Multivariate Expectiles, Expectile Depth and Multiple-Output Expectile Regression

Abdelaati Daouia, Davy Paindaveine

Multivariate Expectiles, Expectile Depth and Multiple-Output Expectile Regression

Abdelaati Daouia
Toulouse School of Economics, Université Toulouse Capitole, France
Davy Paindaveine
Université libre de Bruxelles, Belgium
Toulouse School of Economics, Université Toulouse Capitole, France

Summary. Despite the importance of expectiles in fields such as econometrics, risk management, and extreme value theory, expectile regression — or, more generally, M-quantile regression — unfortunately remains limited to single-output problems. To improve on this, we define hyperplane-valued multivariate M-quantiles that show strong advantages over their point-valued competitors. Our M-quantiles are directional in nature and provide centrality regions when all directions are considered. These regions define new statistical depths, the *halfspace M-depths*, that include the celebrated Tukey depth as a particular case. We study thoroughly the proposed M-quantiles, halfspace M-depths, and corresponding regions. M-depths not only provide a general framework to consider Tukey depth, expectile depth, L_r-depths, etc., but are also of interest on their own. However, since our original motivation was to consider multiple-output expectile regression, we pay more attention to the expectile case and show that expectile depth and multivariate expectiles enjoy distinctive properties that will be of primary interest to practitioners: expectile depth is maximized at the mean vector, is smoother than the Tukey depth, and exhibits surprising monotonicity properties that are key for computational purposes. Finally, our multivariate expectiles allow defining multiple-output expectile regression methods, that, in risk-oriented applications in particular, are preferable to their analogs based on standard quantiles.

1. Introduction

Whenever one wants to assess the impact of a vector of covariates $\mathbf{X}$ on a scalar response Y, mean regression, in its various forms (linear, nonlinear, or nonparametric), remains by far the most popular method. Mean regression, however, only captures the conditional mean

$$\mu(\mathbf{x}) := \mathrm{E}[Y|\mathbf{X}=\mathbf{x}] = \arg\min_{\theta\in\mathbb{R}} \mathrm{E}\big[(Y-\theta)^2|\mathbf{X}=\mathbf{x}\big]$$

of the response, hence fails to describe thoroughly the conditional distribution of Y given $\mathbf{X}$. Such a thorough description is given by the Koenker and Basset (1978) *quantile regression*, that considers the conditional quantiles

$$q_\alpha(\mathbf{x}) := \arg\min_{\theta\in\mathbb{R}} \mathrm{E}\big[\rho_{\alpha,L_1}(Y-\theta)|\mathbf{X}=\mathbf{x}\big], \quad \alpha\in(0,1), \tag{1}$$

where $\rho_{\alpha,L_1}(t) := \{(1-\alpha)\mathbb{I}[t<0] + \alpha\mathbb{I}[t>0]\}|t|$ is the *check function* (throughout, $\mathbb{I}[A]$ stands for the indicator function of A). An alternative to quantile regression is the Newey and Powell (1987) *expectile regression*, that focuses on the conditional expectiles

$$e_\alpha(\mathbf{x}) := \arg\min_{\theta\in\mathbb{R}} \mathrm{E}\big[\rho_{\alpha,L_2}(Y-\theta)|\mathbf{X}=\mathbf{x}\big], \quad \alpha\in(0,1), \tag{2}$$

where $\rho_{\alpha,L_2}(t) := \{(1-\alpha)\mathbb{I}[t<0] + \alpha\mathbb{I}[t>0]\}t^2$ is an asymmetric quadratic loss function, in the same way the check function is an asymmetric absolute loss function. Conditional expectiles, like conditional quantiles, fully characterize the conditional distribution of the response and nicely include the conditional mean $\mu(\mathbf{x})$ as a particular case. Sample conditional expectiles, unlike their quantile counterparts, are sensitive to extreme observations, but this may actually be an asset in some applications; in financial risk management, for instance, quantiles are often criticized for being too liberal (due to their insensitivity to extreme losses) and expectiles are therefore favoured in any prudent and reactive risk analysis (Daouia et al., 2018).

Expectile regression shows other advantages over quantile regression, of which we mention only a few here. First, inference on quantiles requires estimating nonparametrically the conditional density of the response at the considered quantiles, which is notoriously difficult. In contrast, inference on expectiles can be performed without resorting to any smoothing, bootstrap or Bayesian technique, which makes it easy, e.g., to test for homoscedasticity or for conditional symmetry in linear regression models (Newey and Powell, 1987). Second, since expectile regression includes classical mean regression as a particular case, it is closer to the least squares notion of explained variance and, in parametric cases, expectile regression coefficients can be interpreted with respect to variance heteroscedasticity. This is of particular relevance in complex regression specifications including nonlinear, random or spatial effects (Sobotka and Kneib, 2012). Third, expectile smoothing techniques, based on kernel smoothing (Yao and Tong, 1996) or penalized splines (Schnabel and Eilers, 2009), show better smoothness and stability than their quantile counterparts and also make expectile crossings far more rare than quantile crossings; see Schnabel and Eilers (2009), Eilers (2013) and Schulze Waltrup et al. (2015). These points explain why expectiles recently regained much interest in econometrics; see, e.g., Kuan et al. (2009), De Rossi and Harvey (2009), and Embrechts and Hofert (2014).

Despite these nice properties, expectile regression still suffers from an important drawback, namely its limitation to single-output problems. In contrast, many works developed multiple-output *quantile* regression methods. We refer, among others, to Chakraborty (2003), Cheng and De Gooijer (2007), Wei (2008), Hallin et al. (2010), Cousin and Di Bernardino (2013), Waldmann and Kneib (2015), Hallin et al. (2015), Carlier et al. (2016, 2017), and Chavas (2018). This is in line with the fact that defining a satisfactory concept of multivariate quantile is a classical problem that has attracted much attention in the literature (we refer to Serfling (2002) and to the references therein), whereas the literature on multivariate expectiles is much sparser. Some early efforts to define multivariate expectiles can be found in Koltchinski (1997), Breckling et al. (2001) and Kokic et al. (2002), that all define more generally multivariate versions of the *M-quantiles* from Breckling and Chambers (1988) (a first concept of multivariate M-quantile was actually already discussed in Breckling and Chambers (1988) itself). Recently, there has been a renewed interest in defining multivariate expectiles; see, e.g., Cousin and Di Bernardino (2014), Maume-Deschamps et al. (2017a,b), and Herrmann et al. (2018). Multivariate risk handling in finance and actuarial sciences is mostly behind this growing interest, as will be discussed in Section S.3 of the

supplementary materials.

This paper introduces multivariate expectiles—and, more generally, multivariate M-quantiles—that enjoy many desirable properties, particularly in terms of affine equivariance (while this equivariance property is a standard requirement in the companion problem of defining multivariate quantiles, the available concepts of multivariate expectiles or M-quantiles are at best orthogonal-equivariant). Like their competitors, our multivariate M-quantiles are directional quantities, but they are hyperplane-valued rather than point-valued. Despite this different nature, they still generate centrality regions when all directions are considered. While this has not been discussed in the multivariate M-quantile literature (nor in the multivariate expectile one), this defines an M-concept of statistical depth. The resulting *halfspace M-depths* generalize the Tukey (1975) halfspace depth and satisfy the desirable properties of depth from Zuo and Serfling (2000). Remarkably, these M-depths can alternatively be obtained by replacing, in the Tukey depth, standard quantile outlyingness with M-quantile outlyingness, which a posteriori shows that our multivariate M-quantile concept is a (if not the) most natural one. This is a key result that allows us to study the structural properties of M-depths, which will apply generically to the Tukey depth, expectile depth, L_r-depths, etc. While M-depths are of interest on their own (they allow, e.g., achieving a balance between robustness and efficiency), we pay more attention to the particular case of expectile depth, in line with our original motivation to consider multiple-output expectile regression. Expectile depth actually offers many properties that, in comparison with Tukey depth, should be appealing to practitioners: it is maximized at the mean vector, it is smooth, and it shows a surprising monotonicity that is key for its computation. Finally, our multivariate expectiles allow us to define multiple-output expectile regression methods, that, in risk-oriented applications in particular, will be preferable to their analogs based on standard quantiles.

The outline of the paper is as follows. In Section 2, we carefully define univariate M-quantiles by extending a result from Jones (1994). In Section 3, we introduce our concept of multivariate M-quantiles and compare the resulting M-quantile regions with those associated with alternative M-quantile concepts. In Section 4, we define the halfspace M-depths and investigate their properties, whereas, in Section 5, we focus on the particular case of expectile depth. In Section 6, we explain how these expectiles allow performing multiple-output expectile regression, which is illustrated on simulated and real data. For the sake of completeness, a supplement describes some of the main competing multivariate M-quantile concepts and provides asymptotic results for the proposed M-depths. It also discusses the relation between multivariate M-quantiles and risk measures, and it shows that our expectiles satisfy the coherency axioms of multivariate risk measures. Finally, the supplement proves all results of this paper.

2. On univariate M-quantiles

As mentioned in Jones (1994), the Breckling and Chambers (1988) M-quantiles are related to M-estimates (or M-functionals) of location in the same way standard quantiles are related to the median. In line with (1)-(2), the order-α M-quantile of a probability measure P over $\mathbb{R}$ may be thought of as

$$\theta_\alpha^\rho(P) := \arg\min_{\theta\in\mathbb{R}} O_\alpha^\rho(\theta), \qquad \text{with } O_\alpha^\rho(\theta) := \mathrm{E}\big[\rho_\alpha(Z-\theta)-\rho_\alpha(Z)\big], \tag{3}$$

where $\rho_\alpha(t) := \{(1-\alpha)\mathbb{I}[t < 0] + \alpha\mathbb{I}[t > 0]\}\rho(t)$ is based on a suitable symmetric loss function ρ and where the random variable Z has distribution P. Standard quantiles are obtained for the absolute loss function $\rho(t) = |t|$, whereas expectiles are associated with the quadratic loss function $\rho(t) = t^2$. One may also consider the Huber loss functions

$$\rho_c(t) := \frac{t^2}{2c}\,\mathbb{I}[|t| < c] + \Big(|t| - \frac{c}{2}\Big)\mathbb{I}[|t| \geq c], \qquad c > 0, \tag{4}$$

that allow recovering, up to an irrelevant positive scalar factor, the absolute value and quadratic loss functions above. The resulting M-quantiles $\theta_\alpha^{\rho_c}(P)$ thus offer a continuum between quantiles and expectiles.

The M-quantiles in (3) may be non-unique: for instance, if $\rho(t) = |t|$ and $P = P_n$ is the empirical probability measure associated with a sample of size n, then $\theta_\alpha^\rho(P)$, for any $\alpha = 1/n, \ldots, (n-1)/n$, is an interval with non-empty interior. We will therefore adopt an alternative definition of M-quantiles, that results from Theorem 1 below. The result, that is of independent interest, significantly extends the theorem in Jones (1994) (in particular, Jones' result excludes the absolute loss function and all Huber loss functions). To state the result, we define the class $\mathcal{C}$ of loss functions $\rho : \mathbb{R} \to \mathbb{R}^+$ that are convex, symmetric and such that $\rho(t) = 0$ for $t = 0$ only. For any $\rho \in \mathcal{C}$, we write ψ_- for the left-derivative of ρ (existence follows from convexity of ρ) and we denote as $\mathcal{P}^\rho$ the collection of probability measures P over $\mathbb{R}$ such that (i) $P[\{\theta\}] < 1$ for any $\theta \in \mathbb{R}$ and (ii) $\int_{-\infty}^{\infty} |\psi_-(z-\theta)|\, dP(z) < \infty$ for any $\theta \in \mathbb{R}$. We then have the following result.

Theorem 1 *Fix $\alpha \in (0,1)$, $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho$. Let Z be a random variable with distribution P. Then, (i) $\theta \mapsto O_\alpha^\rho(\theta)$ is well-defined for any θ, and it is left- and right-differentiable over $\mathbb{R}$, hence also continuous over $\mathbb{R}$. (ii) The corresponding left- and right-derivatives satisfy $O_{\alpha-}^{\rho\prime}(\theta) \leq O_{\alpha+}^{\rho\prime}(\theta)$ at any θ. (iii) The sign of $O_{\alpha+}^{\rho\prime}(\theta)$ is the same as that of $G^\rho(\theta) - \alpha$, where we let*

$$G^\rho(\theta) := \frac{\mathrm{E}[|\psi_-(Z-\theta)|\mathbb{I}[Z \leq \theta]]}{\mathrm{E}[|\psi_-(Z-\theta)|]};$$

(iv) $\theta \mapsto G^\rho(\theta)$ is a cumulative distribution function over $\mathbb{R}$. (v) The order-α M-quantile *of P, which we define as*

$$\theta_\alpha^\rho(P) := \inf\big\{\theta \in \mathbb{R} : G^\rho(\theta) \geq \alpha\big\}, \tag{5}$$

minimizes $\theta \mapsto O_\alpha^\rho(\theta)$ over $\mathbb{R}$, hence provides a unique representative of the argmin in (3). (vi) If ψ_- is continuous over $\mathbb{R}$ (or if P is non-atomic), then G^ρ is continuous at $\theta_\alpha^\rho(P)$, so that $G^\rho(\theta_\alpha^\rho(P)) = \alpha$.

For the absolute and quadratic loss functions, one has

$$G^\rho(\theta) = \mathbb{P}[Z \leq \theta] \quad \text{and} \quad G^\rho(\theta) = \frac{\mathrm{E}[|Z-\theta|\mathbb{I}[Z \leq \theta]]}{\mathrm{E}[|Z-\theta|]},$$

respectively (throughout, $\mathbb{P}$ refers to the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which all random variables and random vectors are defined). When case (vi) applies (as it does for the quadratic loss function and any Huber loss function), $G^\rho(\theta_\alpha^\rho(P)) = \alpha$ plays the role of the first-order condition associated with (3); when it does not, this first-order condition is to be replaced by the more general one in (5), which provides the usual definition of standard

quantiles. For our later purposes, it is important to note that the larger $G^\rho(\theta)(\leq 1/2)$ (resp., $1 - G^\rho(\theta - 0)(\leq 1/2)$, where $H(\theta - 0)$ denotes the limit of $H(t)$ as $t \nearrow \theta$), the less θ is outlying below (resp., above) the "median" $\theta^\rho_{1/2}(P)$. Therefore, $MD^\rho(\theta, P) := \min(G^\rho(\theta), 1 - G^\rho(\theta - 0))$ measures the centrality—as opposed to outlyingness—of θ with respect to P. In other words, $MD^\rho(\theta, P)$ defines a measure of *statistical depth* over $\mathbb{R}$; see Zuo and Serfling (2000). In the sequel, we will extend this "M-depth" to $\mathbb{R}^d$. Note that, for $d = 1$ and $\rho(t) = |t|$, the depth $MD^\rho(\theta, P)$ reduces to the Tukey (1975) halfspace depth.

3. Our multivariate M-quantiles

The first multivariate M-quantiles were defined in Breckling and Chambers (1988) (and include the celebrated geometric quantiles from Chaudhuri (1996), as well as the geometric expectiles recently studied in Herrmann et al. (2018)). Since then, several concepts of multivariate M-quantiles have been proposed. For the sake of completeness, we describe the multivariate M-quantiles above, as well as those from Breckling et al. (2001) and Kokic et al. (2002), in Section S.1 of the supplement. For now, it is only important to mention that, possibly after an unimportant reparametrization, all aforementioned multivariate M-quantiles can be written as functionals $P \mapsto \boldsymbol{\theta}^\rho_{\alpha,\mathbf{u}}(P)$ that take values in $\mathbb{R}^d$ and are indexed by a scalar order $\alpha \in (0, 1)$ and a direction $\mathbf{u} \in \mathcal{S}^{d-1} := \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\|^2 := \mathbf{z}'\mathbf{z} = 1\}$; here, P is a probability measure over $\mathbb{R}^d$. Typically, $\boldsymbol{\theta}^\rho_{\alpha,\mathbf{u}}(P)$ does not depend on $\mathbf{u}$ for $\alpha = 1/2$, and the resulting common location is the center (the "median") of P. Our multivariate M-quantiles will also be of a directional nature but they will be *hyperplane-valued* rather than *point-valued*. For $d = 1$, it is often important to know whether some test statistic takes a value below or above a given quantile, that is used as a critical value; for $d > 1$, hyperplane-valued quantiles, unlike point-valued ones, could similarly be used as critical values with vector-valued test statistics.

To be able to define our M-quantiles, denote as $P_\mathbf{u}$ the distribution of $\mathbf{u}'\mathbf{Z}$ when the random d-vector $\mathbf{Z}$ has distribution P; here, $\mathbf{u} \in \mathcal{S}^{d-1}$. Consider then the collection $\mathcal{P}^\rho_d$ of probability measures P over $\mathbb{R}^d$ such that (i) no hyperplane of $\mathbb{R}^d$ has P-probability one and such that (ii) $\int_{-\infty}^{\infty} |\psi_-(z-\theta)|\, dP_\mathbf{u}(z) < \infty$ for any $\theta \in \mathbb{R}$ and $\mathbf{u} \in \mathcal{S}^{d-1}$ (note that $\mathcal{P}^\rho_1 = \mathcal{P}^\rho$, where $\mathcal{P}^\rho$ was defined in Section 2). Our concept of multivariate M-quantiles is then the following.

Definition 1 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^\rho_d$. Let $\mathbf{Z}$ be a random d-vector with distribution P. Then, for any $\alpha \in (0, 1)$ and $\mathbf{u} \in \mathcal{S}^{d-1}$, the* order-$\alpha$ M-quantile of P in direction $\mathbf{u}$ *is the hyperplane*

$$\pi^\rho_{\alpha,\mathbf{u}}(P) = \left\{\mathbf{z} \in \mathbb{R}^d : \mathbf{u}'\mathbf{z} = \theta^\rho_\alpha(P_\mathbf{u})\right\},$$

where $\theta^\rho_\alpha(P_\mathbf{u})$ is the order-α M-quantile of $P_\mathbf{u}$; see (5). The corresponding upper-halfspace $H^\rho_{\alpha,\mathbf{u}}(P) = \left\{\mathbf{z} \in \mathbb{R}^d : \mathbf{u}'\mathbf{z} \geq \theta^\rho_\alpha(P_\mathbf{u})\right\}$ will be called order-α M-quantile halfspace of P in direction $\mathbf{u}$.

For $\rho(t) = |t|$, these quantile hyperplanes reduce to those from Paindaveine and Šiman (2011) (see also Kong and Mizera, 2012), whereas $\rho(t) = t^2$ provides the proposed multivariate expectiles. For any loss function ρ, the hyperplanes $\pi^\rho_{\alpha,\mathbf{u}}$ are linked in a straightforward way to the direction $\mathbf{u}$: they are simply orthogonal to $\mathbf{u}$. In contrast, the point-valued

competitors mentioned above typically depend on $\mathbf{u}$ in an intricate way, and in particular $\boldsymbol{\theta}^{\rho}_{\alpha,\mathbf{u}}(P)$ usually does not belong to the halfline with direction $\mathbf{u}$ originating from the corresponding median (see above). Note that the "intercepts" of our M-quantile hyperplanes are the univariate M-quantiles of the projection $\mathbf{u}'\mathbf{Z}$ of $\mathbf{Z}$ onto $\mathbf{u}$, hence also allow for a direct interpretation.

Irrespective of the loss function ρ, competing multivariate M-quantiles fail to be equivariant under affine transformations. Our M-quantiles improve on this.

Theorem 2 *Fix $\rho \in \mathcal{C}_{\text{aff}}$ and $P \in \mathcal{P}^{\rho}_{d}$, where $\mathcal{C}_{\text{aff}}(\subset \mathcal{C})$ is the collection of power loss functions $\rho(t) = |t|^r$ with $r \geq 1$. Let $\mathbf{A}$ be an invertible $d \times d$ matrix and $\mathbf{b}$ be a d-vector. Then, for any $\alpha \in (0,1)$ and $\mathbf{u} \in \mathcal{S}^{d-1}$,*

$$\pi^{\rho}_{\alpha,\mathbf{u}_{\mathbf{A}}}(P_{\mathbf{A},\mathbf{b}}) = \mathbf{A}\pi^{\rho}_{\alpha,\mathbf{u}}(P) + \mathbf{b} \quad \text{and} \quad H^{\rho}_{\alpha,\mathbf{u}_{\mathbf{A}}}(P_{\mathbf{A},\mathbf{b}}) = \mathbf{A}H^{\rho}_{\alpha,\mathbf{u}}(P) + \mathbf{b},$$

where $\mathbf{u}_{\mathbf{A}} := (\mathbf{A}^{-1})'\mathbf{u}/\|(\mathbf{A}^{-1})'\mathbf{u}\|$ and where $P_{\mathbf{A},\mathbf{b}}$ is the distribution of $\mathbf{AZ}+\mathbf{b}$ when $\mathbf{Z}$ is a random d-vector with distribution P.

In the univariate case, the M-quantiles associated with $\rho \in \mathcal{C}_{\text{aff}}$ are known as *L_r-quantiles* and were used for testing symmetry in nonparametric regression (Chen, 1996); the estimation of extreme L_r-quantiles was also recently investigated in Daouia et al. (2019). While Theorem 2 above shows in particular that quantile and expectile hyperplanes are affine-equivariant, the restriction to $\mathcal{C}_{\text{aff}}$ cannot be dropped. For instance, for fixed $c > 0$, the M-quantile hyperplanes $\pi^{\rho_c}_{\alpha,\mathbf{u}}(P)$, associated with the Huber loss functions in (4), fail to be affine-equivariant. Our multivariate extension is not to be blamed for this, however, since the corresponding univariate M-quantiles $\theta^{\rho_c}_{\alpha}(P)$ themselves fail to be scale-equivariant.

At first sight, a possible advantage of any point-valued M-quantiles $\boldsymbol{\theta}^{\rho}_{\alpha,\mathbf{u}}(P)$ is that they naturally generate contours and regions. More precisely, they allow considering, for any $\alpha \in (0, \frac{1}{2}]$, the order-$\alpha$ M-quantile contour $\{\boldsymbol{\theta}^{\rho}_{\alpha,\mathbf{u}}(P) : \mathbf{u} \in \mathcal{S}^{d-1}\}$, the interior part of which is then the corresponding order-α M-quantile region. Our hyperplane-valued M-quantiles, however, also provide centrality regions, hence corresponding contours.

Definition 2 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^{\rho}_{d}$. For any $\alpha \in (0,1)$, the order-α M-quantile region of P is $R^{\rho}_{\alpha}(P) = \bigcap_{\mathbf{u}\in\mathcal{S}^{d-1}} H^{\rho}_{\alpha,\mathbf{u}}(P)$ and the corresponding order-α contour is the boundary $\partial R^{\rho}_{\alpha}(P)$ of $R^{\rho}_{\alpha}(P)$.*

Theorem 1 entails that the univariate M-quantiles $\theta^{\rho}_{\alpha}(P)$ in (5) are monotone non-decreasing functions of α. A direct corollary is that the regions $R^{\rho}_{\alpha}(P)$ are nested (the larger α, the smaller the corresponding region). The proposed regions enjoy many nice properties compared to their competitors resulting from point-valued M-quantiles, as we show on the basis of Theorem 3 below. To state the result, we define the *C-support of P* as $C_P := \{\mathbf{z} \in \mathbb{R}^d : \mathbb{P}[\mathbf{u}'\mathbf{Z} \leq \mathbf{u}'\mathbf{z}] > 0 \text{ for any } \mathbf{u} \in \mathcal{S}^{d-1}\}$, where the random d-vector $\mathbf{Z}$ has distribution P. Clearly, C_P can be thought of as the convex hull of P's support. We then have the following result.

Theorem 3 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^{\rho}_{d}$. Then, for any $\alpha \in (0,1)$, the region $R^{\rho}_{\alpha}(P)$ is a convex and compact subset of C_P. Moreover, if $\rho \in \mathcal{C}_{\text{aff}}$, then $R^{\rho}_{\alpha}(P_{\mathbf{A},\mathbf{b}}) = \mathbf{A}R^{\rho}_{\alpha}(P) + \mathbf{b}$ for any invertible $d \times d$ matrix $\mathbf{A}$ and d-vector $\mathbf{b}$.*

No competing M-quantile regions combine these properties. For instance, the original M-quantile regions from Breckling and Chambers (1988), hence also the geometric quantile regions from Chaudhuri (1996) and their expectile counterparts from Herrmann et al. (2018), may extend far beyond the convex hull of the support; see below. This was actually the motivation for the alternative proposals in Breckling et al. (2001) and Kokic et al. (2002). The regions introduced in these two papers, however, may fail to be convex, which is unnatural. More generally, none of the competing M-quantile or expectile regions are affine-equivariant. This may result in quite pathological behaviors: for instance, Theorem 2.2 from Girard and Stupfler (2017) implies that, if P is an elliptically symmetric probability measure admitting a density f, then, for small α, the geometric quantile contours from Chaudhuri (1996) are "orthogonal" to the principal component structure of P, in the sense that these contours are furthest (resp., closest) to the symmetry center of P in the last (resp., first) principal direction. In contrast, the affine-equivariance result in Theorem 3 ensures that, in such a distributional setup, the shape of our M-quantile contours will match the principal component structure of P.

We illustrate this on the "cigar-shaped" data example from Breckling et al. (2001) and Kokic et al. (2002), for which $P = P_n$ is the empirical probability measure associated with $n = 200$ bivariate observations whose x-values form a uniform grid in $[-1, 1]$ and whose y-values are randomly drawn from the normal distribution with mean 0 and variance .01. Figure 1 draws, for several orders α, the various quantile and expectile contours mentioned in the previous paragraph (our contours $\partial R^{\rho}_{\alpha}(P_n)$ were computed by replacing the intersection in Definition 2 by an intersection over 500 equispaced directions $\mathbf{u}$ in $\mathcal{S}^1$; all competing contours require a similar discretization). Clearly, both the geometric quantiles from Chaudhuri (1996) and geometric expectiles from Herrmann et al. (2018) extend much beyond the convex hull of the data points. Moreover, the aforementioned pathological behavior of the extreme geometric quantiles relative to the principal component structure of P not only shows for these quantiles but also for the corresponding expectiles. Finally, the outer quantile/expectile regions from Breckling et al. (2001) and Kokic et al. (2002) are non-convex in most cases. In line with Theorem 3, our M-quantile regions and contours do not exhibit these deficiencies.

4. Halfspace M-depth

Our M-quantile regions $R^{\rho}_{\alpha}(P)$ are *centrality regions*, in the sense that they group locations $\mathbf{z}$ in the sample space $\mathbb{R}^d$ according to their centrality with respect to the underlying distribution P. This defines the following concept of *statistical depth*.

Definition 3 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}^{\rho}_{d}$. Then, the corresponding* halfspace M-depth *of $\mathbf{z}$ with respect to P is $MD^{\rho}(\mathbf{z}, P) = \sup\{\alpha > 0 : \mathbf{z} \in R^{\rho}_{\alpha}(P)\}$ (here, $\sup \emptyset := 0$).*

Other M-depth concepts could similarly be defined from competing M-quantile regions. However, these M-depths would, irrespective of ρ, fail to meet one of the most classical requirements for depth, namely affine invariance; see Zuo and Serfling (2000). Our M-depths are better in this respect; see Theorem 6(i) below.

For any depth, the corresponding depth regions, that collect locations with depth larger than or equal to a given level α, are of particular interest. The following result shows that

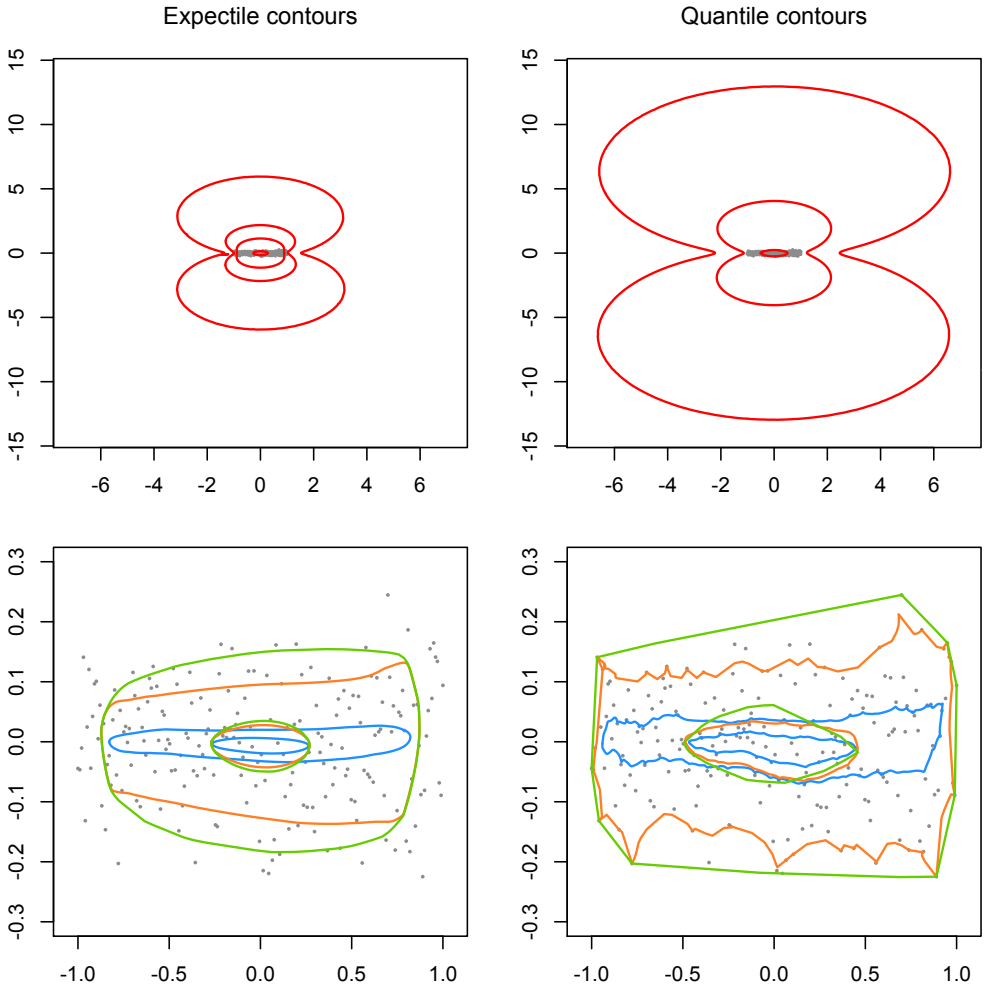

Fig. 1: *(Top:) the geometric expectile contours from Herrmann et al. (2018) (left) and geometric quantile contours from Chaudhuri (1996) (right), for the cigar-shaped data described in Section 3 and for* $\alpha = .00001, .0005, .005,$ *and* $.25$ *(for the smallest* α*, the quantile contour is outside the plot). (Bottom left:) the expectile contours from Breckling et al. (2001) (blue), the* $(\delta = 10)$*-version of the Kokic et al. (2002) expectile contours (orange), and the proposed expectile contours (green), for the same data and for* $\alpha = 1/n = .005$ *and* $.25$*; we use the same value of* δ *as in Kokic et al. (2002). (Bottom right:) the quantile versions of the contours in the bottom left panel. In each panel, the* $(n = 200)$ *data points are shown in grey.*

halfspace M-depth regions strictly coincide with the centrality regions introduced in the previous section.

Theorem 4 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Then, for any $\alpha \in (0,1)$, the level-α depth region $\{\mathbf{z} \in \mathbb{R}^d : MD^\rho(\mathbf{z}, P) \geq \alpha\}$ coincides with $R_\alpha^\rho(P)$.*

This result has several interesting consequences. First, it implies that MD^ρ reduces to the Tukey (1975) halfspace depth for $\rho(t) = |t|$ (since the corresponding centrality regions $R_\alpha^\rho(P)$ are known to be the Tukey depth regions; see, e.g., Theorem 2 in Kong and Mizera, 2012). Second, Theorems 3–4 show that halfspace M-depth regions are convex, so that our M-depth is quasi-concave: for any $\mathbf{z}_0, \mathbf{z}_1 \in \mathbb{R}^d$ and $\lambda \in (0,1)$, one has $MD^\rho((1-\lambda)\mathbf{z}_0 + \lambda\mathbf{z}_1, P) \geq \min(MD^\rho(\mathbf{z}_0, P), MD^\rho(\mathbf{z}_1, P))$. Third, since Theorems 3–4 imply that the mapping $\mathbf{z} \mapsto MD^\rho(\mathbf{z}, P)$ has closed upper level sets, this mapping is upper semicontinuous over $\mathbb{R}^d$ (it is actually continuous over $\mathbb{R}^d$ if P assigns probability zero to all hyperplanes of $\mathbb{R}^d$; see Lemma S.7). Fourth, the compactness of halfspace M-depth regions for $\alpha > 0$, which results again from Theorems 3–4, allows us to establish the existence of an M-deepest location.

Theorem 5 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Then, $\sup_{\mathbf{z} \in \mathbb{R}^d} MD^\rho(\mathbf{z}, P) = MD^\rho(\mathbf{z}_*, P)$ for some $\mathbf{z}_* \in \mathbb{R}^d$.*

The M-deepest location $\mathbf{z}_*$ may fail to be unique. For the halfspace Tukey depth, whenever a unique representative of the deepest locations is needed, a classical solution consists in considering the *Tukey median*, that is defined as the barycenter of the deepest region. The same solution can be adopted for our M-depth and the convexity of the M-deepest region will still ensure that this uniquely defined M-median has indeed maximal M-depth.

The following result shows that, for $\rho \in \mathcal{C}_{\text{aff}}$ (a restriction that is required only for Part (i) of the result), the halfspace M-depth MD^ρ is a *statistical depth function*, in the axiomatic sense of Zuo and Serfling (2000).

Theorem 6 *Fix $\rho \in \mathcal{C}_{\text{aff}}$ and $P \in \mathcal{P}_d^\rho$. Then, $MD^\rho(\mathbf{z}, P)$ satisfies the following properties: (i) (affine invariance:) for any invertible $d \times d$ matrix $\mathbf{A}$ and d-vector $\mathbf{b}$, $MD^\rho(\mathbf{A}\mathbf{z} + \mathbf{b}, P_{\mathbf{A},\mathbf{b}}) = MD^\rho(\mathbf{z}, P)$, where $P_{\mathbf{A},\mathbf{b}}$ was defined in Theorem 2; (ii) (maximality at the center:) if P is centrally symmetric about $\boldsymbol{\theta}_*$ (i.e., $P[\boldsymbol{\theta}_* + B] = P[\boldsymbol{\theta}_* - B]$ for any d-Borel set B), then $MD^\rho(\boldsymbol{\theta}_*, P) \geq MD^\rho(\mathbf{z}, P)$ for any d-vector $\mathbf{z}$; (iii) (monotonicity along rays:) if $\boldsymbol{\theta}$ has maximum M-depth with respect to P, then, for any $\mathbf{u} \in \mathcal{S}^{d-1}$, $r \mapsto MD^\rho(\boldsymbol{\theta} + r\mathbf{u}, P)$ is monotone non-increasing in $r (\geq 0)$ and $MD^\rho(\boldsymbol{\theta} + r\mathbf{u}, P) = 0$ for any $r > r_{\mathbf{u}}(P) := \sup\{r > 0 : \boldsymbol{\theta} + r\mathbf{u} \in C_P\}$ $(\in (0, +\infty])$; (iv) (vanishing at infinity:) as $\|\mathbf{z}\| \to \infty$, $MD^\rho(\mathbf{z}, P) \to 0$.*

As mentioned above, MD^ρ reduces to the Tukey depth for $\rho(t) = |t|$. For any other ρ function, the depth MD^ρ is, to the authors' best knowledge, original. In particular, the (*halfspace*) *expectile depth* obtained for $\rho(t) = t^2$ has not been considered so far. While, as already mentioned, competing concepts of multivariate expectiles would provide alternative concepts of expectile depth (through the corresponding expectile regions as in Definition 3), the following result hints that our construction is a (if not the) most natural one.

Theorem 7 *Fix $\rho \in \mathcal{C}$ and $P \in \mathcal{P}_d^\rho$. Then, for any $\mathbf{z} \in \mathbb{R}^d$,*

$$MD^\rho(\mathbf{z}, P) = \inf_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{\mathrm{E}[|\psi_-(\mathbf{u}'(\mathbf{Z}-\mathbf{z}))|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) \leq 0]]}{\mathrm{E}[|\psi_-(\mathbf{u}'(\mathbf{Z}-\mathbf{z}))|]}, \tag{6}$$

where ψ_- is the left-derivative of ρ and where $\mathbf{Z}$ has distribution P.

For $\rho(t) = |t|$, we have $\psi_-(t) = \mathbb{I}[t > 0] - \mathbb{I}[t \leq 0]$, so that this result confirms that our M-depth then coincides with the halfspace Tukey depth

$$HD(\mathbf{z}, P) = \inf_{\mathbf{u} \in \mathcal{S}^{d-1}} \mathbb{P}[\mathbf{u}'\mathbf{Z} \leq \mathbf{u}'\mathbf{z}],$$

that records the most extreme (lower-)outlyingness of $\mathbf{u}'\mathbf{z}$ with respect to the distribution of $\mathbf{u}'\mathbf{Z}$. The M-depth in (6) can be interpreted in the exact same way but replaces standard quantile outlyingness with M-quantile outlyingness; see the last paragraph of Section 2. For $\rho(t) = t^2$, Theorem 7 states that our expectile depth can be equivalently defined as

$$ED(\mathbf{z}, P) = \inf_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) \leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|]}. \tag{7}$$

Of course, we could similarly consider the continuum of halfspace M-depths associated with the Huber loss functions ρ_c in (4) or the one made of L_r-depths associated with $\rho(t) = |t|^r$ ($r \geq 1$). The M-depth formulation provides a framework that allows deriving results that apply generically for Tukey depth, expectile depth, or these Huber or L_r-depths (in Section S.2 of the supplement, we provide in particular general asymptotic results for M-depth, M-deepest points and M-depth regions). M-depths are of interest on their own, as they may allow achieving a nice balance between robustness and efficiency (this is in line with the fact that Tukey depth is maximized at a multivariate median, whereas, as we will see in the next section, expectile depth is maximized at the mean vector). Rather than pursuing the investigation of M-depths, however, we will mainly focus on expectile depth below, as our work was mainly motivated by expectiles and multiple-output expectile regression.

5. Halfspace expectile depth

Below, we derive further properties of (halfspace) expectile depth that show why this particular M-depth should be particularly appealing to practitioners. Throughout, we write $\mathcal{P}_d$ for the collection $\mathcal{P}_d^\rho$ of probability measures over $\mathbb{R}^d$ for which expectile depth is well-defined, that is, the one associated with $\rho(t) = t^2$. Clearly, $\mathcal{P}_d$ collects the probability measures that (i) do not give P-probability one to any hyperplane of $\mathbb{R}^d$ and that (ii) have finite first-order moments. Note that this moment assumption is required even for $d = 1$; as for (i), it only rules out distributions that are actually over a lower-dimensional Euclidean space.

5.1. Distinctive properties of expectile depth

For halfspace Tukey depth, which is an L_1-concept, the deepest point is not always unique and its unique representative, namely the Tukey median, is a multivariate extension of the univariate median. Also, the depth of the Tukey median may depend on the underlying probability measure P. Our expectile depth, that is rather of an L_2-nature, is much different in these respects.

Theorem 8 *For any $P \in \mathcal{P}_d$, the expectile depth $ED(\mathbf{z}, P)$ is uniquely maximised at $\mathbf{z} = \boldsymbol{\mu}_P := \mathrm{E}[\mathbf{Z}]$ (where $\mathbf{Z}$ is a random d-vector with distribution P) and the corresponding maximum depth is $ED(\boldsymbol{\mu}_P, P) = 1/2$.*

Expectile depth regions therefore always provide nested regions around the mean vector $\boldsymbol{\mu}_P$, which should be appealing to practitioners. Also, the fact that the corresponding maximal depth is always $1/2$ will allow practitioners to better interpret what it means that another location would have expectile depth equal to, say, $1/4$ (this would be, irrespective of the distribution, half as deep as any location can get). In contrast, there is no way to evaluate the "relative" depth of a location with Tukey depth $1/4$ without evaluating the depth of the Tukey median, which, at least in moderate to high dimensions, is notoriously difficult. Moreover, since the maximal expectile depth is $1/2$ for any P, a natural affine-invariant test for $\mathcal{H}_0 : \boldsymbol{\mu}_P = \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0 \in \mathbb{R}^d$ is fixed, is the one rejecting $\mathcal{H}_0$ for large values of $T_n := (1/2) - ED(\boldsymbol{\mu}_0, P_n)$, where P_n is the empirical probability measure associated with the sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ at hand. Due to the relation between expectile depth and the mean vector, this can be regarded as a nonparametric version of the Hotelling test.

We turn to another distinctive aspect of expectile depth. Theorem 6(iii) shows that halfspace M-depth decreases monotonically when one moves away from a deepest point along any ray. This decrease, however, may fail to be strict (in the sample case, for instance, the halfspace Tukey depth is piecewise constant, hence will fail to be strictly decreasing). In contrast, expectile depth always offers a strict decrease (until, of course, the minimal depth value zero is reached, if it is).

Theorem 9 *Fix $P \in \mathcal{P}_d$ and $\mathbf{u} \in \mathcal{S}^{d-1}$. Let $r_{\mathbf{u}}(P) = \sup\{r > 0 : \boldsymbol{\mu}_P + r\mathbf{u} \in C_P\} (\in (0, +\infty])$. Then, $r \mapsto ED(\boldsymbol{\mu}_P + r\mathbf{u}, P)$ is monotone strictly decreasing in $[0, r_{\mathbf{u}}(P)]$ and $ED(\boldsymbol{\mu}_P + r\mathbf{u}, P) = 0$ for $r \geq r_{\mathbf{u}}(P)$.*

This also has practical advantages in several applications of depth, where ties in the depth values of several locations are problematic (if $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are randomly sampled from a distribution P admitting a density, then the corresponding depths $ED(\mathbf{Z}_i, P_n)$ will be pairwise different with probability one). An application where it is desirable to avoid ties is supervised classification: for instance, the *max-depth classifiers* from Ghosh and Chaudhuri (2005) (see also Li et al., 2012) classify $\mathbf{z}$ as arising from P_1 rather that P_2 if $\mathbf{z}$ is deeper with respect to P_1 than it is with respect to P_2, but obviously ties will lead to an unpleasant randomization.

Our M-depths are upper semicontinuous functions of $\mathbf{z}$; see Section 4. However, continuity does not hold in general (in particular, the piecewise constant nature of the Tukey depth for empirical distributions rules out continuity in the sample case). In contrast, expectile depth is smooth even in the sample case, which also should be appealing to practitioners who typically find it unpleasant that a small location change has a strong impact on the corresponding depth.

Theorem 10 *Fix $P \in \mathcal{P}_d$. Then, (i) $\mathbf{z} \mapsto ED(\mathbf{z}, P)$ is uniformly continuous over $\mathbb{R}^d$; (ii) for $d = 1$, $z \mapsto ED(z, P)$ is left- and right-differentiable over $\mathbb{R}$; (iii) for $d \geq 2$, if P is smooth in a neighbourhood $\mathcal{N}$ of $\mathbf{z}_0$ (meaning that for any $\mathbf{z} \in \mathcal{N}$, any hyperplane containing $\mathbf{z}$ has P-probability zero), then $\mathbf{z} \mapsto ED(\mathbf{z}, P)$ admits directional derivatives at $\mathbf{z}_0$ in all directions.*

Before going to another, key, distinctive property of expectile depth in Section 5.3, we illustrate the theoretical results above on the basis of some examples.

5.2. Some examples

We start with both following univariate examples. If P is the uniform measure over the interval $\mathcal{I} = [0, 1]$, then

$$ED(z, P) = \frac{\min(z^2, (1-z)^2)}{z^2 + (1-z)^2} \mathbb{I}[z \in \mathcal{I}] \quad \text{and} \quad HD(z, P) = \min(z, 1-z)\mathbb{I}[z \in \mathcal{I}], \tag{8}$$

whereas if P is the uniform over the pair $\{0, 1\}$, then

$$ED(z, P) = \min(z, 1-z)\mathbb{I}[z \in \mathcal{I}] \quad \text{and} \quad HD(z, P) = \frac{1}{2}\mathbb{I}[z \in \mathcal{I}]; \tag{9}$$

see Figure 2. This illustrates uniform continuity of expectile depth, as well as left- and right-differentiability. Although both distributions are smooth in a neighborhood of $z_0 = 1/2$, plain differentiability does not hold at z_0, which results from the non-uniqueness of the corresponding minimal direction; see Demyanov (2009). For the sake of comparison, the figure also plots the Tukey depth $HD(z, P)$ and the zonoid depth $ZD(z, P)$ from Koshevoy and Mosler (1997). Comparison with the latter depth is natural as it is also maximized at $\boldsymbol{\mu}_P$, hence has an L_2-flavor. Both the Tukey and zonoid depths are less smooth than the expectile one, and in particular, the zonoid depth is not continuous for the discrete uniform. Also, the zonoid depth region $\{z \in \mathbb{R} : ZD(z, P) \geq \alpha\}$ in the left panel is the (L_1) interquantile interval $[\frac{\alpha}{2}, 1 - \frac{\alpha}{2}]$, which is not so natural for a depth of an L_2-nature. In contrast, for any P, the level-α expectile depth region is the interexpectile interval $[\theta^\rho_\alpha(P), \theta^\rho_{1-\alpha}(P)]$, with $\rho(t) = t^2$, which reflects, for any α (rather than for the deepest level only), the L_2-nature of expectile depth.

Before considering multivariate examples, we state a further distinctive property of expectile depth, that is specific to the multivariate case. A direction $\mathbf{u}_0$ is said to be *minimal* for the depth $MD^\rho(\mathbf{z}, P)$ if it achieves the infimum in (6). For the halfspace Tukey depth, such a minimal direction does not always exist (consider, e.g., $\mathbf{z} = (1, 0)' \in \mathbb{R}^2$ and $P = \frac{1}{2}P_1 + \frac{1}{2}P_2$, where P_1 is the bivariate standard normal distribution and P_2 is the Dirac distribution at $(1, 1)'$). In contrast, the continuity, for any $P \in \mathcal{P}_d$, of the function whose infimum is considered over $\mathcal{S}^{d-1}$ in (7) (see Lemma S.10(i)) and the compactness of $\mathcal{S}^{d-1}$ imply that

$$ED(\mathbf{z}, P) = \min_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z} - \mathbf{z}) \leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|]}, \tag{10}$$

so that a minimal direction always exists for expectile depth (in the bivariate mixture example above, $\mathbf{u}_0 = (-1, 0)'$ is a minimal direction for $ED(\mathbf{z}, P)$).

Consider then the case where $P(\in \mathcal{P}_d)$ is the distribution of $\mathbf{Z} = \mathbf{A}\mathbf{Y} + \boldsymbol{\mu}$, where $\mathbf{A}$ is an invertible $d \times d$ matrix, $\boldsymbol{\mu}$ is a d-vector and $\mathbf{Y} = (Y_1, \ldots, Y_d)'$ is a spherically symmetric random vector, meaning that the distribution of $\mathbf{O}\mathbf{Y}$ does not depend on the $d \times d$ orthogonal matrix $\mathbf{O}$. In other words, P is elliptical with mean vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$. In the standard case where $\mathbf{A} = \mathbf{I}_d$ (the d-dimensional identity matrix) and $\boldsymbol{\mu} = 0$,

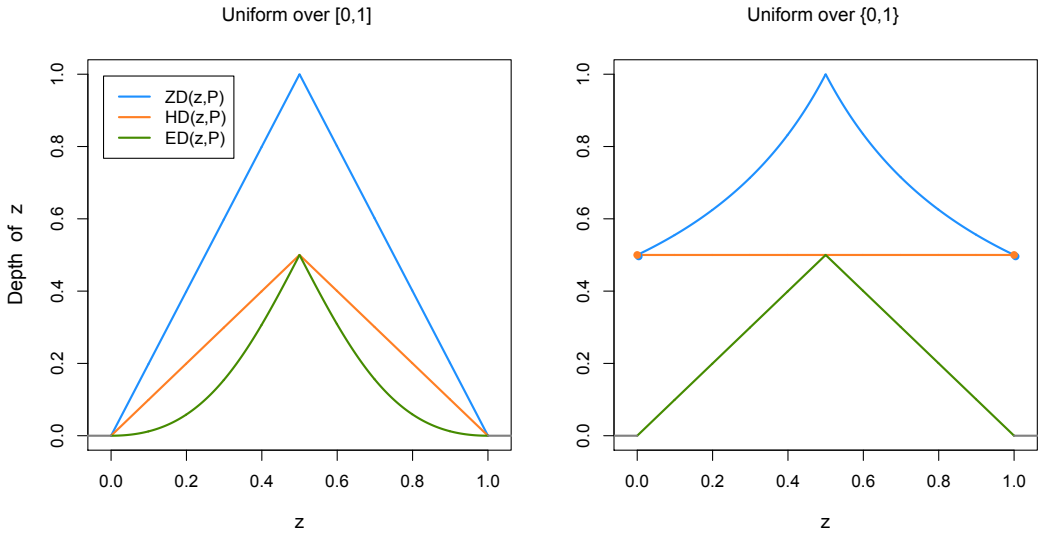

Fig. 2: Plots, as functions of z, of the zonoid depth $ZD(z,P)$ (blue), of the halfspace Tukey depth $HD(z,P)$ (orange) and of the expectile depth $ED(z,P)$ (green), when P is the uniform over the interval $[0,1]$ (left) and the uniform over the pair $\{0,1\}$ (right). For both probability measures, all depth functions take value zero outside $[0,1]$.

Theorem 1(iv) provides

$$ED(\mathbf{z},P) = \min_{\mathbf{u}\in\mathcal{S}^{d-1}} \frac{\mathrm{E}[|Y_1-\mathbf{u}'\mathbf{z}|\mathbb{I}[Y_1\le\mathbf{u}'\mathbf{z}]]}{\mathrm{E}[|Y_1-\mathbf{u}'\mathbf{z}|]}$$

$$= -\frac{\mathrm{E}[(Y_1+\|\mathbf{z}\|)\mathbb{I}[Y_1\le-\|\mathbf{z}\|]]}{\mathrm{E}[|Y_1+\|\mathbf{z}\||]} =: g(\|\mathbf{z}\|),$$

so that, for arbitrary $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, affine invariance entails that $ED(\mathbf{z},P)=g(\|\mathbf{z}\|_{\mu,\Sigma})$, with $\|\mathbf{z}\|^2_{\mu,\Sigma}:=(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})$. Expectile depth regions are thus concentric ellipsoids that, under absolute continuity of P, coincide with equidensity contours. The function g depends on the distribution of $\mathbf{Y}$: if $\mathbf{Y}$ is d-variate standard normal, then it is easy to check that $g(r)=\{1-1/(2\phi(r)/r+2\Phi(r)-1)\}/2$, where ϕ and Φ denote the probability density function and cumulative distribution function of the univariate standard normal distribution, respectively. If $\mathbf{Y}$ is uniform over the unit ball $B^d:=\{\mathbf{z}\in\mathbb{R}^d:\|\mathbf{z}\|\le 1\}$ or on the unit sphere $\mathcal{S}^{d-1}$, then one can show that

$$g(r)=\omega_d(r):=\left(\frac{1}{2}-\frac{\sqrt{\pi}r(1-r^2)^{-(d+1)/2}\Gamma(\frac{d+3}{2})}{2\Gamma(\frac{d+2}{2})(1+(d+1)r^2{}_2F_1(1,\frac{d+2}{2};\frac{3}{2};r^2))}\right)\mathbb{I}[r\le 1]$$

and $g(r)=\omega_{d-2}(r)$, respectively, where Γ is the Euler Gamma function and ${}_2F_1$ is the hypergeometric function. From affine invariance, these expressions agree with those obtained for $d=1$ in (8) and (9), respectively. In all cases considered, thus, the function g is continuous and monotone strictly decreasing on its support, which illustrates the theoretical results of Section 5.1.

Our last multivariate example is a non-elliptical one. Consider the probability measure $P_\alpha(\in \mathcal{P}_d)$ having independent standard (symmetric) α-stable marginals, with $1 < \alpha \leq 2$. If $\mathbf{Z} = (Z_1, \ldots, Z_d)'$ has distribution P_α, then $\mathbf{u}'\mathbf{Z}$ is equal in distribution to $\|\mathbf{u}\|_\alpha Z_1$, where we let $\|\mathbf{x}\|_\alpha^\alpha := \sum_{j=1}^d |x_j|^\alpha$. Thus, (10) provides

$$ED(\mathbf{z}, P_\alpha) = \min_{\mathbf{v} \in \mathcal{S}_\alpha^{d-1}} \frac{\mathrm{E}[|Z_1 - \mathbf{v}'\mathbf{z}|\mathbb{I}[Z_1 \leq \mathbf{v}'\mathbf{z}]]}{\mathrm{E}[|Z_1 - \mathbf{v}'\mathbf{z}|]},$$

where $\mathcal{S}_\alpha^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\|_\alpha = 1\}$ is the unit L_α-sphere. Theorem 1(iv) implies that the minimum is achieved when $\mathbf{v}'\mathbf{z}$ takes its minimal value $-\|\mathbf{z}\|_\beta$, where $\beta = \alpha/(\alpha - 1)$ is the conjugate exponent to α; see Lemma A.1 in Chen and Tyler (2004). Denoting as f_α the marginal density of P_α, this yields

$$ED(\mathbf{z}, P_\alpha) = -\frac{\mathrm{E}[(Z_1 + \|\mathbf{z}\|_\beta)\mathbb{I}[Z_1 \leq -\|\mathbf{z}\|_\beta]]}{\mathrm{E}[|Z_1 + \|\mathbf{z}\|_\beta|]} = -\frac{\int_{-\infty}^0 x f_\alpha(x - \|\mathbf{z}\|_\beta)\, dx}{\int_{-\infty}^\infty |x| f_\alpha(x - \|\mathbf{z}\|_\beta)\, dx},$$

which shows that expectile depth regions are concentric L_β-balls. For $\alpha = 2$, these results agree with those obtained in the Gaussian case above.

5.3. Computational aspects

The sample M-depth regions $R_\alpha^\rho(P_n)$ can be computed by replacing the intersection in Definition 2 with an intersection over finitely many directions $\mathbf{u}_\ell$, $\ell = 1, \ldots, L$, with L large; see Section 3. Many applications, however, do not require computing depth regions but rather the depth of a given location $\mathbf{z}$ only. An important example is supervised classification through the max-depth approach; see Ghosh and Chaudhuri (2005) or Li et al. (2012). While the halfspace M-depth of $\mathbf{z}$ can in principle be obtained from the depth regions (recall that $MD^\rho(\mathbf{z}, P) = \sup\{\alpha > 0 : \mathbf{z} \in R_\alpha^\rho(P)\}$), it will be much more efficient in such applications to compute $MD^\rho(\mathbf{z}, P)$ through the alternative expression in (6). Recall that, for the halfspace Tukey and halfspace expectile depths, this alternative expression reduces to

$$HD(\mathbf{z}, P) = \inf_{\mathbf{u} \in \mathcal{S}^{d-1}} h_\mathbf{z}(\mathbf{u}) \quad \text{and} \quad ED(\mathbf{z}, P) = \min_{\mathbf{u} \in \mathcal{S}^{d-1}} e_\mathbf{z}(\mathbf{u}),$$

respectively, where we let

$$h_\mathbf{z}(\mathbf{u}) := \mathbb{P}[\mathbf{u}'\mathbf{Z} \leq \mathbf{u}'\mathbf{z}] \quad \text{and} \quad e_\mathbf{z}(\mathbf{u}) := \frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z} - \mathbf{z}) \leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|]}. \tag{11}$$

The function $\mathbf{u} \mapsto h_\mathbf{z}(\mathbf{u})$ that is to be minimized to compute Tukey depth does not allow using standard algorithms such as Newton-Raphson-type methods, as such iterative methods may converge to one of the many *local* minima of this function; see Figure 3(a) (another reason why Newton-Raphson methods are ruled out is that, in the sample case, $\mathbf{u} \mapsto h_\mathbf{z}(\mathbf{u})$ is a piecewise constant, hence non-smooth, function). Computation of Tukey depth is actually a very difficult task, that generated a vast literature and can be performed in small to moderate dimensions only.

Remarkably, a further, quite unexpected, distinctive property of expectile depth opens the door to Newton-Raphson-based computation of this depth. More specifically, the following result ensures that the function $\mathbf{u} \mapsto e_\mathbf{z}(\mathbf{u})$ that is to be minimized to compute expectile depth is not only smooth but also has no local-but-not-global minimizers.

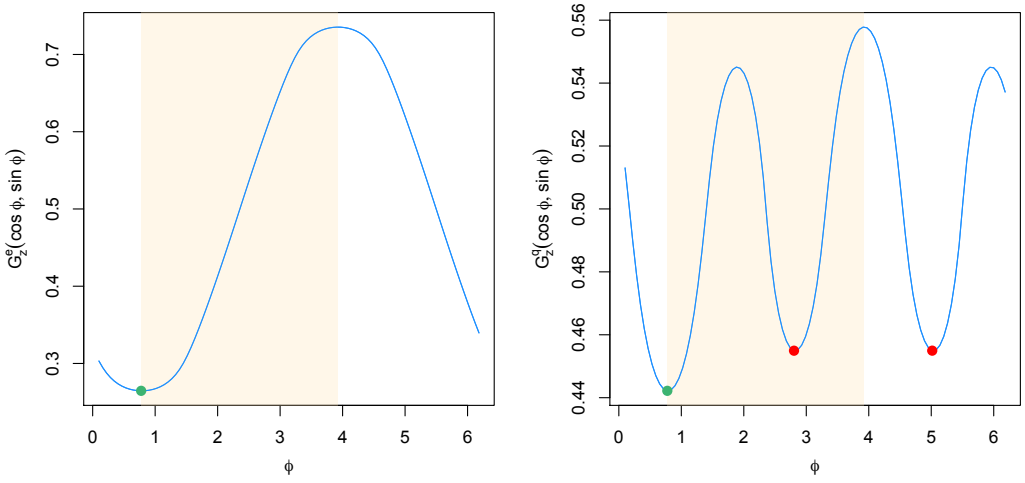

Fig. 3: Plots of the Tukey depth outlyingness (left) and expectile depth outlyingness (right) in (11), with $\mathbf{z} = \left(\frac{3}{4}, \frac{3}{4}\right)'$ and P the probability measure over $\mathbb{R}^2$ whose marginals are independent exponentials with mean one. Global minimizers providing the respective depths are marked in green, whereas local-but-not-global minimizers preventing the use of Newton-Raphson-type methods are marked in red. The shaded area shows the range considered in Theorem 11.

Theorem 11 *Fix $P \in \mathcal{P}_d$ and $\mathbf{z} \in \mathbb{R}^d$ such that $ED(\mathbf{z}, P) > 0$. Assume that $P[\Pi \setminus \{\mathbf{z}\}] = 0$ for any hyperplane Π containing $\mathbf{z}$. Fix a great circle $\mathcal{G}$ of $\mathcal{S}^{d-1}$ and let $\mathbf{u}_0$ be an arbitrary minimizer of $e_{\mathbf{z}}(\cdot)$ on $\mathcal{G}$. Let $\mathbf{u}_t$, $t \in [0, \pi]$, be a path on $\mathcal{G}$ from $\mathbf{u}_0$ to $-\mathbf{u}_0$. Then, there exist t_a, t_b with $0 \leq t_a \leq t_b \leq \pi$ such that $t \mapsto e_{\mathbf{z}}(\mathbf{u}_t)$ is constant over $[0, t_a]$, admits a strictly positive derivative at any $t \in (t_a, t_b)$ (hence is strictly increasing over $[t_a, t_b]$), and is constant over $[t_b, \pi]$.*

The result is illustrated in Figure 3, that draws, for $\mathbf{z} = \left(\frac{3}{4}, \frac{3}{4}\right)'$ and for P the probability measure over $\mathbb{R}^2$ whose marginals are independent exponentials with mean one, the plots of $\theta \mapsto h_{\mathbf{z}}((\cos\theta, \sin\theta)')$ and $\theta \mapsto e_{\mathbf{z}}((\cos\theta, \sin\theta)')$ over $[0, 2\pi]$. Clearly, the figure also shows that the result may fail for Tukey depth. Jointly with the fact that, in the sample case, $t \mapsto e_{\mathbf{z}}(\mathbf{u}_t)$ will be left- and right-differentiable, Theorem 11 opens the door to fast computation of expectile depth, also in high dimensions, through Newton-Raphson-type methods. An algorithm evaluating expectile depth in this way will be developed in a later, more computational, work.

6. Multiple-output expectile regression

We now consider the multiple-output regression framework involving a d-vector $\mathbf{Y}$ of responses and a p-vector $\mathbf{X}$ of (random) covariates. Denoting as $P_{\mathbf{x}}$ the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$, our interest lies in the conditional M-quantile halfspaces and

regions

$$H^{\rho}_{\alpha,\mathbf{u},\mathbf{x}} = H^{\rho}_{\alpha,\mathbf{u}}(P_{\mathbf{x}}) \quad \text{and} \quad R^{\rho}_{\alpha,\mathbf{x}} = R^{\rho}_{\alpha}(P_{\mathbf{x}}),$$

with $\alpha \in (0,1)$ and $\mathbf{u} \in \mathcal{S}^{d-1}$. If a random sample $(\mathbf{X}_1,\mathbf{Y}_1),\ldots,(\mathbf{X}_n,\mathbf{Y}_n)$ is available, then one may consider the estimates

$$\hat{H}^{\rho(n)}_{\alpha,\mathbf{u},\mathbf{x}} := \left\{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'\mathbf{y} \geq \theta^{\rho(n)}_{\alpha,\mathbf{u},\mathbf{x}}\right\} \quad \text{and} \quad \hat{R}^{\rho(n)}_{\alpha,\mathbf{x}} := \cap_{\mathbf{u}\in\mathcal{S}^{d-1}} \hat{H}^{\rho(n)}_{\alpha,\mathbf{u},\mathbf{x}}, \tag{12}$$

where $\theta^{\rho(n)}_{\alpha,\mathbf{u},\mathbf{x}}$ is the estimate of $\theta^{\rho}_{\alpha}(P^{\mathbf{u}'\mathbf{Y}|[\mathbf{X}=\mathbf{x}]})$ obtained from a single-output, linear or nonparametric, regression using the responses $\mathbf{u}'\mathbf{Y}_1,\ldots,\mathbf{u}'\mathbf{Y}_n$ and covariates $\mathbf{X}_1,\ldots,\mathbf{X}_n$ (in the examples below, that focus on $d=2$, the intersection in (12) was replaced with an intersection over $L=200$ equispaced directions in $\mathcal{S}^1$). For expectiles, single-output linear and nonparametric regression can respectively be performed via the functions `expectreg.ls` and `expectreg.boost` from the R package *expectreg* (nonparametric regression here is thus based on the expectile boosting approach from Sobotka and Kneib, 2012). Multiple-output quantile regression can be achieved in the same way, by performing single-output linear quantile regression (via the function `rq` in the R package *quantreg*) or single-output nonparametric quantile regression (via, e.g., the function `cobs` in the R package *cobs*, which relies on the popular quantile smoothing spline approach). Whenever we use `expectreg.boost` and `cobs` below, it is with the corresponding default automatic selection of smoothing parameters.

6.1. Simulated data illustration

To illustrate these regression methods on simulated data, we generated a random sample of size $n=300$ from the heteroscedastic linear regression model

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = 4\begin{pmatrix} X \\ X \end{pmatrix} + \sqrt{\frac{X}{3}}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \tag{13}$$

where the covariate X is uniform over $[0,1]$, $\varepsilon_1+1, \varepsilon_2+1$ are exponential with mean one, and $X, \varepsilon_1, \varepsilon_2$ are mutually independent. For several orders α and several values of x, we evaluated the conditional quantile and expectile regions $\hat{R}^{\rho(n)}_{\alpha,x}$, in each case both from the corresponding linear and nonparametric regression methods above. The resulting contours are provided in Figure 4. While both expectile and quantile methods capture trend and heteroscedasticity, expectiles dominate quantiles in many respects: (i) unlike quantiles, expectiles provide very similar linear and nonparametric regression fits (which is desirable since the model is linear). (ii) Expectiles yield smoother contours than quantiles. (iii) Inner expectile contours, that do not have the same location as their quantile counterparts, are easier to interpret as they relate to conditional means of the marginal responses (inner quantile contours refer to the Tukey median, which is not directly related to marginal medians). (iv) Last but not least, unlike expectile contours, several quantile contours associated with a common value of x do cross (see the bottom right panel of Figure 4), which obviously is incompatible with what occurs at the population level.

6.2. Real data illustration

We now conduct multiple-output expectile (and quantile) regression to investigate risk factors for early childhood malnutrition in India. Prior studies typically focused on children's

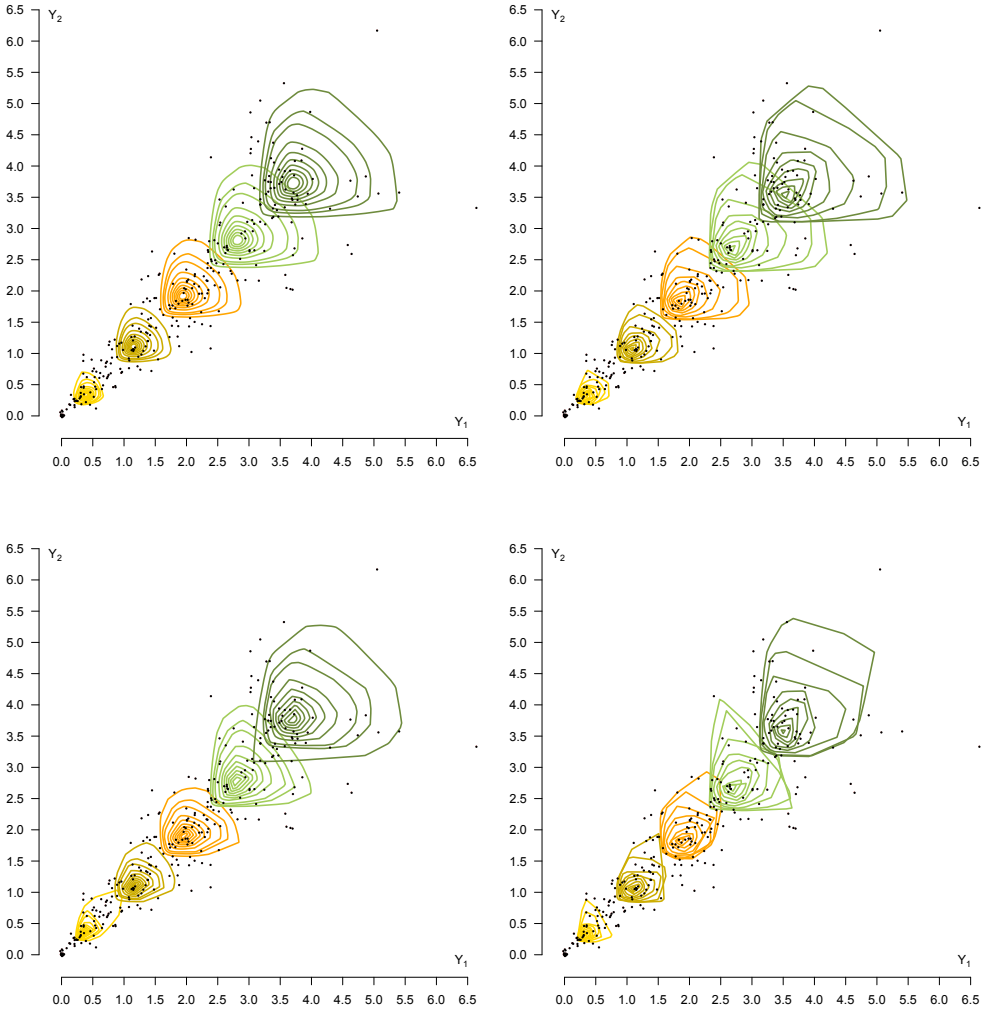

Fig. 4: (Left:) conditional expectile contours $\partial \hat{R}^{\rho(n)}_{\alpha,x}$, for $\alpha \in \{.01, .03, .05, .10, .15, \ldots, .40\}$ and for values of x that are the 10% (yellow), 30% (brown), 50% (orange), 70% (light green) and 90% (dark green) empirical quantiles of $X_1, \ldots, X_n$, obtained by applying a linear (top) or nonparametric (bottom) regression method to a random sample of size $n = 300$ from the linear regression model in (13). (Right:) conditional quantile contours associated with the same values of α (but .01) and the same values of x. Again, both linear regression (top) and nonparametric regression (bottom) are considered; see Section 6.1 for details. Bivariate responses (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, are shown in black.

height as an indicator of nutritional status (Koenker, 2011; Fenske et al., 2011). Given that the prevalence of underweighted children in India is among the highest worldwide, we consider here determinants of children's *weight* (Y_1; in kilograms) and *height* (Y_2; in centimeters) simultaneously. We use a selected sample of 37,623 observations, coming from the 2005/2006 Demographic and Health Survey (DHS) conducted in India for children under five years of age. Since a thorough case study is beyond the scope of this paper, we restrict to assessing the separate effects of the following covariates on the response $\mathbf{Y} = (Y_1, Y_2)'$: (a) the child's age (in months) and (b) the mother's Body Mass Index (defined as BMI := weight/height^2, in kilograms/meters^2). Koenker (2011) investigated the additive effects of these covariates on low levels of the single response *height* through a quantile regression with small α.

Figure 5 plots, for each of the two covariates, nonparametric conditional expectile and quantile contours associated with the extreme levels $\alpha \in \{.005, .01\}$ and several covariate values x (while growth curves are known to be highly nonlinear, the dependence on both other covariates might be linear, so that we further provide linear fits in Figure 6). At the very large sample size considered, quantile contours are not subject to the lack of smoothness nor to the crossing issue they exhibited for the simulated data above, but they would be subject to these for small to moderate sample sizes (see, e.g., the real data example treated in Section S.3 of the supplement, that provides, with sample size $n = 3{,}134$, quantile contours that lack smoothness and touch each other). For the age covariate, it is seen that both expectiles and quantiles capture well the trend (and heteroscedasticity), but it should be noted that expectile and quantile contours are not centered at the same locations—recall from Section 6.2 that the location of expectile contours is easier to interpret than their quantile counterparts. Also, note that quantile contours tend to be more elliptical than the expectile ones, hence are less flexible. In particular, expectiles reveal that mothers with a BMI above median may lead to overweigthed tall children, but not to overweighted short ones; at such BMI levels, both green expectile contours indeed show (for nonparametric and linear fits) an asymmetry to obesity in the upper-right direction, but not in the lower-left one (this is arguably not related to malnutrition, but it is still pointing to some risk factor). The more elliptical, hence more symmetric, quantile contours fail to reveal such subtle risks, which is due to their smaller sensitivity to extreme observations (compared to expectiles). Like the simulated data of the previous section, the present real data therefore shows that multiple-output expectile regression may reveal aspects of the dependence of the response on the covariates to which quantile regression will remain blind. This is particularly the case in risk-oriented applications, where robustness is not a desirable property.

Acknowledgements

The authors acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

References

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

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

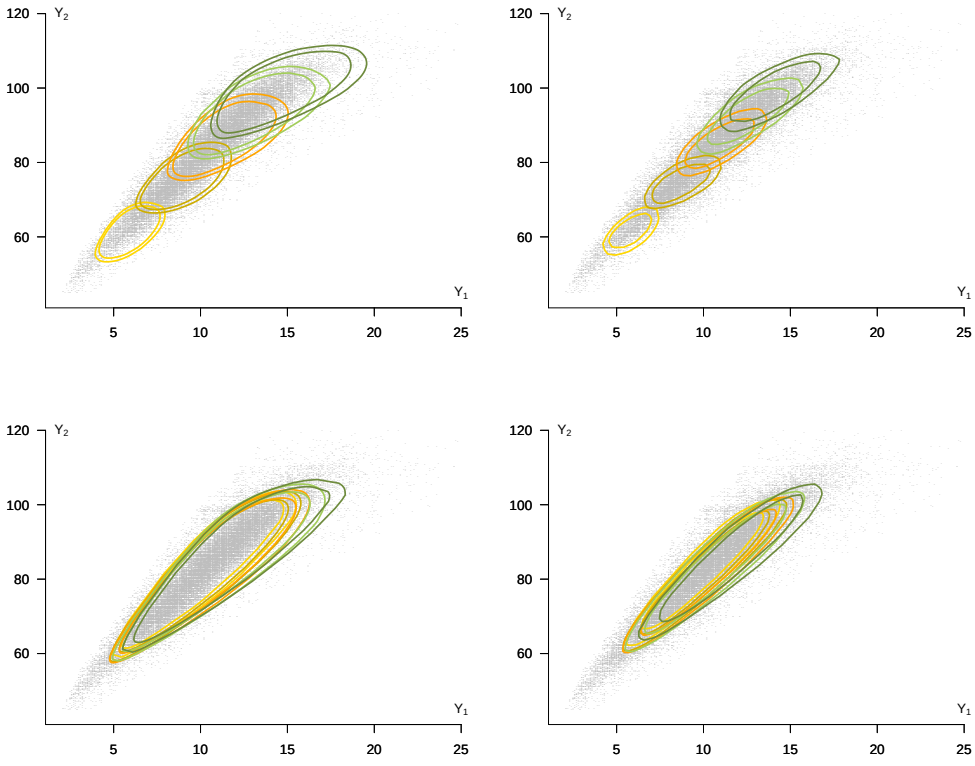

Fig. 5: (Left:) conditional expectile contours $\partial \hat{R}^{\rho(n)}_{\alpha,x}$ for the joint children's weights (Y_1) and heights (Y_2), obtained from *nonparametric* regression as explained at the beginning of Section 6, conditional on their age (top) and on the mother's BMI (bottom) for $\alpha \in \{.005, .01\}$ and for values of x that are the 5% (yellow), 25% (brown), 50% (orange), 75% (light green) and 95% (dark green) empirical quantiles of the covariate. (Right:) the corresponding quantile contours; see Section 6.2 for details. Bivariate responses (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, are shown in gray.

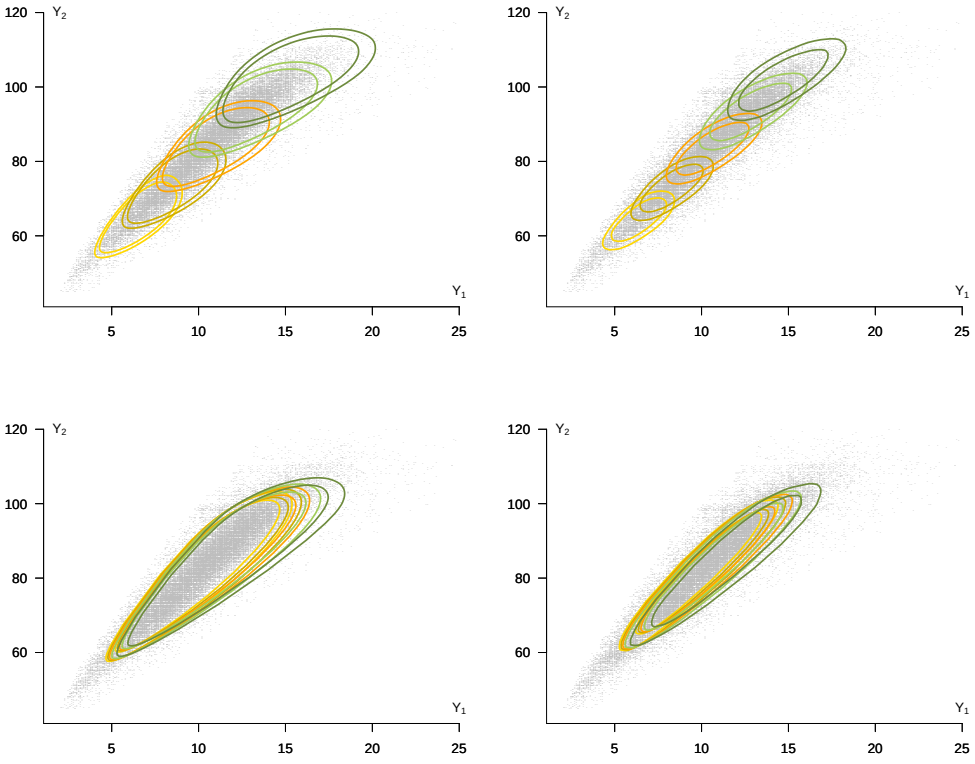

Fig. 6: (Left:) conditional expectile contours $\partial \hat{R}^{\rho(n)}_{\alpha,x}$ for the joint children's weights (Y_1) and heights (Y_2), obtained from *linear* regression, conditional on their age (top) and on the mother's BMI (bottom), for $\alpha \in \{.005, .01\}$ and for values of x that are the 5% (yellow), 25% (brown), 50% (orange), 75% (light green) and 95% (dark green) empirical quantiles of the covariate. (Right:) the corresponding quantile contours; see Section 6.2 for details. Bivariate responses (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, are shown in gray.

Carlier, G., Chernozhukov, V. and Galichon, A. (2016) Vector quantile regression: An optimal transport approach. *Ann. Statist.*, **44**, 1165–1192.

— (2017) Vector quantile regression beyond the specified case. *J. Multivariate Anal.*, **161**, 96–102.

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

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

Chavas, J.-P. (2018) On multivariate quantile regression analysis. *Stat. Methods Appl.*, **27**, 365–384.

Chen, Z. (1996) Conditional L_p-quantiles and their application to testing of symmetry in non-parametric regression. *Statist. Probab. Lett.*, **29**, 107–115.

Chen, Z. and Tyler, D. E. (2004) On the behavior of Tukey's depth and median under symmetric stable distributions. *J. Statist. Plann. Inference*, **122**, 111–124.

Cheng, Y. and De Gooijer, J. (2007) On the uth geometric conditional quantile. *J. Statist. Plann. Inference*, **137**, 1914–1930.

Cousin, A. and Di Bernardino, E. (2013) On multivariate extensions of value-at-risk. *J. Multivariate Anal.*, **119**, 32–46.

— (2014) On multivariate extensions of conditional-tail-expectation. *Insurance Math. Econom.*, **55**, 272–282.

Daouia, A., Girard, S. and Stupfler, G. (2018) Estimation of tail risk based on extreme expectiles. *J. Roy. Statist. Soc. Ser. B*, **80**, 263–292.

— (2019) Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli*, **25**, 264–309.

De Rossi, G. and Harvey, H. (2009) Quantiles, expectiles and splines. *J. Econometrics*, **152**, 179–185.

Demyanov, V. F. (2009) Minimax: directional differentiability. In *Encyclopedia of Optimization* (eds. C. A. Floudas and P. M. Pardalos), 2075–2079. Springer.

Eilers, P. (2013) Discussion: The beauty of expectiles. *Stat. Model.*, **13**, 317–322.

Embrechts, P. and Hofert, M. (2014) Statistics and quantitative risk management for banking and insurance. *Annu. Rev. Stat. Appl.*, **1**, 493–514.

Fenske, N., Kneib, T. and Hothorn, T. (2011) Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *J. Amer. Statist. Assoc.*, **106**, 494–510.

Ghosh, A. K. and Chaudhuri, P. (2005) On maximum depth and related classifiers. *Scand. J. Statist.*, **32**, 327–350.

Girard, S. and Stupfler, G. (2017) Intriguing properties of extreme geometric quantiles. *REVSTAT*, **15**, 107–139.

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

Hallin, M., Paindaveine, D. and Šiman, M. (2010) Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth (with discussion). *Ann. Statist.*, **38**, 635–669.

Herrmann, K., Hofert, M. and Mailhot, M. (2018) Multivariate geometric expectiles. *Scand. Actuar. J.*, **2018**, 629–659.

Jones, M. (1994) Expectiles and M-quantiles are quantiles. *Statist. Probab. Lett.*, **20**, 149–153.

Koenker, R. (2011) Additive models for quantile regression: Model selection and confidence bandaids. *Braz. J. Probab. Stat.*, **25**, 239–262.

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

Kokic, P., Breckling, J. and Lübke, O. (2002) *A new definition of multivariate M-quantiles.*, chap. Statistical data analysis based on the L1-norm and related methods, 15–24. Basel, Switzerland: Birkhäuser.

Koltchinski, V. I. (1997) M-estimation, convexity and quantiles. *Ann. Statist.*, **25**, 435–477.

Kong, L. and Mizera, I. (2012) Quantile tomography: using quantiles with multivariate data. *Statist. Sinica*, **22**, 1589–1610.

Koshevoy, G. and Mosler, K. (1997) Zonoid trimming for multivariate distributions. *Ann. Statist.*, **25**, 1998–2017.

Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009) Assessing value at risk with care, the conditional autoregressive expectile models. *J. Econometrics*, **150**, 261–270.

Li, J., Cuesta-Albertos, J. and Liu, R. Y. (2012) Dd-classifier: Nonparametric classification procedures based on dd-plots. *J. Amer. Statist. Assoc.*, **107**, 737–753.

Maume-Deschamps, V., Rullière, D. and Said, K. (2017a) Asymptotic multivariate expectiles. *ArXiv preprint arXiv:1704.07152v2*.

— (2017b) Multivariate extensions of expectiles risk measures. *Depend. Model.*, **5**, 20–44.

Newey, W. and Powell, J. (1987) Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.

Paindaveine, D. and Šiman, M. (2011) On directional multiple-output quantile regression. *J. Multivariate Anal.*, **102**, 193–212.

Schnabel, S. and Eilers, P. (2009) Optimal expectile smoothing. *Comput. Stat. Data Anal.*, **53**, 4168–4177.

Schulze Waltrup, L., Sobotka, F., Kneib, T. and Kauermann, G. (2015) Expectile and quantile regression - David and Goliath? *Stat. Model.*, **15**, 433–456.

Serfling, R. J. (2002) Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerl.*, **56**, 214–232.

Sobotka, F. and Kneib, T. (2012) Geoadditive expectile regression. *Comput. Stat. Data Anal.*, **56**, 755–767.

Tukey, J. W. (1975) Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, 523–531. Canad. Math. Congress, Montreal, Que.

Waldmann, E. and Kneib, T. (2015) Bayesian bivariate quantile regression. *Stat. Model.*, **15**, 326–344.

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

Yao, Q. and Tong, H. (1996) Asymmetric least squares regression and estimation: a nonparametric approach. *J. Nonparametr. Stat.*, **6**, 273–292.

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