Supplement to "Multivariate Expectiles, Expectile Depth and Multiple-Output Expectile Regression"

Abdelaati Daouia* and Davy Paindaveine†

*,†Toulouse School of Economics, Université Toulouse 1 Capitole
†ECARES and Département de Mathématique, Université Libre de Bruxelles

These supplementary materials provide the following further contributions: for the sake of completeness, we first describe some of the main competing multivariate expectile concepts (Section S.1). We compute expectile depth and expectile depth regions in several multivariate examples (Section S.2). We state asymptotic results for the proposed expectile depth (Section S.3). We illustrate on simulated data the proposed multiple-output expectile regression methods and show that these dominate the corresponding quantile-based methods in terms of crossings (Section S.4). We discuss the relation between multivariate expectiles and risk measures, and we show that our expectiles satisfy the coherency axioms of multivariate risk measures (Section S.5). Finally, we prove all results of the paper (Section S.6).

Below, (n) denotes Equation n from the main paper, whereas $(\mathrm{S}.n)$ refers to Equation n from this supplement. Section S.n, Theorem n or Lemma S.n are used in the same way.

*Abdelaati Daouia is Maître de Conférences, Université Toulouse 1 Capitole TSE, 1, Esplanade de l'Université, 31080 Toulouse Cedex 06, France (E-mail: abdelaati.daouia@tse-fr.eu). Davy Paindaveine is Professor of Statistics, Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F. D. Roosevelt, 50, CP 114/04, B-1050 Bruxelles, Belgium (E-mail: dpaindav@ulb.ac.be).

S.1 Competing expectile concepts

We define below some of the main concepts of multivariate expectiles available in the literature, with a particular emphasis on the concepts we used in the paper for comparison with the proposed multivariate expectiles.

Before proceeding, it is needed to introduce an alternative parametrization of the univariate expectiles $e_\alpha = e_\alpha(P)$ from Section 2 (the dependence on P will play no role in this section, hence will be dropped in the notation). This alternative parametrization is $e_{\tau,u} := e_{(1-\tau u)/2}$ and indexes univariate expectiles by an order $\tau \in [0,1)$ and a direction $u \in \{-1,1\}$, or equivalently by an order τu that belongs to the open unit "ball" $(-1,1)$ of $\mathbb{R}$. In this directional parametrization, the most central expectile corresponds to $\tau = 0$ and the most extreme ones are obtained as $\tau \to 1$. In the d-dimensional case ($d \geq 2$), where there are no left nor right, it is natural to similarly index expectiles by an order $\tau \in [0,1)$ and a direction $\mathbf{u} \in \mathcal{S}^{d-1}$, or equivalently by a vectorial order $\tau\mathbf{u}$ belonging to the open unit ball $\{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| < 1\}$ of $\mathbb{R}^d$, with the same idea that $\tau = 0$ will yield the most central expectile and that $\tau\mathbf{u}$ with $\tau \to 1$ will provide extreme expectiles in direction $\mathbf{u}$. It is then standard (see the references below) to consider contours generated by expectiles of a fixed order τ. As in the body of the paper, these contours are the boundaries of "centrality regions" that provide a center-outward ordering of points in $\mathbb{R}^d$. There are alternative directional parametrizations for expectiles; typically, these involve an order $\alpha \in (0,1)$ and a direction $\mathbf{u} \in \mathcal{S}^{d-1}$, and are such that the expectile of order α in direction $\mathbf{u}$ is equal to the expectile of order $1-\alpha$ in direction $-\mathbf{u}$. For such a parametrization, central expectiles are associated with $\alpha = 1/2$, whereas extreme ones are obtained as $\alpha \to 0$ and $\alpha \to 1$ (this is the parametrization of multivariate expectiles that was used in the paper). In the rest of this section, we discriminate between these two parametrizations by using the notation τ

and α in a consistent way.

These general considerations allow us to review some of the main concepts of multivariate expectiles. The first concept can be found in Breckling and Chambers (1988), a paper whose main contribution was to introduce the concept of univariate *M-quantiles*, that generalize both univariate quantiles and expectiles. There, for any $\alpha \in (0, \frac{1}{2}]$, the order-$\alpha$ M-quantile of P in direction $\mathbf{u}$ is defined as the "geometric" quantity

$$\boldsymbol{\theta}^{\rho,\mathrm{geom}}_{\alpha,\mathbf{u}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \mathrm{E}\left[\left\{1-(1-2\alpha)\frac{\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta})}{\|\mathbf{Z}-\boldsymbol{\theta}\|}\right\}\rho(\|\mathbf{Z}-\boldsymbol{\theta}\|)\right]$$

(throughout this appendix, $\mathbf{Z}$ is a random d-vector with distribution P), where $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex loss function such that $\rho(0) = 0$; the definition for $\alpha \in (\frac{1}{2}, 1)$ results from the identity $\boldsymbol{\theta}^{\rho,\mathrm{geom}}_{\alpha,\mathbf{u}} = \boldsymbol{\theta}^{\rho,\mathrm{geom}}_{1-\alpha,-\mathbf{u}}$. The term "geometric" above is justified by the fact that, for $\rho(t) = |t|$ and $\rho(t) = t^2$, these M-quantiles reduce to the geometric quantiles

$$\mathbf{q}^{\mathrm{geom}}_{\alpha,\mathbf{u}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \mathrm{E}\big[\|\mathbf{Z}-\boldsymbol{\theta}\| - (1-2\alpha)\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta})\big] \tag{S.1}$$

and geometric expectiles

$$\mathbf{e}^{\mathrm{geom}}_{\alpha,\mathbf{u}} := \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \mathrm{E}\big[\|\mathbf{Z}-\boldsymbol{\theta}\|\{\|\mathbf{Z}-\boldsymbol{\theta}\| - (1-2\alpha)\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta})\}\big] \tag{S.2}$$

from Chaudhuri (1996) and Herrmann et al. (2018), respectively; these papers, that actually rather rely on the $(\tau, \mathbf{u})$-directional parametrization, would refer to (S.1) (resp., (S.2)) as quantiles (resp., expectiles) of order $\tau = 1-2\alpha$ in direction $-\mathbf{u}$. As recently showed theoretically in Girard and Stupfler (2017), geometric quantiles exhibit undesirable properties. In particular, (a) the extreme quantile contours obtained as $\alpha \to 0$ may extend far outside the support of the distribution. Also, (b) such extreme quantile contours exhibit a structure that is incompatible with the principal component structure of the underlying distribution: more precisely, they will be furthest (resp., closest) to the center of the distribution in the last (resp., first) principal direction, which is orthogonal to what one would expect from quantile contours. As we show empirically in Figure 1, these pathological features unfortunately extend to geometric expectiles (the empirical version of (S.2) is simply obtained by replacing the expectation with a sample average over the observations $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ at hand).

As we mentioned in the paper, geometric quantiles and expectiles may extend far outside the support of the distribution as $\alpha \to 0$. To improve on this, Breckling et al. (2001) and Kokic et al. (2002) introduced alternative concepts of multivariate M-quantiles, actually only for the case where ρ is a Huber loss function (which covers the loss functions providing quantiles and expectiles as limiting cases). To define these quantiles, we need to introduce the following notation: let $S(t) := \mathbb{I}[t > 0] - \mathbb{I}[t < 0]$ be the sign function, $\psi_c(\mathbf{t}) := (\mathbf{t}/c)\mathbb{I}[\|\mathbf{t}\| < c] + (\mathbf{t}/\|\mathbf{t}\|)\mathbb{I}[\|\mathbf{t}\| \geq c]$ be a d-variate extension of Huber's ψ-function, and further write $h_\alpha(t) := (1-\alpha)\mathbb{I}[t < 0] + (1/2)\mathbb{I}[t = 0] + \alpha\mathbb{I}[t > 0]$. Then, the Kokic et al. (2002) order-α M-quantile $\boldsymbol{\theta}^{\delta,c}_{\alpha,\mathbf{u}}$ of P in direction $\mathbf{u}$ is the solution $\boldsymbol{\theta}(\in \mathbb{R}^d)$ of

$$\mathrm{E}\left[\left\{(1-2\alpha)S(\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta}))\left(1 - \frac{|\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta})|}{\|\mathbf{Z}-\boldsymbol{\theta}\|}\right)^\delta + 2h_\alpha(\mathbf{u}'(\mathbf{Z}-\boldsymbol{\theta}))\right\}\psi_c(\mathbf{Z}-\boldsymbol{\theta})\right] = 0; \quad \text{(S.3)}$$

here, $c, \delta > 0$ are fixed. For $d = 1$, it is easy to check that $\theta^{\delta,c}_{\alpha,1}$, for any $\delta > 0$, reduces to the univariate quantile q_α as $c \to 0$ and to the univariate expectile e_α as $c \to \infty$, so that the limit of $\boldsymbol{\theta}^{\delta,c}_{\alpha,\mathbf{u}}$, as $c \to 0$ and as $c \to \infty$, may be considered as a multivariate quantile and as a multivariate expectile, respectively. The multivariate quantiles/expectiles from Breckling et al. (2001) then simply correspond to the particular case obtained for $\delta = 1$.

As explained in the paper, an important drawback of the aforementioned multivariate expectiles (and of other multivariate expectiles, such as those from Koltchinski, 1997) is their weak equivariance properties. More precisely, these expectiles are equivariant under orthogonal transformations, but they fail to be equivariant under general affine transformations. Actually, other recent proposals enjoy even weaker equivariance properties;

for instance, the multivariate expectiles from Maume-Deschamps et al. (2017a,b) are not equivariant under orthogonal transformations.

S.2 Multivariate examples for expectile depth

In this section, we compute expectile depth for several classical distributions over $\mathbb{R}^d$, with $d > 1$. Consider first 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 $\boldsymbol{\mu} = 0$, Theorem 8 provides

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

where we used the fact that the function G in (3) is a cumulative distribution function, hence is non-decreasing. For arbitrary $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, affine invariance entails that $HED(\mathbf{z}, P) = g(\|\mathbf{z}\|_{\mu,\Sigma})$, 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 $B^d := \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| \leq 1\}$ or on the unit sphere $\mathcal{S}^{d-1}$, then one can show that

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

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 (4). In all cases considered, thus, the function g is continuous and monotone strictly decreasing on its support, which illustrates the theoretical results of Section 4.2.

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, Theorem 8 provides

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

where $\mathcal{S}_\alpha^{d-1} := \{\mathbf{v} \in \mathbb{R}^d : \|\mathbf{v}\|_\alpha = 1\}$ is the unit L_α-sphere. Since the function G in (3) is non-decreasing, 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

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

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.

S.3 Asymptotic results for expectile depth

We provide here consistency results for expectile depth and expectile depth regions, that can be proved on the basis of Theorem 8 (proofs are provided in Section S.6). We start with the following uniform consistency result, that is the expectile depth analog of the classical halfspace depth result from Donoho and Gasko (1992), Section 6.

Theorem S.1. *Fix $P \in \mathcal{P}_d$ and let P_n be the empirical probability measure associated with a random sample of size n from P. Then, for any compact subset K of $\mathbb{R}^d$,*

$$\sup_{\mathbf{z} \in K} |HED(\mathbf{z}, P_n) - HED(\mathbf{z}, P)| \to 0$$

almost surely as $n \to \infty$.

For halfspace depth, this uniform consistency property, jointly with a general result on the consistency of M-estimators (such as Theorem 2.12 in Kosorok, 2008), allows one to establish almost sure consistency of the sample deepest point. For our expectile depth, however, asymptotic theory for the sample deepest point is trivial, since this deepest point is simply the sample average $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$ of the observations. In particular, the asymptotic distribution of the sample expectile deepest point trivially follows from the central limit theorem and can be used for inference on the population deepest point. In sharp contrast, the asymptotic distribution of the sample halfspace deepest point is so complicated (see Massé, 2002) that it is hopeless to perform inference based on it.

We turn to consistency of depth regions. This was first discussed in He and Wang (1997), where the focus was mainly on elliptical distributions. While results under milder conditions were obtained in Kim (2000) and Zuo and Serfling (2000), we will here exploit the general results from Dyckerhoff (2016). We have the following result.

Theorem S.2. *Fix $P \in \mathcal{P}_d$ and let P_n be the empirical probability measure associated with a random sample of size n from P. Then, for any compact interval $\mathcal{I}$ in $(0, \frac{1}{2})$,*

$$\sup_{\alpha \in \mathcal{I}} d_H(R_\alpha(P_n), R_\alpha(P)) \to 0$$

almost surely as $n \to \infty$, where d_H denotes the Hausdorff distance.

Remarkably, this consistency result holds without any assumption on P (beyond the fact that P belongs to $\mathcal{P}_d$). In comparison, the corresponding halfspace depth result requires that P is smooth (in the sense that it assigns probability mass zero to any hyperplane of $\mathbb{R}^d$) and that it has a connected support.

S.4 Multiple-output expectile regression on simulated data

To illustrate on simulated data the multiple-output expectile regression methods described in Section 5.2, we generated a random sample of size $n = 100$ from the multiple-output heteroscedastic linear regression model

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

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 expectile regions $R^{(n)}_{\alpha,x}$ and their quantile analogs from the nonparametric regression methods described in Section 5.2. For the sake of comparison, we also provide the contours obtained from the corresponding linear regression methods. The resulting contours are provided in Figure S.1. 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 S.1), which is incompatible with what occurs at the population level.

S.5 Multivariate expectile risks

The risk of a collection of financial assets is typically assessed by aggregating these assets, using their monetary values, into a combined random univariate portfolio Z. It is then sufficient to consider univariate risk measures $\varrho(Z)$; see Artzner et al. (1999) and Delbaen (2002). More and more often, however, the focus is on the more realistic situation where the risky portfolio is a random d-vector whose components relate to different security markets. In such a context, liquidity problems and/or transaction costs between the various security markets typically prevent investors from aggregating their portfolio into a univariate portfolio (Jouini et al., 2004). This calls for multivariate risk measures $\varrho(\mathbf{Z})$, where $\mathbf{Z}$ is a random d-vector.

Extensions of the axiomatic foundation for *coherent* univariate risk measures to the d-variate framework have been studied in Jouini et al. (2004) and Cascos and Molchanov (2007). Such extensions usually involve set-valued risk measures, as in the following definition (we restrict here to bounded random vectors as in Jouini et al., 2004, but the extension to the general case could be achieved as in Delbaen, 2002).

Definition S.1. *Let L_d^∞ be the set of (essentially) bounded random d-vectors and $\mathcal{B}_d$ be the Borel sigma-algebra on $\mathbb{R}^d$. Then a* coherent d-variate risk measure *is a function $R : L_d^\infty \to \mathcal{B}_d$ satisfying the following properties: (i) (translation invariance:) $R(\mathbf{Z}+\mathbf{z}) = R(\mathbf{Z}) + \mathbf{z}$ for any $\mathbf{Z} \in L_d^\infty$ and $\mathbf{z} \in \mathbb{R}^d$; (ii) (positive homogeneity:) $R(\lambda\mathbf{Z}) = \lambda R(\mathbf{Z})$ for any $\mathbf{Z} \in L_d^\infty$ and $\lambda > 0$; (iii) (monotonicity:) if $\mathbf{X} \leq \mathbf{Y}$ almost surely in the componentwise sense, then $R(\mathbf{Y}) \subset R(\mathbf{X}) \oplus \mathbb{R}_+^d$ and $R(\mathbf{X}) \subset R(\mathbf{Y}) \oplus \mathbb{R}_-^d$, where $\oplus$ denotes the Minkowski*

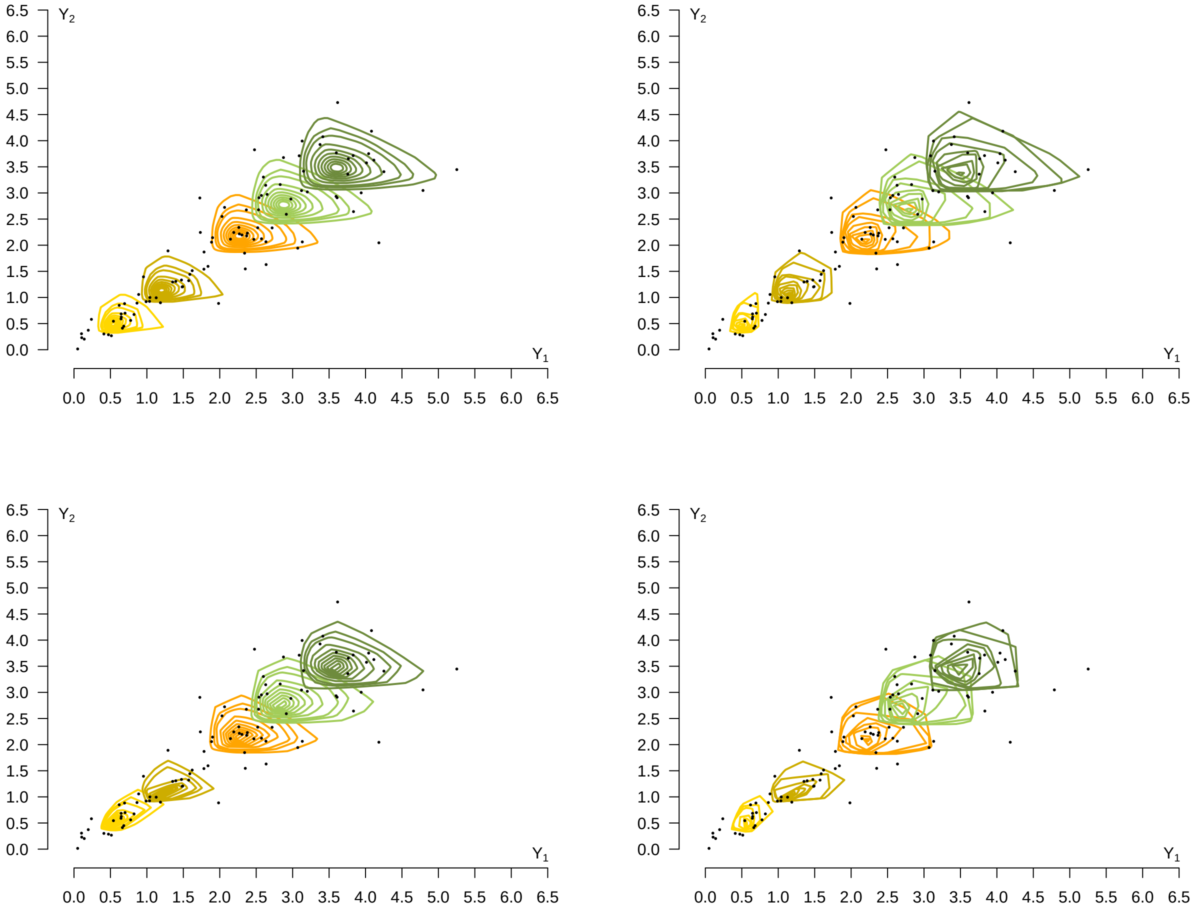

Figure S.1: (Left:) conditional expectile contours $\partial R^{(n)}_{\alpha,x}$, for $\alpha \in \{.01, .03, .05, .10, .15, \ldots, .40\}$ and for values of x that are the 10% (yellow), 30% (brown), 50% (orange), 70% (light green) and 90% (dark green) empirical quantiles of $X_1, \ldots, X_n$, obtained by applying a linear (top) or nonparametric (bottom) regression method to a random sample of size $n = 100$ from the heteroscedastic linear regression model in (S.4). (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 S.4 for details. Bivariate responses (Y_{i1}, Y_{i2}), $i = 1, \ldots, n$, are shown in black.

sum and where we let $\mathbb{R}^d_\pm := \{\mathbf{x} \in \mathbb{R}^d : \pm x_1 \geq 0, \ldots, \pm x_d \geq 0\}$*; (iv) (subadditivity:)* $R(\mathbf{X}+\mathbf{Y}) \subset R(\mathbf{X}) \oplus R(\mathbf{Y})$ *for any* $\mathbf{X}, \mathbf{Y} \in L^\infty_d$*; (v) (connectedness/closedness:)* $R(\mathbf{X})$ *is connected and closed for any* $\mathbf{X} \in L^\infty_d$*.*

In the univariate case, such coherent set-valued risk measures can be obtained as $R(Z) = [-\varrho(Z), \infty)$, where $\varrho(Z)$ is a real-valued coherent risk measure in the sense of Artzner et al. (1999) and Delbaen (2002); see Remark 2.2 in Jouini et al. (2004). For the most classical risk measure, namely the *Value at Risk*, the resulting set is $R(Z) = [-\mathrm{VaR}_\alpha(Z), \infty)$, where $-\mathrm{VaR}_\alpha(Z) = q_\alpha(Z)$ is the standard α-quantile of Z. The sign convention in $\mathrm{VaR}_\alpha(Z)$ corresponds to an implicit specification of the positive direction $u = 1$, which associates a positive risk measure with the typically negative profit—that is, loss—$q_\alpha(Z)$ obtained for small values of α.

In this univariate setting, the M-quantiles from Breckling and Chambers (1988), which encompass both quantiles and expectiles, have recently received a lot of attention since the resulting risk measures share the important property of *elicitability* (which corresponds to the existence of a natural backtesting methodology; Gneiting, 2011). In this framework, expectiles play a special role as they are the only M-quantiles providing coherent risk measures (Bellini et al., 2014). Actually, expectiles define the only coherent risk measure that is also elicitable (Ziegel, 2016). In the d-variate case, a natural expectile set-valued risk measure is given by our expectile halfspace $H_{\alpha,\mathbf{u}}(\mathbf{Z})$ in Definition 1 (in this section, $H_{\alpha,\mathbf{u}}(\mathbf{Z})$, $R_\alpha(\mathbf{Z}), \ldots$ respectively stand for $H_{\alpha,\mathbf{u}}(P)$, $R_\alpha(P)$, $\ldots$, where P is the distribution of $\mathbf{Z}$). Using quantiles rather than expectiles, this set-valued risk measure, for $d = 1$ and the positive direction $u = 1$, would reduce to the risk measure $[-\mathrm{VaR}_\alpha(Z), \infty)$ above, which, as already mentioned, also relies on the choice of a positive direction. For $d > 1$, it is similarly natural to restrict to "positive" directions $\mathbf{u}$, that is, to $\mathbf{u} \in \mathcal{S}^{d-1}_+ := \mathcal{S}^{d-1} \cap \mathbb{R}^d_+$.

Now, already for $d = 1$, the VaR risk measure fails to be subadditive in general (Acerbi, 2002). It is also often criticized for its insensitivity to extreme losses, since it depends on the frequency of tail losses but not on their severity. Denoting as $e_\alpha(Z)$ the order-α expectile of Z, the expectile risk measure $R(Z) = [e_\alpha(Z), \infty)$, with $\alpha \in (0, \frac{1}{2}]$, improves over VaR on both accounts since it is coherent (Bellini et al., 2014) and depends on the severity of tail losses (Kuan et al., 2009). Our expectile d-variate risk measure, namely the halfspace $H_{\alpha,\mathbf{u}}(\mathbf{Z})$ extends this univariate expectile risk measure to the d-variate setup and, quite nicely, turns out to be coherent for any $\alpha \in (0, \frac{1}{2}]$ and any direction $\mathbf{u} \in \mathcal{S}_+^{d-1}$: since connectedness/closedness holds trivially ($H_{\alpha,\mathbf{u}}(\mathbf{Z})$ is a closed halfspace) and since translation invariance and positive homogeneity directly follow from Theorem 1, we focus on monotonicity and subadditivity (see Definition S.1) and further cover some other properties from Dyckerhoff and Mosler (2011).

Theorem S.3. *Let $\mathbf{X}, \mathbf{Y}$ be random d-vectors with respective distributions P, Q in $\mathcal{P}_d$. Then, we have the following properties: (i) (monotonicity) if $\mathbf{X} \leq \mathbf{Y}$ almost surely in a componentwise sense, then $H_{\alpha,\mathbf{u}}(\mathbf{Y}) \subset H_{\alpha,\mathbf{u}}(\mathbf{X}) \oplus \mathbb{R}_+^d$ and $H_{\alpha,\mathbf{u}}(\mathbf{X}) \subset H_{\alpha,\mathbf{u}}(\mathbf{Y}) \oplus \mathbb{R}_-^d$ for any $\alpha \in (0,1)$ and $\mathbf{u} \in \mathcal{S}_+^{d-1}$; (ii) (subadditivity) for any $\alpha \in (0, \frac{1}{2}]$ and $\mathbf{u} \in \mathcal{S}^{d-1}$, $H_{\alpha,\mathbf{u}}(\mathbf{X}+\mathbf{Y}) \subset H_{\alpha,\mathbf{u}}(\mathbf{X}) \oplus H_{\alpha,\mathbf{u}}(\mathbf{Y})$; (iii) (superadditivity) for any $\alpha \in [\frac{1}{2}, 1)$ and $\mathbf{u} \in \mathcal{S}^{d-1}$, $H_{\alpha,\mathbf{u}}(\mathbf{X}) \oplus H_{\alpha,\mathbf{u}}(\mathbf{Y}) \subset H_{\alpha,\mathbf{u}}(\mathbf{X}+\mathbf{Y})$; (iv) (nestedness:) for any $\mathbf{u} \in \mathcal{S}^{d-1}$, $\alpha \mapsto H_{\alpha,\mathbf{u}}(\mathbf{X})$ is non-increasing with respect to inclusion.*

In order to illustrate these d-variate M-quantile risk measures, we briefly consider the daily returns on the IBM and MSFT shares from 03-01-2007 to 27-09-2018. The data were taken from Yahoo Finance using the *quantmod* package in R. Figure S.2 shows the resulting $n = 3{,}134$ bivariate observations along with some of the corresponding expectile risk measures $H_{\alpha,u}(P_n)$ (more precisely, the figure only displays their boundary hyperplanes) and some HED regions $R_\alpha(P_n)$. We also provide there a few halfspace depth regions and zonoid depth regions. For $d = 1$, the latter are related to *expected shortfall*, hence are also connected to risk measures. However, while zonoid regions formally are coherent risk measures (Cascos and Molchanov, 2007; Dyckerhoff and Mosler, 2011), a univariate zonoid depth region is not an interval of the usual form $[-\varrho(Z), \infty)$ but rather a compact interval. Our expectile risk measures $H_{\alpha,\mathbf{u}}(P_n)$, $\mathbf{u} \in \mathcal{S}_+^{d-1}$, offer an intuitive interpretation for the multivariate risk in the sense that the required capital reserve should cover any loss associated with joint returns inside $H_{\alpha,\mathbf{u}}(P_n)$, that is, above the hyperplane $\pi_{\alpha,\mathbf{u}}(P_n)$. Such losses can easily be identified in an automatic way. For these risks, the choice of a suitable security level α and direction $\mathbf{u} \in \mathcal{S}_+^{d-1}$ is a decision that should be made by risk managers and regulators. Other d-variate set-valued risk measures that trim unfavorable returns only yet do not require the choice of a direction $\mathbf{u}$, are the upper envelopes $\cap_{\mathbf{u}\in\mathcal{S}_+^{d-1}} H_{\alpha,\mathbf{u}}(P_n)$ of our directional expectile risk measures, or the full expectile regions themselves: both losses and profits associated with joint returns inside these regions can also be easily determined.

S.6 Proofs

This last section of the supplement presents the proofs of all results stated in the main paper and in previous sections of the supplement. We start with the following result, that extends the result from Jones (1994) (note that Jones' result excludes the sample case) and clarifies the link between both definitions of expectiles provided in Section 2.

Theorem S.4. *Fix $P \in \mathcal{P}_1$ and $\alpha \in (0,1)$. Let Z be a random variable with distribution P. Then, (i) $\theta \mapsto O_\alpha(\theta) = \mathrm{E}[\rho_\alpha(Z-\theta) - \rho_\alpha(Z)]$ is well-defined for any θ, and it is continuously differentiable over $\mathbb{R}$. (ii) The sign of its derivative at θ is the same as that of $G(\theta) - \alpha$, where G was defined in (3); (iii) $\theta \mapsto G(\theta)$ is a continuous cumulative distribution function over $\mathbb{R}$. (iv) The* order-α expectile of P

$$e_\alpha(P) := \min\left\{\theta \in \mathbb{R} : G(\theta) \geq \alpha\right\} \tag{S.5}$$

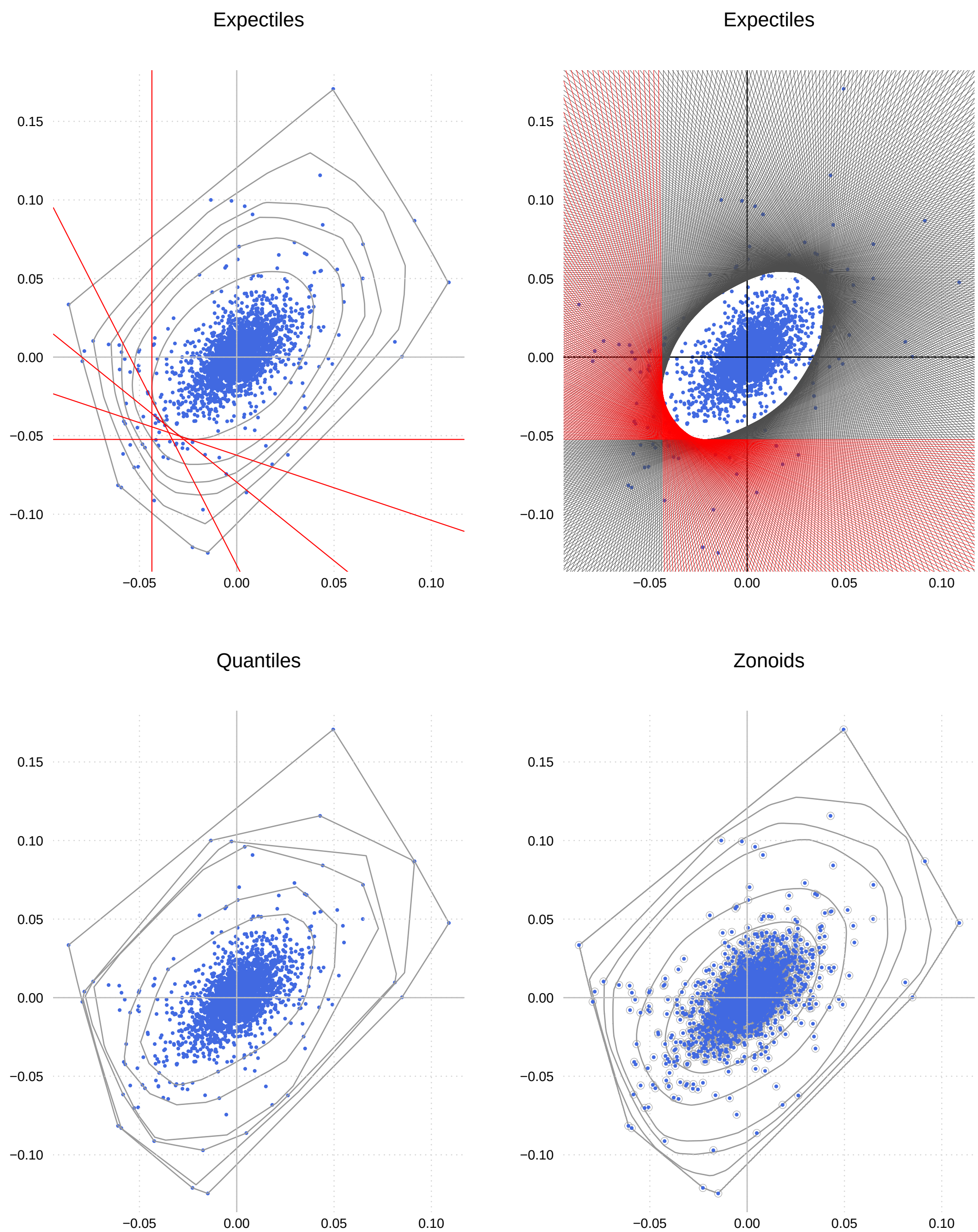

Figure S.2: (Top left:) boundary hyperplanes of the expectile halfspaces $H_{\alpha,\mathbf{u}}(P_n)$, for $\alpha = .003$ and $\mathbf{u} = (\cos\frac{\ell\pi}{8}, \sin\frac{\ell\pi}{8})'$ with $\ell = 0, 1, 2, 3, 4$, along with the HED regions $R_\alpha(P_n)$ associated with the extreme levels $\alpha = .000001, .0001, .0003, .0005, .001, .003$. (Top right): the boundary hyperplanes of the 500 expectile halfspaces that led to the construction of $R_\alpha(P_n)$ with $\alpha = .003$; hyperplanes associated with (positive) directions $\mathbf{u} \in \mathcal{S}^1_+$ are drawn in red. (Bottom left): halfspace depth regions of order $\alpha = .0003, .0005, .0007, .001, .003, .005$. (Bottom right:) zonoid depth regions of order $\alpha = .0003, .001, .002, .003, .01, .03$.

is well-defined and minimizes $\theta \mapsto O_\alpha(\theta)$ over $\mathbb{R}$, hence provides a unique representative of the argmin in (2).

Proof of Theorem S.4. (i) Note that $t \mapsto \rho_\alpha(t) = \{(1-\alpha)\mathbb{I}[t<0] + \alpha\mathbb{I}[t>0]\}t^2$ is differentiable on $\mathbb{R}$ with derivative $t \mapsto 2\{(1-\alpha)\mathbb{I}[t<0] + \alpha\mathbb{I}[t>0]\}t$, so that the mean-value theorem ensures that, for some $\lambda \in (0,1)$,

$$\big|\rho_\alpha(z-\theta) - \rho_\alpha(z)\big| \leq 2|\theta| \max(\alpha, 1-\alpha)|z - \lambda\theta| \leq 2|\theta|(|z| + |\theta|).$$

Consequently, since $P \in \mathcal{P}_1$, we have

$$|O_\alpha(\theta)| \leq \int_{-\infty}^{\infty} |\rho_\alpha(z-\theta) - \rho_\alpha(z)|\, dP(z) \leq 2|\theta|(\mathrm{E}[|Z|] + |\theta|) < \infty$$

for any θ. The mapping $\theta \mapsto O_\alpha(\theta)$ is thus well-defined for any θ.

We turn to differentiability of $\theta \mapsto O_\alpha(\theta)$. To do so, let

$$O'_\alpha(\theta) := 2(1-\alpha)\int_{-\infty}^{\infty} |z-\theta|\mathbb{I}[z<\theta]\, dP(z) - 2\alpha \int_{-\infty}^{\infty} |z-\theta|\mathbb{I}[z>\theta]\, dP(z) \tag{S.6}$$

and fix $\theta_0 \in \mathbb{R}$. For any $h \in [-1,1] \setminus \{0\}$, write then

$$\begin{aligned}
&\frac{O_\alpha(\theta_0+h) - O_\alpha(\theta_0)}{h} - O'_\alpha(\theta_0) \\
&\qquad = \int_{-\infty}^{\infty} \left\{ \frac{\rho_\alpha(z-\theta_0-h) - \rho_\alpha(z-\theta_0)}{h} - O'_\alpha(\theta_0) \right\} dP(z) \\
&\qquad = (1-\alpha)\int_{-\infty}^{\infty} L_{\theta_0}(h,z)\, dP(z) + \alpha \int_{-\infty}^{\infty} R_{\theta_0}(h,z)\, dP(z),
\end{aligned} \tag{S.7}$$

where we let

$$L_{\theta_0}(h,z) := \frac{(z-\theta_0-h)^2\mathbb{I}[z<\theta_0+h] - (z-\theta_0)^2\mathbb{I}[z<\theta_0]}{h} - 2|z-\theta_0|\mathbb{I}[z<\theta_0]$$

and

$$R_{\theta_0}(h,z) := \frac{(z-\theta_0-h)^2\mathbb{I}[z>\theta_0+h] - (z-\theta_0)^2\mathbb{I}[z>\theta_0]}{h} + 2|z-\theta_0|\mathbb{I}[z>\theta_0].$$

For any z, $\theta \mapsto (z-\theta)^2\mathbb{I}[z<\theta]$ is differentiable at θ_0, with derivative $-2(z-\theta_0)\mathbb{I}[z<\theta_0] = 2|z-\theta_0|\mathbb{I}[z<\theta_0]$, so that, for any z, we have that $L_{\theta_0}(h,z) \to 0$ as h converges to zero. The mean-value theorem then implies that, for any $h \in [-1,1]$, there exists $\lambda \in (0,1)$ such that

$$\begin{aligned}|L_{\theta_0}(h,z)| &\leq 2|z-\theta_0-\lambda h|\mathbb{I}[z<\theta_0+\lambda h] + 2|z-\theta_0|\mathbb{I}[z<\theta_0]\\ &\leq 2(\theta_0+\lambda h - z) + 2(\theta_0 - z)\\ &\leq 4|z| + 4|\theta_0| + 1.\end{aligned}$$

Since this upper-bound, which does not depend on h, is a P-integrable function of z, Lebesgue's DCT entails that $\int_{-\infty}^{\infty} L_{\theta_0}(h,z)\, dP(z) \to 0$ as h converges to zero. Using the fact that, for any z, $\theta \mapsto (z-\theta)^2\mathbb{I}[z>\theta]$ is differentiable at θ_0, with derivative $-2|z-\theta_0|\mathbb{I}[z>\theta_0]$, one can similarly show that $\int_{-\infty}^{\infty} R_{\theta_0}(h,z)\, dP(z) \to 0$ as h converges to zero. From (S.7), this establishes that $\theta \mapsto O_\alpha(\theta)$ is differentiable, with derivative

$$\begin{aligned}O'_\alpha(\theta) &= 2(1-\alpha)\int_{-\infty}^{\infty} |z-\theta|\mathbb{I}[z<\theta]\, dP(z) - 2\alpha \int_{-\infty}^{\infty} |z-\theta|\mathbb{I}[z>\theta]\, dP(z)\\ &= 2(1-\alpha)\Big\{ -\int_{-\infty}^{\infty} (z-\theta)\mathbb{I}[z<\theta]\, dP(z)\Big\} - 2\alpha \int_{-\infty}^{\infty} (z-\theta)\mathbb{I}[z>\theta]\, dP(z)\\ &=: 2(1-\alpha)H_2(\theta) - 2\alpha H_1(\theta).\end{aligned} \tag{S.8}$$

A trivial application of Lebesgue's DCT shows that both H_1 and H_2 are continuous on $\mathbb{R}$, so that O_α is continuously differentiable.

(ii) The assumption that $P[\{\theta\}] < 1$ for any $\theta \in \mathbb{R}$ ensures that

$$H_1(\theta) + H_2(\theta) = \int_{-\infty}^{\infty} |z-\theta|\, dP(z) = \int_{-\infty}^{\infty} |z-\theta|\mathbb{I}[z \neq \theta]\, dP(z) > 0$$

for any $\theta \in \mathbb{R}$. Therefore, we may write

$$O'_\alpha(\theta) = 2(1-\alpha)H_2(\theta) - 2\alpha H_1(\theta) = 2\,(G(\theta)-\alpha)\,(H_1(\theta)+H_2(\theta)),$$

where

$$G(\theta) = \frac{H_2(\theta)}{H_1(\theta) + H_2(\theta)}, \tag{S.9}$$

is the function defined in (3). It follows that $O'_\alpha(\theta)$ and $G(\theta) - \alpha$ have the same sign.

(iii) Fix $\theta_b > \theta_a$. We have $(z-\theta_b)\mathbb{I}[z > \theta_b] \le (z-\theta_a)\mathbb{I}[z > \theta_a]$, so that $H_1(\theta_b) \le H_1(\theta_a)$. Similarly, $-(z-\theta_b)\mathbb{I}[z < \theta_b] \ge -(z-\theta_a)\mathbb{I}[z < \theta_a]$, so that $H_2(\theta_b) \ge H_2(\theta_a)$. Therefore, H_1 and H_2 are monotone non-increasing and non-decreasing, respectively. Note also that both H_1 and H_2 take their values in $\mathbb{R}_+$. Since a direct computation shows that

$$\begin{aligned}\{H_1(\theta_a) + H_2(\theta_a)\}&\{H_1(\theta_b) + H_2(\theta_b)\}\{G(\theta_b) - G(\theta_a)\} \\ &= H_1(\theta_a)\{H_2(\theta_b) - H_2(\theta_a)\} - H_2(\theta_a)\{H_1(\theta_b) - H_1(\theta_a)\},\end{aligned} \tag{S.10}$$

we thus conclude that G is monotone non-decreasing.

The Monotone Convergence Theorem implies that $\lim_{\theta\to-\infty} H_2(\theta) = 0$ and $\lim_{\theta\to\infty} H_1(\theta) = 0$. Since H_1 is a monotone non-increasing function of θ, $H_1(\theta)$ will stay away from zero for large negative values of θ, which implies that

$$\lim_{\theta\to-\infty} G(\theta) = \lim_{\theta\to-\infty} \frac{H_2(\theta)}{H_1(\theta) + H_2(\theta)} = 0.$$

Similarly, since H_2 is a monotone non-decreasing function of θ, $H_2(\theta)$ will stay away from zero for large positive values of θ, so that

$$\lim_{\theta\to\infty} (1 - G(\theta)) = \lim_{\theta\to\infty} \frac{H_1(\theta)}{H_1(\theta) + H_2(\theta)} = 0.$$

Finally, in view of (S.9), the continuity of G trivially follows from that of H_1 and H_2.

(iv) Since G is a continuous cumulative distribution function, the set $S_\alpha := \{\theta \in \mathbb{R} : G(\theta) \ge \alpha\}$ is non-empty and is lower-bounded. Thus, S_α admits an infimum, which, from continuity, is a minimum. This guarantees existence of $e_\alpha(P)$. We thus have $G(e_\alpha(P)) \ge \alpha$.

If $G(e_\alpha(P)) > \alpha$, then continuity of G will guarantee the existence of $c < e_\alpha(P)$ such that $G(c) \geq \alpha$, which would contradict the definition of $e_\alpha(P)$. Therefore, $G(e_\alpha(P)) = \alpha$. Hence, monotonicity of G provides $G(\theta) \geq \alpha$ for any $\theta \geq e_\alpha(P)$, which implies that $O'_\alpha(\theta) \geq 0$ for any $\theta \geq e_\alpha(P)$, hence that $O_\alpha(\theta) \geq O_\alpha(e_\alpha(P))$ for any $\theta \geq e_\alpha(P)$. Now, the definition of $e_\alpha(P)$ ensures that $G(\theta) < \alpha$ for any $\theta < e_\alpha(P)$, which implies that $O'_\alpha(\theta) < 0$ for any such θ, hence that $O_\alpha(\theta) > O_\alpha(e_\alpha(P))$ for any $\theta < e_\alpha(P)$. We conclude that $e_\alpha(P)$ is a minimizer of $\theta \mapsto O_\alpha(\theta)$. $\square$

Lemma S.1 below will be needed in subsequent proofs, but we present it here since its proof uses the notation introduced in the proof of Theorem S.4 above.

Lemma S.1. *Fix $P \in \mathcal{P}_1$. Let $\theta' < \theta''$ with $\theta', \theta'' \in C(P) = \{\theta \in \mathbb{R} : \min(\mathbb{P}[Z \leq \theta], \mathbb{P}[Z \geq \theta]) > 0\}$, where Z has distribution P. Then G is monotone strictly increasing over $[\theta', \theta'']$.*

Proof of Lemma S.1. Fix θ_a, θ_b with $\theta' < \theta_a < \theta_b < \theta''$. Then we have $\mathbb{P}[Z > \theta_a] \geq \mathbb{P}[Z \geq \theta''] > 0$ and $\mathbb{P}[Z < \theta_b] \geq \mathbb{P}[Z \leq \theta'] > 0$, or (in terms of the cumulative distribution function F of P:) $1 - F(\theta_a) > 0$ and $F(\theta_b - 0) > 0$. Now, (S.10) provides

$$\{H_1(\theta_a) + H_2(\theta_a)\}\{H_1(\theta_b) + H_2(\theta_b)\}\{G(\theta_b) - G(\theta_a)\} \geq H_1(\theta_a)\{H_2(\theta_b) - H_2(\theta_a)\}. \quad \text{(S.11)}$$

Since $1 - F(\theta_a) > 0$, we have

$$H_1(\theta_a) = \int_{-\infty}^{\infty} (z - \theta_a)\mathbb{I}[z > \theta_a]\, dP(z) > 0,$$

whereas, since $F(\theta_b - 0) > 0$, we have

$$\begin{aligned}
H_2(\theta_b) - H_2(\theta_a) &= \int_{-\infty}^{\infty} \big\{ - (z - \theta_b)\mathbb{I}[z < \theta_b] + (z - \theta_a)\mathbb{I}[z < \theta_a] \big\}\, dP(z) \\
&= \int_{-\infty}^{\infty} (\theta_b - \theta_a)\mathbb{I}[z \leq \theta_a]\, dP(z) + \int_{-\infty}^{\infty} (\theta_b - z)\mathbb{I}[\theta_a < z < \theta_b]\, dP(z) \\
&= \int_{-\infty}^{\infty} \min(\theta_b - \theta_a, \theta_b - z)\mathbb{I}[z < \theta_b]\, dP(z) > 0.
\end{aligned}$$

Thus, it follows from (S.11) that $G(\theta_b) > G(\theta_a)$. Consequently, G is monotone strictly increasing over (θ', θ''), hence also over $[\theta', \theta'']$. $\square$

In the rest of this supplement, $\theta_\alpha(Z)$, $H_{\alpha,\mathbf{u}}(\mathbf{Z}), \ldots$, will respectively stand for $\theta_\alpha(P)$, $H_{\alpha,\mathbf{u}}(P), \ldots$, where P is the distribution of Z or $\mathbf{Z}$.

Proof of Theorem 1. Let $\mathbf{Z}$ be a random d-vector with distribution P. Denote as S_α the set of real numbers ϕ such that

$$\frac{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \phi|\mathbb{I}[\mathbf{u}'\mathbf{Z} - \phi \leq 0]]}{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \phi|]} \geq \alpha$$

and as T_α the set of real numbers θ such that

$$\frac{\mathrm{E}[|\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) - \theta|\mathbb{I}[\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) \leq \theta]]}{\mathrm{E}[|\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) - \theta|]} \geq \alpha.$$

Note that

$$\begin{aligned}
&\frac{\mathrm{E}[|\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) - \theta|\mathbb{I}[\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) \leq \theta]]}{\mathrm{E}[|\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b}) - \theta|]} \\
&\quad = \frac{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \{\|(\mathbf{A}^{-1})'\mathbf{u}\|\theta - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b}\}|\mathbb{I}[\mathbf{u}'\mathbf{Z} - \{\|(\mathbf{A}^{-1})'\mathbf{u}\|\theta - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b}\} \leq 0]]}{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \{\|(\mathbf{A}^{-1})'\mathbf{u}\|\theta - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b}\}|]},
\end{aligned}$$

so that $\theta \in T_\alpha$ if and only if $\|(\mathbf{A}^{-1})'\mathbf{u}\|\theta - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b} \in S_\alpha$. Thus, $S_\alpha = \|(\mathbf{A}^{-1})'\mathbf{u}\|T_\alpha - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b}$. Since $e_\alpha(\mathbf{u}'\mathbf{Z}) = \min S_\alpha$ and $e_\alpha(\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z}+\mathbf{b})) = \min T_\alpha$ by definition (see (S.5)), this implies that $e_\alpha(\mathbf{u}'\mathbf{Z}) = \|(\mathbf{A}^{-1})'\mathbf{u}\|e_\alpha(\mathbf{u}_{\mathbf{A}}'(\mathbf{A}\mathbf{Z} + \mathbf{b})) - \mathbf{u}'\mathbf{A}^{-1}\mathbf{b}$. We conclude that $H_{\alpha,\mathbf{u}_{\mathbf{A}}}(\mathbf{A}\mathbf{Z} + \mathbf{b})$ collects the d-vectors $\mathbf{z}$ satisfying

$$\mathbf{u}_{\mathbf{A}}'\mathbf{z} \geq \frac{e_\alpha(\mathbf{u}'\mathbf{Z})}{\|(\mathbf{A}^{-1})'\mathbf{u}\|} + \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{b}}{\|(\mathbf{A}^{-1})'\mathbf{u}\|},$$

or equivalently, $\mathbf{u}'(\mathbf{A}^{-1}(\mathbf{z} - \mathbf{b})) \geq e_\alpha(\mathbf{u}'\mathbf{Z})$. This establishes the result. $\square$

Let $\mathbf{Z}$ be a random d-vector with distribution $P \in \mathcal{P}_d$. In the subsequent proofs, we let

$$G_{\mathbf{u}}(\theta) = G_{\mathbf{u}}^e(\theta) := \frac{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \theta|\mathbb{I}[\mathbf{u}'\mathbf{Z} - \theta \leq 0]]}{\mathrm{E}[|\mathbf{u}'\mathbf{Z} - \theta|]}. \tag{S.12}$$

With this notation, $e_\alpha(P_{\mathbf{u}}) = e_\alpha(\mathbf{u}'\mathbf{Z})$ is given by $\min\{\theta \in \mathbb{R} : G_{\mathbf{u}}(\theta) \geq \alpha\}$.

Proof of Theorem 2. Fix $\alpha \in (0,1)$. By definition, $R_\alpha(P)$ is an intersection of closed convex subsets of $\mathbb{R}^d$, so that it is itself closed and convex. Since the affine-equivariance relation $R_\alpha(P_{\mathbf{A},\mathbf{b}}) = \mathbf{A}R_\alpha(P)+\mathbf{b}$ is a direct corollary of Theorem 1, it only remains to show that (i) $R_\alpha(P) \subset C(P)$ and that (ii) $R_\alpha(P)$ is bounded. Let us start with (i). Fix $\mathbf{z} \notin C(P)$ and let $\mathbf{Z}$ be a random d-vector with distribution P. Then there exists $\mathbf{u}_0 \in \mathcal{S}^{d-1}$ such that $\mathbb{P}[\mathbf{u}_0'\mathbf{Z} \leq \mathbf{u}_0'\mathbf{z}] = 0$, so that $G_{\mathbf{u}_0}(\mathbf{u}_0'\mathbf{z}) = 0$. Continuity of $G_{\mathbf{u}_0}$ (Theorem S.4(iii)) then entails that $\mathbf{u}_0'\mathbf{z} < e_\alpha(\mathbf{u}_0'\mathbf{Z})$. This implies that $\mathbf{z} \notin H_{\alpha,\mathbf{u}_0}(P)$, hence that $\mathbf{z} \notin R_\alpha(P)$. (ii) Fix $\mathbf{z} \in R_\alpha(P)$. For any $j \in \{1,\ldots,d\}$, we must have $\mathbf{z} \in H_{\alpha,\mathbf{e}_j}(P) \cap H_{\alpha,-\mathbf{e}_j}(P)$, where $\mathbf{e}_j$ denotes the jth vector of the canonical basis of $\mathbb{R}^d$. This implies that, for any j, we have $z_j \geq e_\alpha(Z_j)$ and $-z_j \geq e_\alpha(-Z_j)$, that is $z_j \in [e_\alpha(Z_j), -e_\alpha(-Z_j)]$. Since the definition in (3) (or (S.5)) entails that $e_\alpha(Y)$ is finite for any random variable Y, it follows that $R_\alpha(P)$ is a subset of the bounded hyperrectangle $\bigtimes_{j=1}^d [e_\alpha(Z_j), -e_\alpha(-Z_j)]$, hence is itself bounded. $\square$

Proof of Theorem 3. In this proof, we let $\tilde{R}_\alpha(P) := \{\mathbf{y} \in \mathbb{R}^d : \mathit{HED}(\mathbf{y},P) \geq \alpha\}$. Assume first that $\mathbf{z} \in R_\alpha(P)$. Then $\mathit{HED}(\mathbf{z},P) = \sup\{\beta > 0 : \mathbf{z} \in R_\beta(P)\} \geq \alpha$, so that $\mathbf{z} \in \tilde{R}_\alpha(P)$. Now, assume that $\mathbf{z} \notin R_\alpha(P)$. Then there exists $\mathbf{u}$ such that $\mathbf{u}'\mathbf{z} < e_\alpha(P_{\mathbf{u}})$. By definition of $e_\alpha(P_{\mathbf{u}})$, we must have $G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}) < \alpha$. Fix then $\alpha' \in (G_{\mathbf{u}}(\mathbf{u}'z), \alpha)$. Since $G_{\mathbf{u}}$ is continuous, there exists $\delta \in (0, e_\alpha(P_{\mathbf{u}}) - \mathbf{u}'\mathbf{z})$ such that $G_{\mathbf{u}}(t) < \alpha'$ for any $t \in [\mathbf{u}'\mathbf{z}, \mathbf{u}'\mathbf{z}+\delta]$. The monotonicity of $G_{\mathbf{u}}$ then implies that $\mathbf{u}'\mathbf{z} < \mathbf{u}'\mathbf{z}+\delta \leq e_{\alpha'}(P_{\mathbf{u}})$, which entails that $\mathbf{z} \notin R_{\alpha'}(P)$. Recalling that the expectile regions $R_\beta(P)$ are nested, this implies that $\mathit{HED}(\mathbf{z},P) \leq \alpha' < \alpha$, so that $\mathbf{z} \notin \tilde{R}_\alpha(P)$. $\square$

In the main paper, our theorems were presented in an order that was fixed by pedagogical considerations. In this supplement, however, we will need to prove theorems in the following order: Theorem 8, Theorem 5, Theorem 7, Theorem 6, Theorem 4, Theorem S.1, Theorem S.2, Theorem 9, and finally Theorem S.3. We start with Theorem 8 as it provides an alternative of the HED that is needed in many subsequent proofs. The proof of Theorem 8 requires the following preliminary result.

Lemma S.2. *Let* $\mathbf{Z}$ *be a random* d*-vector with distribution* $P \in \mathcal{P}_d$*. Then, (i)* $(\mathbf{u}, \mathbf{z}) \mapsto \ell_-(\mathbf{u}, \mathbf{z}) := \mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) < 0]]$ *and* $(\mathbf{u}, \mathbf{z}) \mapsto \ell_+(\mathbf{u}, \mathbf{z}) := \mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) > 0]]$ *are continuous over* $\mathcal{S}^{d-1} \times \mathbb{R}^d$*, so that (ii)*

$$(\mathbf{u}, \mathbf{z}) \mapsto G_{\mathbf{z}}(\mathbf{u}) = \frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|\mathbb{I}[\mathbf{u}'\mathbf{Z} \leq \mathbf{u}'\mathbf{z}]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|]}$$

(see (5)) is continuous over $\mathcal{S}^{d-1} \times \mathbb{R}^d$*.*

Proof of Lemma S.2. (i) We only prove the result for ℓ_-, as the proof for ℓ_+ is entirely similar. Fix $(\mathbf{u}_0, \mathbf{z}_0) \in \mathcal{S}^{d-1} \times \mathbb{R}^d$ and write $B_{\mathbf{z}_0}(r) := \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z} - \mathbf{z}_0\| \leq r\}$. For any $\mathbf{y} \in \mathbb{R}^d$, we have that $(\mathbf{u}, \mathbf{z}) \mapsto \mathbf{u}'(\mathbf{y}-\mathbf{z})\mathbb{I}[\mathbf{u}'\mathbf{y} < \mathbf{u}'\mathbf{z}]$ is continuous at $(\mathbf{u}_0, \mathbf{z}_0)$. Moreover, for any $(\mathbf{u}, \mathbf{z}) \in \mathcal{S}^{d-1} \times B_{\mathbf{z}_0}(1)$, the function $\mathbf{y} \mapsto \mathbf{u}'(\mathbf{y} - \mathbf{z})\mathbb{I}[\mathbf{u}'\mathbf{y} < \mathbf{u}'\mathbf{z}]$ is upper-bounded by the function $\mathbf{y} \mapsto \|\mathbf{z}_0\| + 1 + \|\mathbf{y}\|$ that is P-integrable and does not depend on $(\mathbf{u}, \mathbf{z})$. The Lebesgue Dominated Convergence Theorem therefore yields the result. (ii) Since

$$G_{\mathbf{z}}(\mathbf{u}) = \frac{\ell_-(\mathbf{u}, \mathbf{z})}{\ell_-(\mathbf{u}, \mathbf{z}) + \ell_+(\mathbf{u}, \mathbf{z})}, \tag{S.13}$$

the result readily follows from Part (i) (note that the assumption that $P \in \mathcal{P}_d$ ensures that $\ell_-(\mathbf{u}, \mathbf{z}) + \ell_+(\mathbf{u}, \mathbf{z}) > 0$). □

Proof of Theorem 8. Fix $z \in \mathbb{R}^d$ and let $\alpha := \min_{\mathbf{u} \in \mathcal{S}^{d-1}} G_{\mathbf{z}}(\mathbf{u})$ (existence of the minimum follows from Lemma S.2(ii) and the compactness of $\mathcal{S}^{d-1}$). Then, $G_{\mathbf{z}}(\mathbf{u}) \geq \alpha$ for any $\mathbf{u} \in \mathcal{S}^{d-1}$. By definition, we must then have $\mathbf{u}'\mathbf{z} \geq e_\alpha(\mathbf{u}'\mathbf{Z})$ for any $\mathbf{u}$, that is, $\mathbf{z} \in H_{\alpha,\mathbf{u}}(P)$ for any $\mathbf{u}$. This implies that $\mathbf{z} \in R_\alpha(P)$, hence that $\mathit{HED}(\mathbf{z}, P) \geq \alpha$. By contradiction, assume now that $\alpha' := \mathit{HED}(\mathbf{z}, P) > \alpha$. Then $\mathbf{z} \in R_{\alpha'}(P)$, so that $\mathbf{z} \in H_{\alpha',\mathbf{u}}(P)$

for any $\mathbf{u}$, i.e., that $\mathbf{u}'\mathbf{z} \geq e_{\alpha'}(\mathbf{u}'\mathbf{Z})$ for any $\mathbf{u}$. Since $G_{\mathbf{u}}$ is continuous and monotone non-decreasing (in view of (S.12) and Theorem S.4(iii)), this entails that $G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}) \geq G_{\mathbf{u}}(e_{\alpha'}(\mathbf{u}'\mathbf{Z})) = \alpha'$ for any $\mathbf{u}$. Consequently, we must have $\alpha = \min_{\mathbf{u}\in\mathcal{S}^{d-1}} G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}) \geq \alpha'$, a contradiction. $\square$

Proof of Theorem 5. Let Y be a random variable with a distribution in $\mathcal{P}_1$. It follows from Theorem S.4 and Lemma S.1 that the order-1/2 expectile of Y is $\mathrm{E}[Y]$ and that this expectile is the only value of θ such that $\mathrm{E}[|Y-\theta|\mathbb{I}[Y-\theta\leq 0]]/\mathrm{E}[|Y-\theta|] = 1/2$. Letting $\mathbf{Z}$ be a random d-vector with distribution P, we thus have that

$$\frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathrm{E}[\mathbf{Z}])|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathrm{E}[\mathbf{Z}])\leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathrm{E}[\mathbf{Z}])|]} = \frac{1}{2}$$

for any $\mathbf{u}\in\mathcal{S}^{d-1}$. Theorem 8 thus entails that $HED(\mathrm{E}[\mathbf{Z}], P) = 1/2$.

Assume now that there exists $\mathbf{z}\in\mathbb{R}^d$ such that $HED(\mathbf{z}, P) = e > 1/2$. Then, for an arbitrarily fixed $\mathbf{u}\in\mathcal{S}^{d-1}$, Theorem 8 implies that

$$\frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z})\leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|]} \geq e \tag{S.14}$$

and

$$\frac{\mathrm{E}[|-\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[-\mathbf{u}'(\mathbf{Z}-\mathbf{z})\leq 0]]}{\mathrm{E}[|-\mathbf{u}'(\mathbf{Z}-\mathbf{z})|]} \geq e. \tag{S.15}$$

Adding up these two inequalities yields $1 \geq 2e$, a contradiction. We conclude that $HED(\mathrm{E}[\mathbf{Z}], P) \geq HED(\mathbf{z}, P)$ for any $\mathbf{z}\in\mathbb{R}^d$, and it only remains to show that $\mathrm{E}[\mathbf{Z}]$ is the only maximizer of HED. For that purpose, assume that $\mathbf{z}$ is such that $HED(\mathbf{z}, P) = 1/2$. Then for any $\mathbf{u}\in\mathcal{S}^{d-1}$, the inequalities in (S.14)–(S.15) hold with $e = 1/2$ and are actually equalities (indeed, would there be a direction $\mathbf{u}$ for which at least one of these inequalities would be strict, then adding up both inequalities as above would provide $1 > 1$, a contradiction). Thus, for any $\mathbf{u}\in\mathcal{S}^{d-1}$, $\mathbf{u}'\mathbf{z}$ is the order-1/2 expectile of $\mathbf{u}'\mathbf{Z}$, that is,

$\mathbf{u}'\mathbf{z} = \mathrm{E}[\mathbf{u}'\mathbf{Z}]$ (see above). This means that $\mathbf{u}'(\mathbf{z} - \mathrm{E}[\mathbf{Z}]) = 0$ for any $\mathbf{u} \in \mathcal{S}^{d-1}$, which shows that $\mathbf{z} = \mathrm{E}[\mathbf{Z}]$. □

The proof of Theorem 7 still requires the following preliminary result.

Lemma S.3. *Let* $\mathbf{Z}$ *be a random* d*-vector with distribution* $P \in \mathcal{P}_d$ *and consider the functions* $(\mathbf{u}, \mathbf{z}) \mapsto \ell_-(\mathbf{u}, \mathbf{z})$ *and* $(\mathbf{u}, \mathbf{z}) \mapsto \ell_+(\mathbf{u}, \mathbf{z})$ *introduced in Lemma S.2. Then, (i) for any* $\mathbf{u} \in \mathcal{S}^{d-1}$*, the functions* $\mathbf{z} \mapsto \ell_\pm(\mathbf{u}, \mathbf{z})$ *admit, at any* $\mathbf{z} \in \mathbb{R}^d$*, directional derivatives in any direction; (ii) if, moreover,* P *is smooth in a neighbourhood of* $\mathbf{z}_0$ *(in the sense defined in Theorem 7), then, for any* $\mathbf{u} \in \mathcal{S}^{d-1}$*, the functions* $\mathbf{z} \mapsto \ell_\pm(\mathbf{u}, \mathbf{z})$ *are continuously differentiable in a neighbourhood of* $\mathbf{z}_0$*.*

Proof of Lemma S.3. (i) We will show that

$$\frac{\partial \ell_-}{\partial \mathbf{v}}(\mathbf{z}_0) = (\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z} < \mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v} < 0] + (\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z} \leq \mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v} > 0]. \tag{S.16}$$

To do so, note that, for any $h > 0$,

$$\begin{aligned}
&m_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h, \mathbf{y}) \\
&\quad := \frac{1}{h}\big\{\mathbf{u}'(\mathbf{z}_0 + h\mathbf{v} - \mathbf{y})\mathbb{I}[\mathbf{u}'\mathbf{y} < \mathbf{u}'(\mathbf{z}_0 + h\mathbf{v})] - \mathbf{u}'(\mathbf{z}_0 - \mathbf{y})\mathbb{I}[\mathbf{u}'\mathbf{y} < \mathbf{u}'\mathbf{z}_0]\big\} \\
&\qquad\quad - \big\{\mathbf{u}'\mathbf{v}\mathbb{I}[\mathbf{u}'\mathbf{y} < \mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v} < 0] + \mathbf{u}'\mathbf{v}\mathbb{I}[\mathbf{u}'\mathbf{y} \leq \mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v} > 0]\big\} \\
&\quad = \frac{1}{h}\,\big(\mathbf{u}'(\mathbf{z}_0 + h\mathbf{v}) - \mathbf{u}'\mathbf{y}\big)S(\mathbf{u}'\mathbf{v})\mathbb{I}[\mathbf{u}'\mathbf{y} \in \mathcal{I}_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h)],
\end{aligned}$$

where the sign function S was defined on page 4 and where $\mathcal{I}_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h)$ denotes the open interval with endpoints $\mathbf{u}'\mathbf{z}_0$ and $\mathbf{u}'(\mathbf{z}_0 + h\mathbf{v})$. This shows that, for any $\mathbf{y} \in \mathbb{R}^d$, $m_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h, \mathbf{y})$ converges to zero as h goes to zero from above and that the function $\mathbf{y} \mapsto |m_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h, \mathbf{y})|$ is upper-bounded by the function $\mathbf{y} \mapsto |\mathbf{u}'\mathbf{v}|$ that is P-integrable and does not depend on h.

Consequently, the Lebesgue Dominated Convergence Theorem entails that, as h goes to zero from above,

$$\frac{\ell_-(\mathbf{u},\mathbf{z}_0+h\mathbf{v})-\ell_-(\mathbf{u},\mathbf{z}_0)}{h}-\Big\{(\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z}<\mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v}<0]$$
$$+(\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z}\leq\mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v}>0]\Big\}=\int_{\mathbb{R}^d}m_{\mathbf{z}_0,\mathbf{u},\mathbf{v}}(h,\mathbf{y})\,dP(\mathbf{y})\to 0,$$

which establishes (S.16). The exact same reasoning allows to show that

$$\frac{\partial\ell_+}{\partial\mathbf{v}}(\mathbf{z}_0)=-(\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z}\geq\mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v}<0]-(\mathbf{u}'\mathbf{v})\mathbb{P}[\mathbf{u}'\mathbf{Z}>\mathbf{u}'\mathbf{z}_0]\mathbb{I}[\mathbf{u}'\mathbf{v}>0]. \qquad \text{(S.17)}$$

(ii) It trivially follows from the Lebesgue Dominated Convergence Theorem that, under the smoothness assumption considered, the functions $\mathbf{z}\mapsto\frac{\partial\ell_\pm}{\partial\mathbf{v}}(\mathbf{z})$ in (S.16)-(S.17) are continuous in a neighborhood of $\mathbf{z}_0$. $\square$

Proof of Theorem 7. Fix an arbitrary compact set $K\subset\mathbb{R}^d$ and $\varepsilon>0$. Since Lemma S.2(ii) implies that $(\mathbf{u},\mathbf{z})\mapsto G_{\mathbf{z}}(\mathbf{u})$ is continuous over the compact set $\mathcal{S}^{d-1}\times K$, it is also uniformly continuous on that set. Hence, there exists $\delta>0$ such that for any $\mathbf{u}_1,\mathbf{u}_2\in\mathcal{S}^{d-1}$ and $\mathbf{z}_1,\mathbf{z}_2\in K$ satisfying $\max(\|\mathbf{u}_1-\mathbf{u}_2\|,\|\mathbf{z}_1-\mathbf{z}_2\|)<\delta$, we have $|G_{\mathbf{z}_1}(\mathbf{u}_1)-G_{\mathbf{z}_2}(\mathbf{u}_2)|<\varepsilon$. For any $\mathbf{z}\in\mathbb{R}^d$, pick arbitrarily $\mathbf{u}_{\mathbf{z}}\in\mathcal{S}^{d-1}$ such that $HED(\mathbf{z},P)=G_{\mathbf{z}}(\mathbf{u}_{\mathbf{z}})$; existence follows from Theorem 8. Then, for any $\mathbf{z}_1,\mathbf{z}_2\in K$ with $\|\mathbf{z}_1-\mathbf{z}_2\|<\delta$, we have

$$HED(\mathbf{z}_1,P)=G_{\mathbf{z}_1}(\mathbf{u}_{\mathbf{z}_1})>G_{\mathbf{z}_2}(\mathbf{u}_{\mathbf{z}_1})-\varepsilon\geq HED(\mathbf{z}_2,P)-\varepsilon.$$

By symmetry, we also have $HED(\mathbf{z}_2,P)>HED(\mathbf{z}_1,P)-\varepsilon$, which yields $|HED(\mathbf{z}_2,P)-HED(\mathbf{z}_1,P)|<\varepsilon$. Consequently, $\mathbf{z}\mapsto HED(\mathbf{z},P)$ is uniformly continuous over K.

We now show that uniform continuity extends to $\mathbb{R}^d$. To do so, fix $\varepsilon>0$ and pick C large enough to have $HED(\mathbf{z},P)<\varepsilon/2$ as soon as $\mathbf{z}\notin B_0(C):=\{\mathbf{z}\in\mathbb{R}^d:\|\mathbf{z}\|\leq C\}$ (existence of C follows from the boundedness result in Theorem 2; we refer to the proof of

Theorem 4(iv) for details). Since $\mathbf{z} \mapsto HED(\mathbf{z}, P)$ is uniformly continuous over any $B_0(r)$, there exists $\delta > 0$ such that for any $\mathbf{z}_1, \mathbf{z}_2 \in B_0(C+1)$ such that $\|\mathbf{z}_1 - \mathbf{z}_2\| < \delta$, one has $|HED(\mathbf{z}_2, P) - HED(\mathbf{z}_1, P)| < \varepsilon$. Letting $\tilde{\delta} := \min(\delta, 1)$, it is then easy to check that for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$ such that $\|\mathbf{z}_1 - \mathbf{z}_2\| < \tilde{\delta}$, we must have $|HED(\mathbf{z}_2, P) - HED(\mathbf{z}_1, P)| < \varepsilon$ (note that as soon as one of such $\mathbf{z}_1, \mathbf{z}_2$ belongs to $B_0(C)$, then they both belong to $B_0(C+1)$).

(ii) Fix $\mathbf{z}_0 \in \mathbb{R}^d$ and $\mathbf{u}, \mathbf{v} \in \mathcal{S}^{d-1}$. Lemma S.3(i) implies that $\mathbf{z} \mapsto G_{\mathbf{z}}(\mathbf{u})$ admits a directional derivative at $\mathbf{z}_0$ in direction $\mathbf{v}$. In dimension $d = 1$, there are finitely many $\mathbf{u}$'s that are to be considered in (5), so that the aforementioned differentiability readily entails equidifferentiability in the sense of Milgrom and Segal (2002). The result then follows from Theorem 3 of Milgrom and Segal (2002). (iii) By assumption, P is smooth over a neighbourhood $\mathcal{N}$ of $\mathbf{z}_0$. Lemma S.3(ii) then yields that, for any $\mathbf{u} \in \mathcal{S}^{d-1}$, $\mathbf{z} \mapsto G_{\mathbf{z}}(\mathbf{u})$ is continuously differentiable over $\mathcal{N}$. The result then follows from Theorem 1 in Danskin (1966) or Proposition 1 in Demyanov (2009). □

The proof of Theorem 6 requires the following *strict* quasi-concavity property.

Lemma S.4. *Let P be a probability measure in $\mathcal{P}_d$ and denote as $\boldsymbol{\mu}(P)$ the corresponding mean vector. Then,*

$$HED((1-\lambda)\boldsymbol{\mu}(P) + \lambda \mathbf{z}, P) > HED(\mathbf{z}, P) \tag{S.18}$$

for any $\lambda \in [0, 1)$ and $z (\neq \boldsymbol{\mu}(P))$ in the c-support $C(P)$ of P.

Proof of Lemma S.4. Fix $\mathbf{z}_\lambda := (1-\lambda)\boldsymbol{\mu}(P) + \lambda \mathbf{z}$, with $\mathbf{z}(\neq \boldsymbol{\mu}(P)) \in C(P)$ and $\lambda \in (0, 1)$ (for $\lambda = 0$, the result directly follows from Theorem 5). Let $A := \{\mathbf{u} \in \mathcal{S}^{d-1} : \mathbf{u}'(\mathbf{z} - \boldsymbol{\mu}(P)) = 0\}$. First note that the proof of Theorem 5 entails that, for any $u \in A$,

$$G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}_\lambda) = G_{\mathbf{u}}(\mathbf{u}'\boldsymbol{\mu}(P)) = \frac{1}{2} = HED(\boldsymbol{\mu}(P), P) > HED(\mathbf{z}_\lambda, P),$$

so that (5) yields

$$HED(\mathbf{z}_\lambda, P) = \min_{u \in \mathcal{S}^{d-1} \setminus A} G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}_\lambda). \tag{S.19}$$

Fix then $\mathbf{u} \in \mathcal{S}^{d-1} \setminus A$. Since both $\mathbf{z}$ and $\boldsymbol{\mu}(P)$ belong to $C(P)$, we have $\mathbb{P}[\mathbf{u}'\mathbf{Z} \leq \min(\mathbf{u}'\boldsymbol{\mu}(P), \mathbf{u}'\mathbf{z})] > 0$ and $\mathbb{P}[\mathbf{u}'\mathbf{Z} \geq \max(\mathbf{u}'\boldsymbol{\mu}(P), \mathbf{u}'\mathbf{z})] > 0$. Recalling that $\mathbf{u} \notin A$, we have that $\mathbf{u}'\mathbf{z}_\lambda$ belongs to the open interval with endpoints $\mathbf{u}'\boldsymbol{\mu}(P)$ and $\mathbf{u}'\mathbf{z}$. Lemma S.1 then yields

$$\begin{aligned} G_{\mathbf{u}}(\mathbf{u}'\mathbf{z}_\lambda) > G_{\mathbf{u}}(\min(\mathbf{u}'\boldsymbol{\mu}(P), \mathbf{u}'\mathbf{z})) &= \min(G_{\mathbf{u}}(\mathbf{u}'\boldsymbol{\mu}(P)), G_{\mathbf{u}}(\mathbf{u}'\mathbf{z})) \\ &\geq \min(HED(\boldsymbol{\mu}(P), P), HED(\mathbf{z}, P)) = HED(\mathbf{z}, P) \end{aligned}$$

for any $\mathbf{u} \in \mathcal{S}^{d-1} \setminus A$. The result thus follows from (S.19). □

Proof of Theorem 6. Fix $0 \leq r_1 < r_2 < r_{\mathbf{u}}(P)$. Then, $\boldsymbol{\mu}(P) + r\mathbf{u} \in C(P)$ for any $r \in [0, r_2]$. Lemma S.4 thus yields $HED(\boldsymbol{\mu}(P) + r_1\mathbf{u}, P) > HED(\boldsymbol{\mu}(P) + r_2\mathbf{u}, P)$, so that $r \mapsto HED(\boldsymbol{\mu}(P) + r\mathbf{u}, P)$ is monotone strictly decreasing in $[0, r_{\mathbf{u}}(P))$. Now, fix $r > r_{\mathbf{u}}(P)$. By definition of $r_{\mathbf{u}}(P)$, $\boldsymbol{\theta} + r\mathbf{u} \notin C(P)$. Theorem 2 then ensures that there is no $\alpha \in (0, 1)$ for which $\boldsymbol{\theta} + r\mathbf{u} \in R_\alpha(P)$. Thus, by definition, $HED(\boldsymbol{\theta} + r\mathbf{u}, P) = 0$. Finally, continuity of $z \mapsto HED(z, P)$ (Theorem 7(i)) implies that $HED(\boldsymbol{\theta} + r_{\mathbf{u}}(P)\mathbf{u}, P) = 0$. □

Proof of Theorem 4. (i) The claim directly follows from the affine-equivariance result in Theorem 2. (ii) if $P (\in \mathcal{P}_d)$ is centrally symmetric about $\boldsymbol{\theta}$, then $\boldsymbol{\theta} = \mu(P)$, the mean vector of P, so that the result is a corollary of Theorem 5. (iii) This is a direct consequence of Theorem 6. (iv) Fix $\varepsilon > 0$. In view of Theorem 2, there exists $M > 0$ such that $R_\varepsilon(P)$ is included in $\{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| \leq M\}$. Consequently, Theorem 3 entails that, as soon as $\|\mathbf{z}\| > M$, one has $HED(\mathbf{z}, P) < \varepsilon$, as was to be shown. □

The proof of Theorem S.1 requires the following result.

Lemma S.5. *Let $\mathbf{Z}$ be a random d-vector with distribution $P \in \mathcal{P}_d$. Then, (i) there exist $c > 0$ and $\varepsilon > 0$ such that $\mathbb{P}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})| < c] \leq 1 - \varepsilon$ for any $\mathbf{u} \in \mathcal{S}^{d-1}$ and $\mathbf{z} \in \mathbb{R}^d$; (ii) $\inf_{(\mathbf{u},\mathbf{z}) \in \mathcal{S}^{d-1} \times \mathbb{R}^d} \mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|] > 0$.*

Proof of Lemma S.5. (i) First pick r so large that $\mathbb{P}[\|\mathbf{Z}\| \geq r/2] \leq 1/2$. For any $\mathbf{u} \in \mathcal{S}^{d-1}$ and $a > r$, we then have

$$\mathbb{P}[|\mathbf{u}'\mathbf{Z} - a| < r/2] \leq \mathbb{P}[\|\mathbf{Z}\| > r/2] \leq 1/2. \tag{S.20}$$

It is thus sufficient to show that there exist $c > 0$ and $\varepsilon > 0$ such that $\mathbb{P}[|\mathbf{u}'\mathbf{Z} - a| < c] \leq 1 - \varepsilon$ for any $\mathbf{u} \in \mathcal{S}^{d-1}$ and $a \in [0, r]$.

By contradiction, assume that for any $c > 0$ and $\varepsilon > 0$, there exist $\mathbf{u} \in \mathcal{S}^{d-1}$ and $a \in [0, r]$ such that $\mathbb{P}[|\mathbf{u}'\mathbf{Z} - a| < c] > 1 - \varepsilon$. We can thus construct a sequence $((\mathbf{u}_n, a_n))$ in $K = \mathcal{S}^{d-1} \times [0, r]$ such that

$$\mathbb{P}[|\mathbf{u}_n'\mathbf{Z} - a_n| < 1/n] > 1 - (1/n).$$

Compactness of K entails that there exists a subsequence $((\mathbf{u}_{n_\ell}, a_{n_\ell}))$ that converges in K, to $(\mathbf{u}_0, a_0)$ say. Clearly, we may assume that $(\mathbf{u}_{n_\ell}'\mathbf{u}_0)$ is a monotone non-decreasing sequence and that $(|a_{n_\ell} - a_0|)$ is a monotone non-increasing sequence (if that is not the case, one can always extract a further subsequence meeting these monotonicity properties). Let then $I_\ell := [a_0 - |a_{n_\ell} - a_0|, a_0 + |a_{n_\ell} - a_0|]$ and $C_\ell := \{\mathbf{u} \in \mathcal{S}^{d-1} : \mathbf{u}'\mathbf{u}_0 \geq \mathbf{u}_{n_\ell}'\mathbf{u}_0\}$. Note that the sequences of sets (I_ℓ) and (C_ℓ) are monotone non-increasing with respect to inclusion, with $\cap_\ell I_\ell = \{a_0\}$ and $\cap_\ell C_\ell = \{\mathbf{u}_0\}$. Therefore,

$$\lim_{\ell\to\infty} s_\ell := \lim_{\ell\to\infty} \mathbb{P}[\mathbf{Z} \in \cup_{a \in I_\ell} \cup_{\mathbf{u} \in C_\ell} \{\mathbf{y} : |\mathbf{u}'\mathbf{y} - a| \leq 1/n_\ell\}] = \mathbb{P}[\mathbf{u}_0'\mathbf{Z} - a_0 = 0].$$

But, for any ℓ, $s_\ell \geq \mathbb{P}[|\mathbf{u}_{n_\ell}'\mathbf{Z} - a_{n_\ell}| \leq 1/n_\ell] \geq 1 - (1/n_\ell)$, which implies that (s_ℓ) converges to one as ℓ diverges to infinity. Therefore, $\mathbb{P}[\mathbf{u}_0'\mathbf{Z} - a_0 = 0] = 1$, which, since $P \in \mathcal{P}_d$, is a contradiction.

(ii) Fix c and $\varepsilon > 0$ as in Part (i) of the lemma, that is, such that, letting $A_{\mathbf{u},\mathbf{z}} := \{\mathbf{y} \in \mathbb{R}^d : |\mathbf{u}'(\mathbf{y} - \mathbf{z})| \geq c\}$, we have $P[A_{\mathbf{u},\mathbf{z}}] \geq \varepsilon$ for any $\mathbf{u} \in \mathcal{S}^{d-1}$ and $\mathbf{z} \in \mathbb{R}^d$. Then,

$$\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|] \geq \int_{A_{\mathbf{u},\mathbf{z}}} |\mathbf{u}'(\mathbf{y} - \mathbf{z})| \, dP(\mathbf{y}) \geq c\varepsilon > 0$$

for any $\mathbf{u} \in \mathcal{S}^{d-1}$ and $\mathbf{z} \in \mathbb{R}^d$, which establishes the result. $\square$

PROOF OF THEOREM S.1. Let

$$m_{\mathbf{z},\mathbf{u}}(P) = G_{\mathbf{z}}(\mathbf{u}) = \frac{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})| \mathbb{I}[\mathbf{u}'(\mathbf{Z} - \mathbf{z}) \leq 0]]}{\mathrm{E}[|\mathbf{u}'(\mathbf{Z} - \mathbf{z})|]}.$$

For any $Q \in \mathcal{P}_d$, let $\mathbf{u}_{\mathbf{z}}(Q)$ be such that $HED(\mathbf{z}, Q) = m_{\mathbf{z},\mathbf{u}_{\mathbf{z}}(Q)}(Q)$ (existence follows from Theorem 8). Then,

$$\begin{aligned}
&|HED(\mathbf{z}, P_n) - HED(\mathbf{z}, P)| \, \mathbb{I}[HED(\mathbf{z}, P_n) \geq HED(\mathbf{z}, P)] \\
&\qquad \leq (m_{\mathbf{z},\mathbf{u}_{\mathbf{z}}(P)}(P_n) - m_{\mathbf{z},\mathbf{u}_{\mathbf{z}}(P)}(P)) \mathbb{I}[HED(\mathbf{z}, P_n) \geq HED(\mathbf{z}, P)] \\
&\qquad \leq \Big(\sup_{\mathbf{u}} |m_{\mathbf{z},\mathbf{u}}(P_n) - m_{\mathbf{z},\mathbf{u}}(P)| \Big) \mathbb{I}[HED(\mathbf{z}, P_n) \geq HED(\mathbf{z}, P)]
\end{aligned}$$

and

$$\begin{aligned}
&|HED(\mathbf{z}, P_n) - HED(\mathbf{z}, P)| \, \mathbb{I}[HED(\mathbf{z}, P_n) < HED(\mathbf{z}, P)] \\
&\qquad \leq (m_{\mathbf{z},\mathbf{u}_{\mathbf{z}}(P_n)}(P) - m_{\mathbf{z},\mathbf{u}_{\mathbf{z}}(P_n)}(P_n)) \mathbb{I}[HED(\mathbf{z}, P_n) < HED(\mathbf{z}, P)] \\
&\qquad \leq \Big(\sup_{\mathbf{u}} |m_{\mathbf{z},\mathbf{u}}(P) - m_{\mathbf{z},\mathbf{u}}(P_n)| \Big) \mathbb{I}[HED(\mathbf{z}, P_n) < HED(\mathbf{z}, P)]
\end{aligned}$$

(in this proof, all infima/suprema in $\mathbf{u}$ are over $\mathcal{S}^{d-1}$, whereas those in $\mathbf{z}$ are over K). Adding up these inequalities, we obtain

$$|HED(\mathbf{z}, P_n) - HED(\mathbf{z}, P)| \leq \sup_{\mathbf{u}} |m_{\mathbf{z},\mathbf{u}}(P_n) - m_{\mathbf{z},\mathbf{u}}(P)|,$$

which provides

$$\sup_{\mathbf{z}} |HED(\mathbf{z}, P_n) - HED(\mathbf{z}, P)| \leq \sup_{\mathbf{z},\mathbf{u}} |m_{\mathbf{z},\mathbf{u}}(P_n) - m_{\mathbf{z},\mathbf{u}}(P)|.$$

Now, writing $q^-_{\mathbf{z},\mathbf{u}}(P) := \mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) < 0]]$, $q^+_{\mathbf{z},\mathbf{u}}(P) := \mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|\mathbb{I}[\mathbf{u}'(\mathbf{Z}-\mathbf{z}) > 0]]$ and $q_{\mathbf{z},\mathbf{u}}(P) := q^-_{\mathbf{z},\mathbf{u}}(P) + q^+_{\mathbf{z},\mathbf{u}}(P) = \mathrm{E}[|\mathbf{u}'(\mathbf{Z}-\mathbf{z})|]$, we have

$$\begin{aligned}
|m_{\mathbf{z},\mathbf{u}}(P_n) - m_{\mathbf{z},\mathbf{u}}(P)| &= \Big| \frac{q^-_{\mathbf{z},\mathbf{u}}(P_n)}{q_{\mathbf{z},\mathbf{u}}(P_n)} - \frac{q^-_{\mathbf{z},\mathbf{u}}(P)}{q_{\mathbf{z},\mathbf{u}}(P)} \Big| \\
&\leq \frac{|q^-_{\mathbf{z},\mathbf{u}}(P_n) - q^-_{\mathbf{z},\mathbf{u}}(P)|}{q_{\mathbf{z},\mathbf{u}}(P_n)} + q^-_{\mathbf{z},\mathbf{u}}(P) \Big| \frac{1}{q_{\mathbf{z},\mathbf{u}}(P_n)} - \frac{1}{q_{\mathbf{z},\mathbf{u}}(P)} \Big| \\
&\leq \frac{|q^-_{\mathbf{z},\mathbf{u}}(P_n) - q^-_{\mathbf{z},\mathbf{u}}(P)| + |q_{\mathbf{z},\mathbf{u}}(P_n) - q_{\mathbf{z},\mathbf{u}}(P)|}{q_{\mathbf{z},\mathbf{u}}(P_n)} \\
&\leq \frac{2\sup_{\mathbf{z},\mathbf{u}} |q^-_{\mathbf{z},\mathbf{u}}(P_n) - q^-_{\mathbf{z},\mathbf{u}}(P)| + \sup_{\mathbf{z},\mathbf{u}} |q^+_{\mathbf{z},\mathbf{u}}(P_n) - q^+_{\mathbf{z},\mathbf{u}}(P)|}{\inf_{\mathbf{z},\mathbf{u}} q_{\mathbf{z},\mathbf{u}}(P) - \sup_{\mathbf{z},\mathbf{u}} |q_{\mathbf{z},\mathbf{u}}(P_n) - q_{\mathbf{z},\mathbf{u}}(P)|}
\end{aligned}$$

for any $z \in K$ and $u \in \mathcal{S}^{d-1}$. Since $\inf_{\mathbf{z},\mathbf{u}} q_{\mathbf{z},\mathbf{u}}(P) > 0$ (Lemma S.5), it only remains to prove that

$$\sup_{\mathbf{z},\mathbf{u}} |q^-_{\mathbf{z},\mathbf{u}}(P_n) - q^-_{\mathbf{z},\mathbf{u}}(P)| \stackrel{a.s.}{\to} 0 \quad \text{and} \quad \sup_{\mathbf{z},\mathbf{u}} |q^+_{\mathbf{z},\mathbf{u}}(P_n) - q^+_{\mathbf{z},\mathbf{u}}(P)| \stackrel{a.s.}{\to} 0 \tag{S.21}$$

as $n \to \infty$. Let us focus on the first convergence in (S.21). Clearly, we are after a Glivenko-Cantelli theorem for the classes of functions

$$\mathcal{G} := \Big\{ \mathbf{y} \mapsto g_{\mathbf{z},\mathbf{u}}(\mathbf{y}) := -(\mathbf{u}'(\mathbf{y}-\mathbf{z}))\mathbb{I}[\mathbf{u}'(\mathbf{y}-\mathbf{z}) < 0] : \mathbf{z} \in K, \mathbf{u} \in \mathcal{S}^{d-1} \Big\}$$

(the restriction to a compact K for z ensures that this class posesses an integrable envelope). The collection $\mathcal{H}$ of all halfspaces in $\mathbb{R}^{d+1}$ is a Vapnik-Chervonenkis class; see, e.g., Page 152 of Van der Vaart and Wellner (1996). Consequently, defining the subgraph of a function $f : \mathbb{R}^d \to \mathbb{R}$ as $s_f := \{(\mathbf{y}, t) \in \mathbb{R}^{d+1} : t < f(\mathbf{y})\}$ and letting

$$\mathcal{F} := \Big\{ \mathbf{y} \mapsto f_{\mathbf{z},\mathbf{u}}(\mathbf{y}) := \mathbf{u}'(\mathbf{y} - \mathbf{z}) : \mathbf{z} \in K, \mathbf{u} \in \mathcal{S}^{d-1} \Big\},$$

the collection of subgraphs $\{s_f : f \in \mathcal{F}\}$, as a subset of $\mathcal{H}$, is a Vapnik-Chervonenkis class. In other words, $\mathcal{F}$ is a VC-subgraph class (see, e.g., Section 2.6.2 of Van der Vaart and Wellner, 1996). Now, since $t \mapsto -t\mathbb{I}[t \leq 0]$ is a monotone function, Lemma 2.6.18(viii) of Van der Vaart and Wellner (1996) implies that $\mathcal{G}$ is itself a VC-subgraph class, hence is Glivenko-Cantelli, which implies the first convergence in (S.21). Since the same reasoning establishes the second convergence in (S.21), the result is proved. $\square$

Proof of Theorem S.2. In view of Theorem S.1 and the result in Theorem 4.5 from Dyckerhoff (2016) (more precisely, its corollary in a random sampling scheme as discussed in page 13 of that paper), it is sufficient to prove that $HED(\,\cdot\,, P)$ is *strictly monotone*, in the sense that, for any $\alpha \in (0, \alpha_*)$, with $\alpha_* := \max_{\mathbf{y} \in \mathbb{R}^d} HED(\mathbf{y}, P) = \frac{1}{2}$, the region $R_\alpha(P)$ is the closure $\bar{R}_{\alpha,>}(P)$ of $R_{\alpha,>}(P) := \{\mathbf{z} \in \mathbb{R}^d : HED(\mathbf{z}, P) > \alpha\}$.

Now, since $R_\alpha(P)$ is closed and contains $R_{\alpha,>}(P)$, we have that $\bar{R}_{\alpha,>}(P) \subset R_\alpha(P)$. To show that $R_\alpha(P) \subset \bar{R}_{\alpha,>}(P)$, fix $\mathbf{z} \in R_\alpha(P)$. If $HED(\mathbf{z}, P) > \alpha$, then $\mathbf{z}$ trivially belongs to $\bar{R}_{\alpha,>}(P)$, so that we may assume that $HED(\mathbf{z}, P) = \alpha$. Consider then the line segment associated with $\mathbf{z}_\lambda := (1 - \lambda)\mu(P) + \lambda\mathbf{z}$, $\lambda \in (0, 1)$, from the mean vector $\mu(P)$ of P (the deepest point of $HED(\,\cdot\,, P)$) to $\mathbf{z}$. Theorem 6 guarantees that $(\mathbf{z}_{1-(1/n)})$ is a sequence in $R_{\alpha,>}(P)$ that converges to $\mathbf{z}$, so that $\mathbf{z} \in \bar{R}_{\alpha,>}(P)$. We conclude that we also have $R_\alpha(P) \subset \bar{R}_{\alpha,>}(P)$, hence that $HED(\,\cdot\,, P)$ is strictly monotone. This establishes the result. $\square$

Proof of Theorem 9. From affine invariance, there is no loss of generality in assuming that $\mathbf{z} = 0$, $\mathbf{u}_0 = (0, \ldots, 0, 1)' \in \mathbb{R}^d$ and that the path $\mathbf{u}_t$ is of the form $\mathbf{u}_t =$

$(0, \ldots, 0, \cos(t + \frac{\pi}{2}), \sin(t + \frac{\pi}{2}))'$, $t \in [0, \pi]$. For any $t \in [0, \pi]$, we have

$$\gamma(t) := e_0(\mathbf{u}_t) = \frac{-h_<(t)}{-h_<(t) + h_>(t)} = \frac{h_<(t)}{h_<(t) - h_>(t)},$$

with

$$h_<(t) := \int_{\mathbb{R}^d} \mathbf{u}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} < 0]\, dP(\mathbf{y}) \quad \text{and} \quad h_>(t) := \int_{\mathbb{R}^d} \mathbf{u}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} > 0]\, dP(\mathbf{y}).$$

Throughout the proof, we will use the notation $\boldsymbol{\mu} = \mathrm{E}[\mathbf{Z}]$, $\boldsymbol{\mu}_{t,<} := \mathrm{E}[\mathbf{Z}\mathbb{I}[\mathbf{u}_t'\mathbf{Z} < 0]]$, and $\boldsymbol{\mu}_{t,>} := \mathrm{E}[\mathbf{Z}\mathbb{I}[\mathbf{u}_t'\mathbf{Z} > 0]]$. Note that under the assumptions of the theorem, we have $\boldsymbol{\mu} = \boldsymbol{\mu}_{0,>} + \boldsymbol{\mu}_{0,<}$.

We start by considering differentiability of (a) $h_<(t) = \mathbf{u}_t'\boldsymbol{\mu}_{t,<}$ and (b) $h_>(t) = \mathbf{u}_t'\boldsymbol{\mu}_{t,>}$. (a) Since $P[\Pi \setminus \{0\}] = 0$ for any hyperplane Π containing 0, the mapping $t \mapsto \mathbf{u}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} < 0]$ is P-almost everywhere differentiable at any $t \in [0, \pi]$, with derivative $t \mapsto \dot{\mathbf{u}}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} < 0]$, where we let $\dot{\mathbf{u}}_t := (0, \ldots, 0, -\sin(t+\frac{\pi}{2}), \cos(t+\frac{\pi}{2}))'$. Since the function $(t, \mathbf{y}) \mapsto \dot{\mathbf{u}}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} < 0]$ is upper-bounded by the t-independent P-integrable function $\mathbf{y} \mapsto \|\mathbf{y}\|$, the mapping $t \mapsto h_<(t)$ is differentiable at any $t \in [0, \pi]$, with derivative $\dot{h}_<(t) := \dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,<}$. (b) Similarly, for any $\mathbf{y} \in \mathbb{R}^d$, the mapping $t \mapsto \mathbf{u}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} > 0]$ is P-almost everywhere differentiable at any $t \in [0, \pi]$, with derivative $t \mapsto \dot{\mathbf{u}}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} > 0]$. Since the function $(t, \mathbf{y}) \mapsto \dot{\mathbf{u}}_t'\mathbf{y}\mathbb{I}[\mathbf{u}_t'\mathbf{y} > 0]$ is still upper-bounded by the t-independent P-integrable function $\mathbf{y} \mapsto \|\mathbf{y}\|$, the mapping $t \mapsto h_>(t)$ is differentiable at any $t \in [0, \pi]$, with derivative $\dot{h}_>(t) := \dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,>}$.

We conclude that $t \mapsto \gamma(t)$ is differentiable at any $t \in [0, \pi]$, with a derivative $\dot{\gamma}(t)$ that satisfies

$$\begin{aligned}(h_<(t) - h_>(t))^2\dot{\gamma}(t) &= \dot{h}_<(t)(h_<(t) - h_>(t)) - h_<(t)(\dot{h}_<(t) - \dot{h}_>(t)) \\ &= h_<(t)\dot{h}_>(t) - h_>(t)\dot{h}_<(t) = (\mathbf{u}_t'\boldsymbol{\mu}_{t,<})(\dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,>}) - (\mathbf{u}_t'\boldsymbol{\mu}_{t,>})(\dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,<}).\end{aligned}$$

Let us introduce some further notation. For any $t \in [0, \pi]$, write the projections of $\boldsymbol{\mu}_{t,<}$, $\boldsymbol{\mu}_{t,>}$, and $\boldsymbol{\mu}$ onto the plane spanned by the last two vectors of the canonical basis of $\mathbb{R}^d$ as $(0, \ldots, 0, r_{t,<} \cos \alpha_{t,<}, r_{t,<} \sin \alpha_{t,<})'$, $(0, \ldots, 0, r_{t,>} \cos \alpha_{t,>}, r_{t,>} \sin \alpha_{t,>})'$, and $(0, \ldots, 0, r \cos \alpha, r \sin \alpha)'$, respectively, where all r's are nonnegative and all α's belong to $[0, 2\pi)$. Since it is assumed that $HED(\mathbf{z}, P) > 0$, we must have

$$r_{t,>} > 0 \quad \text{and} \quad \alpha_{t,>} \in (t, t + \pi) \tag{S.22}$$

for any $t \in [0, \pi]$ and

$$r_{t,<} > 0 \quad \text{and} \quad \alpha_{t,<} \in [0, t) \cup (t + \pi, 2\pi) \tag{S.23}$$

for any $t \in [0, \pi]$. Note that $t \mapsto \alpha_{t,>}$ is monotone non-decreasing over $[0, \pi]$ and that $t \mapsto \alpha_{t,<}$ is monotone non-decreasing "modulo 2π" over the same range. Finally, note also that

$$e_0(\mathbf{u}_0) = \frac{\mathrm{E}[|\mathbf{u}_0'\mathbf{Z}|\mathbb{I}[\mathbf{u}_0'\mathbf{Z} \leq 0]]}{\mathrm{E}[|\mathbf{u}_0'\mathbf{Z}|]} \leq \frac{1}{2} \tag{S.24}$$

(if not, then $e_0(\mathbf{u}_\pi) = 1 - e_0(\mathbf{u}_0) < e_0(\mathbf{u}_0)$, which contradicts the definition of $\mathbf{u}_0$). If $e_0(\mathbf{u}_0) = 1/2$, then $e_0(\mathbf{u}_t) = 1/2$ for any $t \in [0, \pi]$ (if $e_0(\mathbf{u}) > 1/2$ for some $\mathbf{u} \in \mathcal{C}$, then $e_0(-\mathbf{u}) = 1 - e_0(\mathbf{u}) < 1/2 = e_0(\mathbf{u}_0)$, a contradiction), so that the result holds with $t_a = t_b = \pi$. We may thus assume that the inequality in (S.24) is strict, which implies that $\mathbf{u}_0'\boldsymbol{\mu} = \mathrm{E}[\mathbf{u}_0'\mathbf{Z}] > 0$, hence that $r > 0$ and $\alpha \in [0, \pi]$.

With the notation introduced above, we have

$$
\begin{aligned}
(h_<(t) - h_>(t))^2\dot{\gamma}(t) &= (\mathbf{u}_t'\boldsymbol{\mu}_{t,<})(\dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,>}) - (\mathbf{u}_t'\boldsymbol{\mu}_{t,>})(\dot{\mathbf{u}}_t'\boldsymbol{\mu}_{t,<}) \\
&= r_{t,<}r_{t,>}(\cos(t+\tfrac{\pi}{2})\cos\alpha_{t,<} + \sin(t+\tfrac{\pi}{2})\sin\alpha_{t,<}) \\
&\qquad\qquad \times(-\sin(t+\tfrac{\pi}{2})\cos\alpha_{t,>} + \cos(t+\tfrac{\pi}{2})\sin\alpha_{t,>}) \\
&\quad - r_{t,<}r_{t,>}(\cos(t+\tfrac{\pi}{2})\cos\alpha_{t,>} + \sin(t+\tfrac{\pi}{2})\sin\alpha_{t,>}) \\
&\qquad\qquad \times(-\sin(t+\tfrac{\pi}{2})\cos\alpha_{t,<} + \cos(t+\tfrac{\pi}{2})\sin\alpha_{t,<}) \\
&= r_{t,<}r_{t,>}\cos(\alpha_{t,<} - (t+\tfrac{\pi}{2}))\sin(\alpha_{t,>} - (t+\tfrac{\pi}{2})) \\
&\quad - r_{t,<}r_{t,>}\cos(\alpha_{t,>} - (t+\tfrac{\pi}{2}))\sin(\alpha_{t,<} - (t+\tfrac{\pi}{2})) \\
&= r_{t,<}r_{t,>}\sin(\alpha_{t,>} - \alpha_{t,<}) =: \ell(t).
\end{aligned}
$$

Now, since $\mathbf{u}_0$ is a minimizer of $e_0(\cdot)$ on $\mathcal{C}$, $\ell(0) = -r_{0,<}r_{0,>}\sin(\alpha_{0,<} - \alpha_{0,>}) = 0$. Since (S.22)-(S.23) entail that $\alpha_{0,<} > \alpha_{0,>}$, this yields $\alpha_{0,<} = \alpha_{0,>} + \pi$. By using the identity $\boldsymbol{\mu} = \boldsymbol{\mu}_{0,>} + \boldsymbol{\mu}_{0,<}$ and the fact that $\alpha \in [0,\pi]$, we conclude that $\alpha = \alpha_{0,>} \in (0,\pi)$. Similarly, using the fact that $\mathbf{u}_\pi = -\mathbf{u}_0$ is a maximizer of $e_0(\cdot)$ on $\mathcal{C}$ (this follows from the fact that $e_0(-\mathbf{u}) = 1 - e_0(\mathbf{u})$ for any $\mathbf{u}$), we have $\ell(\pi) = -r_{\pi,<}r_{\pi,>}\sin(\alpha_{\pi,<} - \alpha_{\pi,>}) = 0$, which implies that $\alpha_{\pi,>} = \alpha_{\pi,<} + \pi$ (recall that we cannot have $\alpha_{\pi,<} = \pi$, nor 0). Thus $\pi < \alpha_{\pi,>} < 2\pi$. By using the identity $\boldsymbol{\mu} = \boldsymbol{\mu}_{\pi,>} + \boldsymbol{\mu}_{\pi,<}$ and the fact that $\alpha \in (0,\pi)$, we conclude that $\alpha_{\pi,<} = \alpha$.

Now, fix $t_0 \in [0,\pi]$ with $\ell(t_0) = -r_{t_0,<}r_{t_0,>}\sin(\alpha_{t_0,<} - \alpha_{t_0,>}) \neq 0$ (if there is no such t_0, then the result holds with $t_a = t_b = \pi$). Monotonicity of $t \mapsto \alpha_{t,>}$ yields $\alpha_{t_0,>} \geq \alpha_{0,>} = \alpha$. Since $\alpha_{t_0,>} = \alpha$ would lead to $\alpha_{t_0,<} = \alpha_{t_0,>} + \pi$ (due to $\mu = \mu_{t_0,>} + \mu_{t_0,<}$), hence to $\ell(t_0) = 0$, we must actually have $\alpha_{t_0,>} > \alpha$.

Pick then an arbitrary $t \in [t_0, \pi)$. Monotonicity of $t \mapsto \alpha_{t,>}$ then yields $\alpha_{t,>} \geq \alpha_{t_0,>} > \alpha$, hence $\alpha_{t,>} \in (\alpha, t+\pi)$. Since $\boldsymbol{\mu} = \boldsymbol{\mu}_{t,>} + \boldsymbol{\mu}_{t,<}$, we must have $\alpha_{t,<} \in (\alpha_{t,>}+\pi, t+2\pi] \mod 2\pi$, that is, (a) $\alpha_{t,<} \in (\alpha_{t,>}+\pi, 2\pi)$ or (b) $\alpha_{t,<} \in [0, t)$. In case (a), we have $\alpha_{t,<} - \alpha_{t,>} \in (\pi, 2\pi)$, so that $\ell(t) = -r_{t,<} r_{t,>} \sin(\alpha_{t,<} - \alpha_{t,>}) > 0$. In case (b), in view of the monotonicity (modulo 2π) of $t \mapsto \alpha_{t,<}$, we have $\alpha_{t,<} \in [0, \alpha_{\pi,<} = \alpha]$. Therefore, the identity $\boldsymbol{\mu} = \boldsymbol{\mu}_{t,>} + \boldsymbol{\mu}_{t,<}$ implies that $\alpha_{t,<} < t < \alpha_{t,>} \leq \alpha_{t,<} + \pi$. Therefore, $\ell(t) = r_{t,<} r_{t,>} \sin(\alpha_{t,>} - \alpha_{t,<}) \geq 0$.

Assume that $\ell(t) = 0$. As shown above, we must thus be in case (b). Then $\alpha_{t,>} = \alpha_{t,<} + \pi$, so that (still due to $\boldsymbol{\mu} = \boldsymbol{\mu}_{t,>} + \boldsymbol{\mu}_{t,<}$) we must have $\alpha_{t,<} = \alpha = \alpha_{\pi,<}$. Monotonicity then implies that for any $t' \in [t, \pi)$, we have $\alpha_{t',<} = \alpha$, which, in view of $\boldsymbol{\mu} = \boldsymbol{\mu}_{t',>} + \boldsymbol{\mu}_{t',<}$ yields $\alpha_{t',>} = \alpha + \pi$. Consequently, we have $\ell(t') = 0$, which establishes the result. $\square$

Proof of Theorem S.3. (i) First note that since $\mathbf{u} \in \mathbb{R}^d_+$, we have $\mathbf{u}'\mathbf{X} \leq \mathbf{u}'\mathbf{Y}$ almost surely, so that the monotonicity of (univariate) expectiles entails that $e_\alpha(\mathbf{u}'\mathbf{X}) \leq e_\alpha(\mathbf{u}'\mathbf{Y})$. It trivially follows that $H_{\alpha,\mathbf{u}}(\mathbf{Y}) \subset H_{\alpha,\mathbf{u}}(\mathbf{X}) \subset H_{\alpha,\mathbf{u}}(\mathbf{X}) \oplus \mathbb{R}^d_+$. To establish the other inclusion, fix $\mathbf{z} \in H_{\alpha,\mathbf{u}}(\mathbf{X})$. We may assume that $\mathbf{z} \notin H_{\alpha,\mathbf{u}}(\mathbf{Y})$ (if $\mathbf{z} \in H_{\alpha,\mathbf{u}}(\mathbf{Y})$, then $\mathbf{z} = \mathbf{z} + 0 \in H_{\alpha,\mathbf{u}}(\mathbf{Y}) \oplus \mathbb{R}^d_-$). We then have

$$\mathbf{z} = (e_\alpha(\mathbf{u}'\mathbf{Y})\mathbf{u} + (I_d - \mathbf{u}\mathbf{u}')\mathbf{z}) + (\mathbf{u}'\mathbf{z} - e_\alpha(\mathbf{u}'\mathbf{Y}))\mathbf{u} =: \mathbf{z}_0 + \mathbf{z}_1.$$

Since $\mathbf{u}'\mathbf{z}_0 = e_\alpha(\mathbf{u}'\mathbf{Y})$, we have that $\mathbf{z}_0 \in H_{\alpha,\mathbf{u}}(\mathbf{Y})$. Since $\mathbf{u}'\mathbf{z} - e_\alpha(\mathbf{u}'\mathbf{Y}) < 0$ (recall that $\mathbf{z} \notin H_{\alpha,\mathbf{u}}(\mathbf{Y})$) and $\mathbf{u} \in \mathbb{R}^d_+$, we also have $\mathbf{z}_1 \in \mathbb{R}^d_-$. This shows that $H_{\alpha,\mathbf{u}}(\mathbf{X}) \subset H_{\alpha,\mathbf{u}}(\mathbf{Y}) \oplus \mathbb{R}^d_-$. (ii) Let $\mathbf{z} \in H_{\alpha,\mathbf{u}}(\mathbf{X}+\mathbf{Y})$ and decompose it into $\mathbf{z} = (e_\alpha(\mathbf{u}'\mathbf{X})\mathbf{u} + (I_d - \mathbf{u}\mathbf{u}')\mathbf{z}) + (\mathbf{u}'\mathbf{z} - e_\alpha(\mathbf{u}'\mathbf{X}))\mathbf{u} =: \mathbf{z}_0 + \mathbf{z}_1$. Obviously, $\mathbf{z}_0 \in H_{\alpha,\mathbf{u}}(\mathbf{X})$. As for $\mathbf{z}_1$, the superadditivity of univariate expectiles for $\alpha \in (0, \frac{1}{2}]$ implies that $\mathbf{u}'\mathbf{z}_1 = \mathbf{u}'\mathbf{z} - e_\alpha(\mathbf{u}'\mathbf{X}) \geq e_\alpha(\mathbf{u}'(\mathbf{X}+\mathbf{Y})) - e_\alpha(\mathbf{u}'\mathbf{X}) \geq e_\alpha(\mathbf{u}'\mathbf{Y})$, which shows that $\mathbf{z}_1 \in H_{\alpha,\mathbf{u}}(\mathbf{Y})$. (iii) If $\mathbf{z}_0 \in H_{\alpha,\mathbf{u}}(\mathbf{X})$ and $\mathbf{z}_1 \in H_{\alpha,\mathbf{u}}(\mathbf{Y})$, then the subadditivity of univariate expectiles for $\alpha \in [\frac{1}{2}, 1)$ readily yields $\mathbf{u}'(\mathbf{z}_0 + \mathbf{z}_1) \geq e_\alpha(\mathbf{u}'\mathbf{X}) + e_\alpha(\mathbf{u}'\mathbf{Y}) \geq e_\alpha(\mathbf{u}'(\mathbf{X}+\mathbf{Y}))$, which shows that $\mathbf{z}_0 + \mathbf{z}_1 \in H_{\alpha,\mathbf{u}}(\mathbf{X}+\mathbf{Y})$. (iv) The result trivially follows from the monotonicity of univariate expectiles with respect to their order α. $\square$

References

Acerbi, C. (2002), "Spectral measures of risk: A coherent representation of subjective risk aversion," *J. Banking & Finance*, 26, 1505–1518.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999), "Coherent Measures of Risk," *Math. Finance*, 9, 203–228.

Bellini, F., Klar, B., Müller, A., and Gianina, E. (2014), "Generalized quantiles as risk measures," *Insurance Math. Econom.*, 54, 41–48.

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.

Cascos, I. and Molchanov, I. (2007), "Multivariate risks and depth-trimmed regions," *Finance Stoch.*, 11, 373–397.

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

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.

Danskin, J. M. (1966), "The theory of Max-Min, with applications," *SIAM J. Appl. Math.*, 14, 641–664.

Delbaen, F. (2002), "Coherent risk measures on general probability spaces," in *Advances in Finance and Stochastics*, eds. Sandmann, K. and Schönbucher, P., Berlin: Springer, pp. 1–37.

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

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

Dyckerhoff, R. (2016), "Convergence of depths and depth-trimmed regions," *ArXiv preprint arXiv:1611.08721v2*.

Dyckerhoff, R. and Mosler, K. (2011), "Weighted-mean trimming of multivariate data," *J. Multivariate Anal.*, 102, 405–421.

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

Gneiting, T. (2011), "Making and evaluating point forecasts," *J. Amer. Statist. Assoc.*, 106, 746–762.

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

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.

Jouini, E., Meddeb, M., and Touzi, N. (2004), "Vector-valued coherent risk measures," *Finance Stoch.*, 8, 531–552.

Kim, J. (2000), "Rate of convergence of depth contours: with application to a multivariate metrically trimmed mean," *Statist. Probab. Lett.*, 49, 393–400.

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

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

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer Series in Statistics, New York: Springer.

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.

Massé, J.-C. (2002), "Asymptotics for the Tukey Median," *J. Multivariate Anal.*, 81, 286–300.

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.

Milgrom, P. and Segal, I. (2002), "Envelope theorems for arbitrary choice sets," *Econometrica*, 79, 583–601.

Van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer Series in Statistics, New York: Springer.

Ziegel, J. (2016), "Coherence and elicitability," *Math. Finance*, 26, 901–918.

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