, 1998–2017 (1997)
- Liu, R.Y.: On a notion of data depth based on random simplices. *Ann. Stat.* **18**, 405–414 (1990)
- Liu, R.Y., Parelius, J.M., Singh, K.: Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh). *Ann. Stat.* **27**, 783–858 (1999)
- Mosler, K.: *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*. Springer, New York (2002)
- Oja, H.: Descriptive statistics for multivariate distributions. *Stat. Probab. Lett.* **1**, 327–332 (1983)
- Pokotylo, O., Mozharovskyi, P., Dyckerhoff, R.: *ddalpha: Depth-Based Classification and Calculation of Data Depth*. R package version 1.3.1 (2015)
- Serfling, R.: *Approximation Theorems of Mathematical Statistics*. Series in Probability and Mathematical Statistics. Wiley, New York (1980)
- Serfling, R.: Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerl.* **56**, 214–232 (2002)
- Small, C.G.: A survey of multidimensional medians. *Int. Stat. Rev.* **58**, 263–277 (1990)
- Tukey, J.: Mathematics and the picturing of data. In: *Proceedings of the 1975 International Congress of Mathematicians*, vol. 2, pp. 523–531 (1975)
- Zuo, Y., Serfling, R.: General notions of statistical depth function. *Ann. Stat.* **28**, 461–482 (2000)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.