12

Multiple-Output Quantile Regression

Marc Hallin

ECARES, Université Libre de Bruxelles, Belgium

Miroslav Šiman

Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic

CONTENTS

12.1 Multivariate quantiles, and the ordering of $\mathbb{R}^d$, $d \geqslant 2$

Quantile regression is about estimating the quantiles of some d-dimensional response $\mathbf{Y}$ conditional on the values $\mathbf{x} \in \mathbb{R}^p$ of some covariates $\mathbf{X}$. The problem is well understood when $d = 1$ (*single-output* case, where Y is used instead of $\mathbf{Y}$): for a (conditional) probability distribution $\mathrm{P}^Y = \mathrm{P}^{\mathrm{Y}}_{\mathbf{X}=\mathbf{x}}$ on $\mathbb{R}$, with distribution function $F = F_{\mathbf{X}=\mathbf{x}}$, the (conditional on $\mathbf{X} = \mathbf{x}$) quantile of order τ of $\mathbf{Y}$ is

$$q_\tau(\mathbf{x}) := \inf\{y \ : \ F(y) \geqslant \tau\}, \quad \tau \in [0, 1).$$

This, under absolute continuity with nonvanishing density (which, for simplicity, we henceforth assume throughout), yields, for $\tau \in (0, 1)$, what we will call the "traditional definition:"

(a) *Regression quantiles: traditional definition.* The *regression quantile of order* τ of Y (relative to the vector of covariates $\mathbf{X}$ with values in $\mathbb{R}^p$) is the mapping

$$\mathbf{x} \mapsto q_\tau(\mathbf{x}) := F^{-1}(\tau), \quad \mathbf{x} \in \mathbb{R}^p, \tag{12.1}$$

where $F(y) := \mathrm{P}(Y \leqslant y | \mathbf{X} = \mathbf{x})$.

The same concept, under finite moments of order 1, also admits the L_1 characterization:

(b) *Regression quantiles: L_1 definition.* The *regression quantile of order* τ of Y (relative to the vector of covariates $\mathbf{X}$ with values in $\mathbb{R}^p$) is the mapping $\mathbf{x} \mapsto q_\tau(\mathbf{x})$, where $q_\tau(\mathbf{x})$ minimizes, for $\mathbf{x} \in \mathbb{R}^p$,

$$\mathrm{E}[\rho_\tau(Y - q) | \mathbf{X} = \mathbf{x}] \tag{12.2}$$

over $q \in \mathbb{R}$; the function $z \mapsto \rho_\tau(z) := (1 - \tau)|z|\, I_{[z<0]} + \tau z\, I_{[z \geqslant 0]}$, as usual, stands for the so-called *check function.*

Neither this L_1 definition (b), nor the "traditional" one (a), leads to a straightforward empirical version (as there are no empirical versions of conditional distributions). The L_1 characterization, however, allows for a linear version of quantile regression:

(c) *Linear regression quantiles.* The *regression quantile hyperplane of order* τ of Y (relative to the vector of covariates $\mathbf{X}$ with values in $\mathbb{R}^p$) is the hyperplane with equation

$$y = q_\tau(\mathbf{x}) = \alpha_\tau + \boldsymbol{\beta}'_\tau \mathbf{x},$$

where α_τ and $\boldsymbol{\beta}_\tau$ are the minimizers, over $(a, \mathbf{b}') \in \mathbb{R}^{p+1}$, of

$$\mathrm{E}[\rho_\tau(Y - a - \mathbf{b}'\mathbf{X})]. \tag{12.3}$$

Contrary to the general concept defined in (a)–(b), regression quantile hyperplanes, thanks to their parametric form, also admit straightforward (merely substitute empirical distributions for the theoretical ones) empirical counterparts:

(d) *Empirical linear regression quantiles.* Denote by $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$ an n-tuple of points in $\mathbb{R}^{p+1}$. The corresponding empirical regression quantile hyperplane of order τ is the hyperplane with equation

$$y = q^{(n)}_\tau(\mathbf{x}) = \alpha^{(n)}_\tau + \boldsymbol{\beta}^{(n)\prime}_\tau \mathbf{x},$$

where $\alpha^{(n)}_\tau$ and $\boldsymbol{\beta}^{(n)}_\tau$ are the minimizers, over $(a, \mathbf{b}') \in \mathbb{R}^{p+1}$, of

$$\sum_{i=1}^{n} \rho_\tau(Y_i - a - \mathbf{b}'\mathbf{X}_i). \tag{12.4}$$

Note that, for $p = 0$, all definitions above yield "location quantiles," as opposed to "regression quantiles."

Now, a response seldom comes as an isolated quantity, and $\mathbf{Y}$, in most situations of practical interest, takes values in $\mathbb{R}^d$, with $d \geqslant 2$ (*multiple-output* case). An extension to $d \geqslant 2$ of the definitions above is thus extremely desirable. Unfortunately, those definitions all exploit the canonical ordering of $\mathbb{R}$. Such an ordering no longer exists in $\mathbb{R}^d$, $d \geqslant 2$. As a consequence, (location and regression) quantiles and check functions, but also equally basic univariate concepts such as distribution functions, signs, and ranks – all playing a fundamental role in statistical inference – do not straightforwardly extend to higher dimensions.

That problem of ordering $\mathbb{R}^d$ – hence that of defining pertinent concepts of multivariate quantiles, ranks and signs – has attracted much interest in the literature, and many solutions have been proposed. The whole theory of statistical depth and also, in a sense, the theory of copulas, have a similar objective. For obvious reasons of space constraints, we cannot provide here an extensive coverage of those theories. For insightful and extensive surveys of statistical depth, we refer to Zuo and Serfling (2000) or Serfling (2002a, 2012).

12.2 Directional approaches

Since the univariate concept of a quantile is well understood and carries all the properties one is expecting, a natural idea, in dimension $d \geqslant 2$, involves trying to reduce the multivariate problem to a collection of univariate ones by considering univariate distributions (such as marginal or projected distributions) associated with the d-dimensional ones. This is what we call a *directional approach*, as opposed to *global* approaches, where definitions are of a direct nature. We start with the pure location case ($p = 0$, no covariates) and projection ideas.

12.2.1 Projection methods

12.2.1.1 Marginal (coordinatewise) quantiles

If a coordinate system is adopted, $\mathbf{Y}$ is written as $(Y_1, \ldots, Y_d)'$. The d marginal distribution functions characterize marginal quantiles $q_{\tau;j}$, $j = 1, \ldots, d$, hence a coordinatewise multivariate quantile

$$\mathbf{q}_{\tau_1,\ldots,\tau_d} := (q_{\tau_1;1}, \ldots, q_{\tau_d;d})'.$$

The mapping $(\tau_1, \ldots, \tau_d) \mapsto \mathbf{q}_{\tau_1,\ldots,\tau_d}$ is actually the inverse of the copula transform; in particular, $\mathbf{q}_{1/2,\ldots,1/2}$ yields the componentwise median. An empirical version of that mapping is readily obtained by considering the empirical marginal distributions of any observed n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$.

This definition, however, does not provide an ordering of $\mathbb{R}^d$, but rather a d-tuple of (marginal) orderings. Moreover, $\mathbf{q}_{\tau_1,\ldots,\tau_d}$ crucially depends on the coordinate system adopted, and is not even rotation-equivariant.

The concept being unsatisfactory in the location case, its regression extensions will not be examined.

12.2.1.2 Quantile biplots

Marginal quantiles actually are obtained by projecting P (equivalently, $\mathbf{Y}$) on d mutually orthogonal straight lines (characterized by the canonical orthonormal basis $(\mathbf{u}_1, \ldots, \mathbf{u}_d)$)

through some origin. If the influence of the arbitrary choice of a basis is to be removed, one might also like to look at projections onto *all* unit vectors $\mathbf{u} \in \mathcal{S}^{d-1}$ through some given origin. This approach is investigated in Kong and Mizera (2008) and, in some detail, in Ahidar (2015, Section 2); see also Ahidar-Coutrix and Berthet (2016).

For each $\mathbf{u}$ and $\tau \in (0,1)$, the univariate distribution of $\mathbf{u}'\mathbf{Y}$ yields a well-defined quantile of order τ ($q_{\tau\mathbf{u}}$, say). Define, for $\tau \in [1/2, 1)$, the *directional quantile* of order τ for direction $\mathbf{u}$ as the point $\mathbf{q}_{\tau\mathbf{u}} := q_{\tau\mathbf{u}}\mathbf{u}$. Substituting empirical quantiles $q^{(n)}_{\tau\mathbf{u}}$ for the theoretical ones yields empirical counterparts $\mathbf{q}^{(n)}_{\tau\mathbf{u}}$. The collection, for $\mathbf{u}$ ranging over the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$, of all those directional quantiles yields what Kong and Mizera (2008) call a *quantile biplot*.

Intuitively appealing as it may be, however, this concept exhibits somewhat weird properties: quantile biplots are very sensitive to the (arbitrary) choice of an origin; they are neither translation- nor rotation-equivariant, and yield strange, often self-intersecting contours. Although the computation of each particular $\mathbf{q}^{(n)}_{\tau\mathbf{u}}$ is quite straightforward, the construction of empirical biplots, in principle, requires considering "infinitely many" directions $\mathbf{u}$. For all those reasons, the concept – which no longer appears in Kong and Mizera (2012) – will not be examined any further.

12.2.1.3 Directional quantile hyperplanes and contours

Instead of quantile biplots associated with the point-valued quantiles $q_{\tau\mathbf{u}}\mathbf{u}$, Kong and Mizera (2008, 2012) also suggest considering, for each direction $\mathbf{u}$ in $\mathcal{S}^{d-1}$ and each $\tau \in (0, 1/2)$, the *directional quantile hyperplane* $H_{\tau\mathbf{u}}$, with equation $\mathbf{u}'\mathbf{y} = q_{\tau\mathbf{u}}$.

Intuitively, that hyperplane is obtained by looking at the collection of all hyperplanes orthogonal to $\mathbf{u}$; the quantile hyperplane $H_{\tau\mathbf{u}}$ of order τ is the (uniquely defined, for an absolutely continuous distribution with nonvanishing density) hyperplane in that collection dividing $\mathbb{R}^d$ into halfspaces with $\mathrm{P}^{\mathbf{Y}}$-probabilities τ ("below" $H_{\tau\mathbf{u}}$) and $1-\tau$ ("above" $H_{\tau\mathbf{u}}$), respectively.

Denote by $\mathcal{H}_{\tau\mathbf{u}}$ the halfspace lying above $H_{\tau\mathbf{u}}$. The intersection $\mathcal{H}(\tau) := \bigcap_{\mathbf{u}\in\mathcal{S}^{d-1}} \mathcal{H}_{\tau\mathbf{u}}$ of those halfspaces characterizes (for given $\tau \in (0, 1/2)$) an inner envelope. We propose the convenient terminology "quantile region" and "quantile contour" for those inner envelopes and their boundaries $H(\tau)$, which enjoy much nicer properties than the quantile biplots: quantile hyperplanes do not depend on any origin; quantile regions and contours are unique (population case) under Lebesgue-absolutely continuous distributions with connected support; they are convex and nested as τ increases, and affine-equivariant.

The index τ associated with a contour $H(\tau)$ or a region $\mathcal{H}(\tau)$ represents a "tangent probability mass;" indexation by "probability content" might be preferable, and, in view of nestedness, is quite possible – but there is no canonical relation between τ and the $\mathrm{P}^{\mathbf{Y}}$-probability of the quantile region $\mathcal{H}(\tau)$.

Empirical versions $H^{(n)}_{\tau\mathbf{u}}$, $\mathcal{H}^{(n)}(\tau)$, and $H^{(n)}(\tau)$, as usual, are obtained by replacing the distribution $\mathrm{P}^{\mathbf{Y}}$ of $\mathbf{Y}$ with the empirical measure associated with some observed n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. However, the characterization of a given quantile contour $H^{(n)}(\tau)$ involves an infinite number of directions $\mathbf{u}$, which of course is not implementable. In order to overcome this, one can compute the N hyperplanes $H^{(n)}_{\tau\mathbf{u}_i}$ associated with a sample (random or systematic) of directions $\mathbf{u}_i$, $i = 1, \ldots, N$: for absolutely continuous $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, the resulting region $\mathcal{H}^{(n)}_N(\tau) := \bigcap_{i=1,\ldots,N} \mathcal{H}^{(n)}_{\tau\mathbf{u}_i}$ is an approximation of the actual quantile region $\mathcal{H}^{(n)}(\tau)$ (to which, under mild conditions, it converges as $N \to \infty$) – a "biased" one, though, since, with probability 1, $\mathcal{H}^{(n)}_N(\tau)$ strictly includes $\mathcal{H}^{(n)}(\tau)$ for all N.

12.2.1.4 Relation to halfspace depth

Kong and Mizera (2012) then establish a most interesting result that the quantile contours/regions thus defined (as envelopes) and the halfspace depth contours/regions, coincide (in the empirical case as well as in the population case).

Recall that the halfspace depth of a point $\mathbf{y}$ with respect to a probability distribution $\mathrm{P}^{\mathbf{Y}}$ (Tukey, 1975) is the minimum, over all hyperplanes running through $\mathbf{y}$, of the $\mathrm{P}^{\mathbf{Y}}$-probabilities of the halfspaces determined by those hyperplanes. The halfspace depth regions $\mathcal{D}(\delta)$ (halfspace depth contours $D(\delta)$) are the collections of points with halfspace depth larger than or equal to δ (with given halfspace depth δ); those regions are convex and nested as depth decreases. The empirical depth of $\mathbf{y}$ with respect to the n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ is defined similarly, with the empirical distribution of the $\mathbf{Y}_i$ playing the role of $\mathrm{P}^{\mathbf{Y}}$. The empirical halfspace depth contours $D^{(n)}(\delta)$ are polyhedrons, the facet hyperplanes of which typically run through d sample points.

An important byproduct of that result is the hint that only a finite number of directions do characterize a given empirical contour $H^{(n)}(\delta)$, namely, those directions that are orthogonal to the facets of $H^{(n)}(\delta)$. The definition adopted so far, which is related to the traditional univariate definition (12.1), does not readily provide a way to identify those directions, though. The directional Koenker–Bassett approach of Section 12.2.2, which extends the univariate L_1 definition (12.2), also leads to a numerical determination of the relevant directions.

In the presence of covariates ($p \geqslant 1$), the connection with halfspace depth does not help much, as most existing regression depth concepts, inspired by Rousseeuw and Hubert (1999), are limited to the single-output setting; a few exceptions can be found, however, in Mizera (2002) and Šiman (2011).

12.2.2 Directional Koenker–Bassett approach

12.2.2.1 Location case ($p = 0$)

Another directional approach is proposed in Hallin et al. (2010), based on a directional adaptation of the L_1 definition (12.2) rather than on Kong and Mizera's directional version of the traditional definition (12.1).

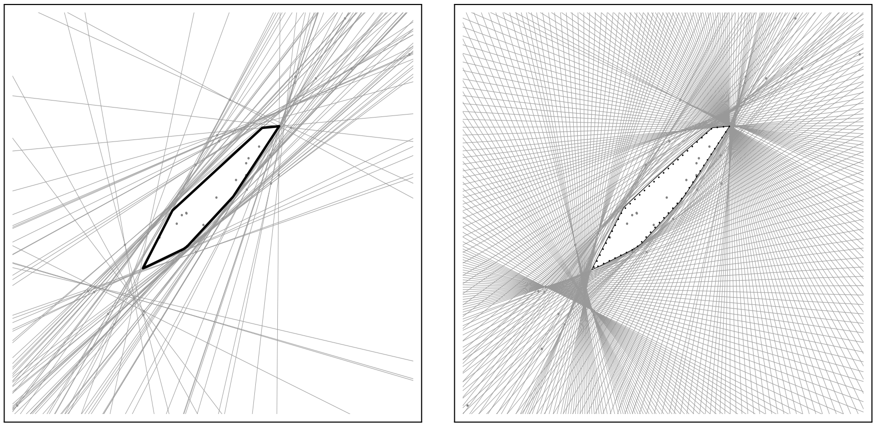
Instead of projecting $\mathbf{Y}$ on a direction $\mathbf{u} \in \mathcal{S}^{d-1}$, Hallin et al. (2010) propose minimizing the usual L_1 residual distance along a direction $\mathbf{u}$ ranging over $\mathcal{S}^{d-1}$: the usual Koenker–Bassett quantile hyperplane construction (Koenker and Bassett, 1978), with "vertical direction" $\mathbf{u}$. More precisely, denoting by $\boldsymbol{\Gamma}_{\mathbf{u}}$ an arbitrary $d \times (d-1)$ matrix of unit vectors such that $(\mathbf{u}, \boldsymbol{\Gamma}_{\mathbf{u}})$ constitutes an orthonormal basis of $\mathbb{R}^d$, decompose $\mathbf{Y}$ into $\mathbf{Y}_{\mathbf{u}} + \mathbf{Y}_{\mathbf{u}}^{\perp}$, where $\mathbf{Y}_{\mathbf{u}} := \mathbf{u}'\mathbf{Y}$ and $\mathbf{Y}_{\mathbf{u}}^{\perp} := \boldsymbol{\Gamma}_{\mathbf{u}}'\mathbf{Y}$. Hallin et al. (2010) define the directional τ-quantile hyperplane of $\mathbf{Y}$ (equivalently, of $\mathrm{P}^{\mathbf{Y}}$) in the direction $\mathbf{u}$ as the hyperplane $\Pi_{\tau\mathbf{u}}$ with equation $\mathbf{u}'\mathbf{y} = \mathbf{b}_{\tau\mathbf{u}}'\boldsymbol{\Gamma}_{\mathbf{u}}'\mathbf{y} + a_{\tau\mathbf{u}}$, where ($\rho_\tau$ as usual stands for the τ-quantile *check function*)

$$(a_{\tau\mathbf{u}}, \mathbf{b}_{\tau\mathbf{u}}')' = \mathrm{argmin}_{(a,\mathbf{b}')' \in \mathbb{R}^d} \mathrm{E}[\rho_\tau(\mathbf{Y}_{\mathbf{u}} - \mathbf{b}'\mathbf{Y}_{\mathbf{u}}^{\perp} - a)]. \tag{12.5}$$

The empirical version $\Pi_{\tau\mathbf{u}}^{(n)}$ of $\Pi_{\tau\mathbf{u}}$, with equation $\mathbf{u}'\mathbf{y} = \mathbf{b}_{\tau\mathbf{u}}^{(n)\prime}\boldsymbol{\Gamma}_{\mathbf{u}}'\mathbf{y} + a_{\tau\mathbf{u}}^{(n)}$, is obtained by replacing the distribution $\mathrm{P}^{\mathbf{Y}}$ of $\mathbf{Y}$ with the empirical measure associated with an observed n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$:

$$(a_{\tau\mathbf{u}}^{(n)}, \mathbf{b}_{\tau\mathbf{u}}^{(n)\prime})' = \mathrm{argmin}_{(a,\mathbf{b}')' \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\tau(\mathbf{Y}_{i,\mathbf{u}} - \mathbf{b}'\mathbf{Y}_{i,\mathbf{u}}^{\perp} - a). \tag{12.6}$$

For any fixed τ, the hyperplanes $\{\Pi_{\tau\mathbf{u}} : \mathbf{u} \in \mathcal{S}^{q-1}\}$ determine a quantile contour $R(\tau)$



(a) (b)

FIGURE 12.1
(a) All 0.2-quantile hyperplanes (thin gray lines) obtained, in a sample of $n = 49$ observations, from the directional Koenker–Bassett definition and the resulting directional 0.2-quantile contours (thick black lines). (b) Approximation (thin black lines) of the same directional 0.2-quantile contours (thick black dotted lines), based on Kong and Mizera's directional quantile hyperplanes (thin gray lines) corresponding to $N = 256$ equispaced $\mathbf{u}$ values on $\mathcal{S}^1$.

and a quantile region $\mathcal{R}(\tau)$ (for the empirical hyperplanes $\{\Pi^{(n)}_{\tau\mathbf{u}} : \mathbf{u} \in \mathcal{S}^{d-1}\}$, $R^{(n)}(\tau)$ and $\mathcal{R}^{(n)}(\tau)$). The hyperplanes constituting an empirical contour can be obtained as the solutions of a linear program parametrized by $\mathbf{u}$; see Hallin et al. (2010) for details, Paindaveine and Šiman (2012a,b) for further computational insights, and Boček and Šiman (2016) for software implementation. The linear programming structure of the problem implies that only a few critical values of $\mathbf{u}$ play a role. More precisely, the unit ball in $\mathbb{R}^d$ is partitioned into a finite number of cones with vertex at the origin, with all the $\mathbf{u}$ in a given cone determining the same quantile hyperplane $\Pi^{(n)}_{\tau\mathbf{u}}$. Those cones are obtained via standard parametric linear programming algorithms.

Hallin et al. (2010) show, moreover, that those quantile contours also coincide with the halfspace depth contours, hence with Kong and Mizera's directional quantile contours. The huge advantage with respect to the projection approach of Section 12.2.1.3 is that, thanks to the "analytical" nature of the L_1 definition, a given empirical contour here can be computed exactly in a finite number of steps. That advantage may disappear, however, as the size of the problem increases: when n and d become too large, linear programming algorithms eventually run into problems, and the approximate contours of Section 12.2.1.3 may be the only feasible solution. The two points of view, however, can be reconciled (see Paindaveine and Šiman, 2011).

Figure 12.1 shows (a) the empirical quantile contour of order $\tau = 0.2$, obtained from a data set of $n = 49$ observations and the directional Koenker–Bassett definition just described, and (b) the approximation of the same contour, based on the Kong and Mizera's directional quantile hyperplanes associated with $N = 256$ equispaced $\mathbf{u}$ values on $\mathcal{S}^1$.

The multivariate quantile contours resulting from this directional Koenker–Bassett approach inherit from their relation to halfspace depth the rich geometric features – convexity, connectedness, nestedness, affine equivariance – of the latter, while bringing to halfspace depth the nice analytical, computational, and probabilistic features of L_1 optimization –

tractable asymptotics (Bahadur representation, root-n consistency, asymptotic normality, etc.), L_1 characterization/optimality, implementable linear programming algorithms, and byproducts of optimization problems (duality and Lagrange multipliers); see Hallin et al. (2010) for explicit results and details.

12.2.2.2 (Nonparametric) regression case ($p \geqslant 1$)

A linear regression extension ($p \geqslant 1$), based on the L_1 approach, of the location concept described in the previous section is quite straightforward. Denote by $(\mathbf{X}_1', \mathbf{Y}_1')', \ldots, (\mathbf{X}_n', \mathbf{Y}_n')'$ an observed n-tuple of independent copies of $(\mathbf{X}', \mathbf{Y}')'$, where $\mathbf{Y} := (Y_1, \ldots, Y_d)'$ is a d-dimensional response and the random vector $\mathbf{X} := (X_1, \ldots, X_p)'$ is a p-tuple of covariates. Definitions (12.5) and (12.6) readily generalize into

$$(a_{\tau\mathbf{u}}, \mathbf{b}_{\tau\mathbf{u}}', \boldsymbol{\beta}_{\tau\mathbf{u}}')' = \text{argmin}_{(a,\mathbf{b}',\boldsymbol{\beta}')' \in \mathbb{R}^{d+p}} \mathrm{E}[\rho_\tau(\mathbf{Y}_\mathbf{u} - \mathbf{b}'\mathbf{Y}_\mathbf{u}^\perp - \boldsymbol{\beta}'\mathbf{X} - a)] \tag{12.7}$$

and

$$(a_{\tau\mathbf{u}}^{(n)}, \mathbf{b}_{\tau\mathbf{u}}^{(n)\prime}, \boldsymbol{\beta}_{\tau\mathbf{u}}^{(n)\prime})' = \text{argmin}_{(a,\mathbf{b}',\boldsymbol{\beta}')' \in \mathbb{R}^{d+p}} \sum_{i=1}^{n} \rho_\tau(\mathbf{Y}_{i,\mathbf{u}} - \mathbf{b}'\mathbf{Y}_{i,\mathbf{u}}^\perp - \boldsymbol{\beta}'\mathbf{X}_i - a), \tag{12.8}$$

characterizing hyperplanes (now in $\mathbb{R}^{d+p}$) $\Pi_{\tau\mathbf{u}}$, with equation

$$\mathbf{u}'\mathbf{y} = \mathbf{b}_{\tau\mathbf{u}}'\boldsymbol{\Gamma}_\mathbf{u}'\mathbf{y} + \boldsymbol{\beta}_{\tau\mathbf{u}}'\mathbf{x} + a_{\tau\mathbf{u}},$$

and empirical hyperplanes $\Pi_{\tau\mathbf{u}}^{(n)}$, with equation

$$\mathbf{u}'\mathbf{y} = \mathbf{b}_{\tau\mathbf{u}}^{(n)\prime}\boldsymbol{\Gamma}_\mathbf{u}'\mathbf{y} + \boldsymbol{\beta}_{\tau\mathbf{u}}^{(n)\prime}\mathbf{x} + a_{\tau\mathbf{u}}^{(n)},$$

respectively.

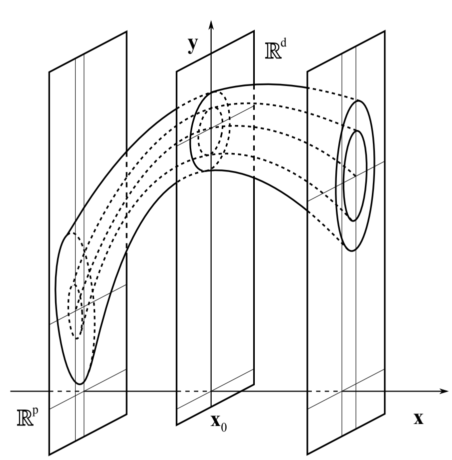


FIGURE 12.2
Directional Koenker–Bassett quantile regression: population quantile regression tubes ($d = 2$, $p = 1$).

To the best of our knowledge, the asymptotic properties of such extensions have not been worked out in the literature, although asymptotic results of the same type as those obtained for the location case certainly can be established under appropriate assumptions. The major problem is the interpretation of the contours characterized by the collection of hyperplanes $\Pi_{\tau\mathbf{u}}$ as $\mathbf{u}$ ranges over the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$. The relevant quantile hyperplanes, quantile/depth regions and contours of interest are the location quantile hyperplanes, quantile/depth regions and contours associated with the d-dimensional distributions of $\mathbf{Y}$ conditional on $\mathbf{X}$ – namely, the collection, for $\mathbf{x}$ ranging over $\mathbb{R}^p$, of the hyperplanes, regions and contours associated with the distributions $\mathrm{P}^{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$. When plotted against $\mathbf{x}$ (which is possible for $d + p \leqslant 3$ only), those contours yield quantile regression "tubes" (see Figure 12.2). Unless very severe restrictions[1] are put on the data-generating process, the contours resulting from (12.7) and (12.8) are not the depth contours associated with the conditional distributions $\mathrm{P}^{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$, but some averaged version of the latter; their interpretation is thus somewhat problematic. This is the reason why Hallin et al. (2015) consider a fully general nonparametric regression setup (rather than linear regression and definitions (12.7) and (12.8)) with the objective of reconstructing the collection of *conditional* (on the value $\mathbf{X} = \mathbf{x}$) quantile contours of $\mathbf{Y}$.

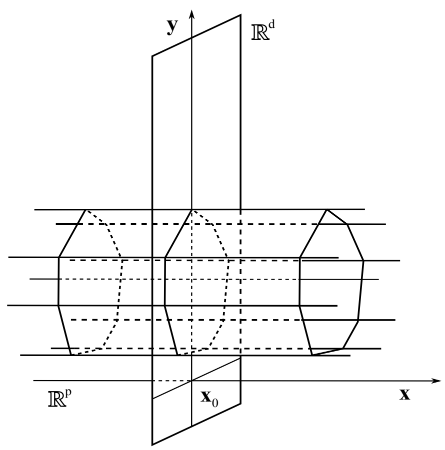


FIGURE 12.3
Directional Koenker–Bassett quantile regression: local constant empirical quantile regression tube at $\mathbf{x}_0$.

Two consistent estimation methods are provided in Hallin et al. (2015): a local constant method and a local bilinear one. In both cases, the estimators are based on a weighted (kernel-based) version of the location contour estimators, computationally still leading to a parametrized linear programming problem with directional parameter $\mathbf{u}$ ranging over the unit sphere $\mathcal{S}^{d-1}$. Bahadur representations of the resulting estimators are established under appropriate technical conditions on the joint distributions of $(\mathbf{X}', \mathbf{Y}')'$, the kernel and the bandwidth defining the weights; those representations entail consistency and asymptotic normality.

Local constant quantile contours yield, for given τ and a selected value $\mathbf{x}_0$ of $\mathbf{x}$, a "horizontal polygonal tube" (Figure 12.3) in $\mathbb{R}^{d+p}$, the interpretation of which is valid at $\mathbf{x}_0$ only, and provides no information on the way the conditional distribution of $\mathbf{Y}$ is varying in the neighborhood of $\mathbf{x}_0$.

[1]For instance, requiring that the distribution of $\mathbf{Y} - \mathbf{B}\mathbf{x}$ (for $\mathbf{B}$ some $d \times p$ regression matrix), conditional on $\mathbf{X} = \mathbf{x}$, does not depend on $\mathbf{x}$.

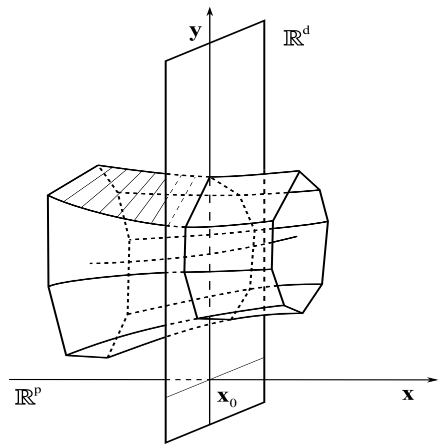


FIGURE 12.4
Directional Koenker–Bassett quantile regression: local bilinear empirical quantile regression tube at $\mathbf{x}_0$.

Local bilinear quantile contours are more informative, since they incorporate information on the derivatives with respect to $\mathbf{x}$ of the coefficients of the conditional quantile hyperplanes; they should also be more reliable at boundary points. The price to be paid is an increase of the number of free parameters involved. Note, however, that the smoothing features of the problem, namely the dimension p of kernels, remains unaffected, irrespective of d). The resulting empirical tubes, as shown in Figure 12.4, are no longer polygonal cylinders, but piecewise ruled quadrics.

Figure 12.5 shows the empirical contours ($\tau = 0.2$ and 0.4) constructed, via (a) the local constant method and (b) the local bilinear method, for various values of x_0, for a set of $n = 4899$ observations ($d = 2$, $p = 1$) simulated from the bivariate heteroskedastic regression model

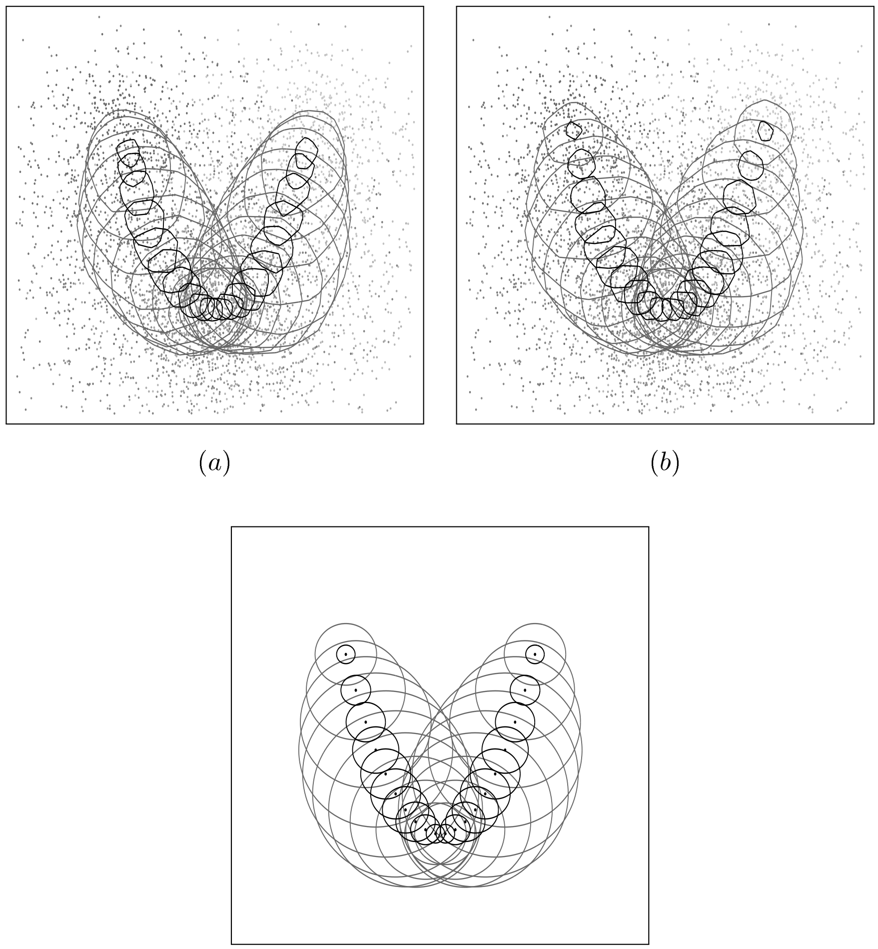
$$(Y_1, Y_2)' = (X, X^2)' + 0.5\big(1 + 3|\sin(\pi X/2)|\big)(e_1, e_2)', \tag{12.9}$$

where X is uniform over $(-2, 2)$ and $(e_1, e_2)' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is bivariate normal. The axes are those of the response space, and (unlike in Figure 12.4) the contours associated with various values of X are shown side by side. The "parabolic regression median" and the periodic conditional scale are well picked up by both methods. The local bilinear contours are less sensitive, as expected, to boundary effects.

We refer the reader to Hallin et al. (2015) for details.

12.3 Direct approaches

All multiple-output quantile regression concepts presented so far were based on directional extensions of the usual single-output ones. Direct approaches are possible, though, along three main lines. The first is based on a spatial extension of the definition of the check function, leading to "spatial" quantiles (also called "geometric" quantiles). The second



(a) (b)

(c)

FIGURE 12.5
Directional Koenker–Bassett quantile regression: the empirical contours ($\tau = 0.2$ and 0.4) obtained, for various values of x_0, via (a) the local constant method and (b) the local bilinear method, for $n = 4899$ observations from model (12.9), along with (c) their (exact) population counterparts; the dot at the center of the population contours is both the conditional mean and the conditionally deepest point.

involves substituting, in the traditional Koenker and Bassett (1978) definition, ellipsoids for hyperplanes, and the "above/below" indicators with "outside/inside" ones. The third approach, inspired by the relation between directional quantiles and halfspace depth (Section 12.2.1.4), is based on recent measure transportation-related concepts (Carlier et al., 2016; Chernozhukov et al., 2017) of Monge–Kantorovich depth and quantiles.

12.3.1 Spatial (geometric) quantile methods

The concepts of *spatial* or *geometric* quantiles, generalizing the well-known spatial median (Haldane, 1948; Gower, 1974; Brown, 1983), are based on a multivariate extension of the L_1 approach outlined in Section 12.1; see Chaudhuri (1996), Koltchinskii (1997), and the references therein.

12.3.1.1 A spatial check function

The concepts of spatial median and quantiles are based on an alternative form for the check function that "naturally" extends to a multivariate context, possibly combined with transformation–retransformation ideas.

It is easy to see that the check function ρ_τ, in univariate or single-output regression settings, can be rewritten as

$$\begin{aligned}\rho_\tau(z) &:= (1-\tau)|z|I_{[z<0]} + \tau|z|I_{[z\geqslant 0]}\\ &= \frac{1}{2}(|z| + vz) =: \frac{1}{2}\Phi_v(z)\end{aligned}$$

with $v := 2\tau - 1$. While τ ranges over $(0,1)$, v ranges over the open unit ball $(-1,1)$ in $\mathbb{R}$. Substituting Φ_v for ρ_τ in (12.2), (12.3), and (12.4) thus leads to the same concepts as in Section 12.1, with, however, a different index v. That index v has a centre-outward directional flavour, with $|v|$ measuring centrality and $v/|v|$ characterizing a direction (either $+1$ or -1).

It is tempting, therefore, to extend the univariate and single-output regression concepts (12.2)–(12.4) to the multivariate and multiple-output context by minimizing a criterion based on the "spatial check function"

$$\Phi_{\mathbf{v}}(\mathbf{z}) := \|\mathbf{z}\| + \mathbf{v}'\mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^d, \tag{12.10}$$

where $\mathbf{v}$ ranges over the open unit ball in $\mathbb{R}^d$, hence takes the form $\mathbf{v} = \tau\mathbf{u}$, where $\tau = \|\mathbf{v}\|$ and $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\| \in \mathcal{S}^{d-1}$.

This, in the location case ($p = 0$, no covariates), yields the population spatial quantile of order $\mathbf{v}$ and its empirical counterpart,

$$\mathbf{Q}_{\mathbf{v}} := \mathrm{argmin}_{\mathbf{q}\in\mathbb{R}^d}\mathrm{E}[\Phi_{\mathbf{v}}(\mathbf{Y}-\mathbf{q})], \quad \mathbf{Q}_{\mathbf{v}}^{(n)} := \mathrm{argmin}_{\mathbf{q}\in\mathbb{R}^d}\sum_{i=1}^n[\Phi_{\mathbf{v}}(\mathbf{Y}_i-\mathbf{q})], \tag{12.11}$$

respectively. For $\tau = 0$ (hence $\mathbf{v} = \mathbf{0}$), one obtains the traditional *spatial median.* Although an intuitive justification of (12.10) is not straightforward, the solutions of (12.11) are such that

$$\mathrm{E}\Big[(\mathbf{Y}-\mathbf{Q}_{\mathbf{v}})/\|\mathbf{Y}-\mathbf{Q}_{\mathbf{v}}\|\Big] = -\mathbf{v} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i-\mathbf{Q}_{\mathbf{v}}^{(n)})/\|(\mathbf{Y}_i-\mathbf{Q}_{\mathbf{v}}^{(n)})\| = -\mathbf{v},$$

which provides an interpretation in terms of the unit vectors originating in $\mathbf{Q}_{\mathbf{v}}$ or $\mathbf{Q}_{\mathbf{v}}^{(n)}$ and pointing at $\mathbf{Y}$ or the observations $\mathbf{Y}_i$. An interesting discussion of this can be found in Serfling (2002b).

Finally, spatial quantiles are also quite robust. They characterize the underlying distribution, and their definition nicely extends (Chakraborty and Chaudhuri, 2014) to Banach spaces.

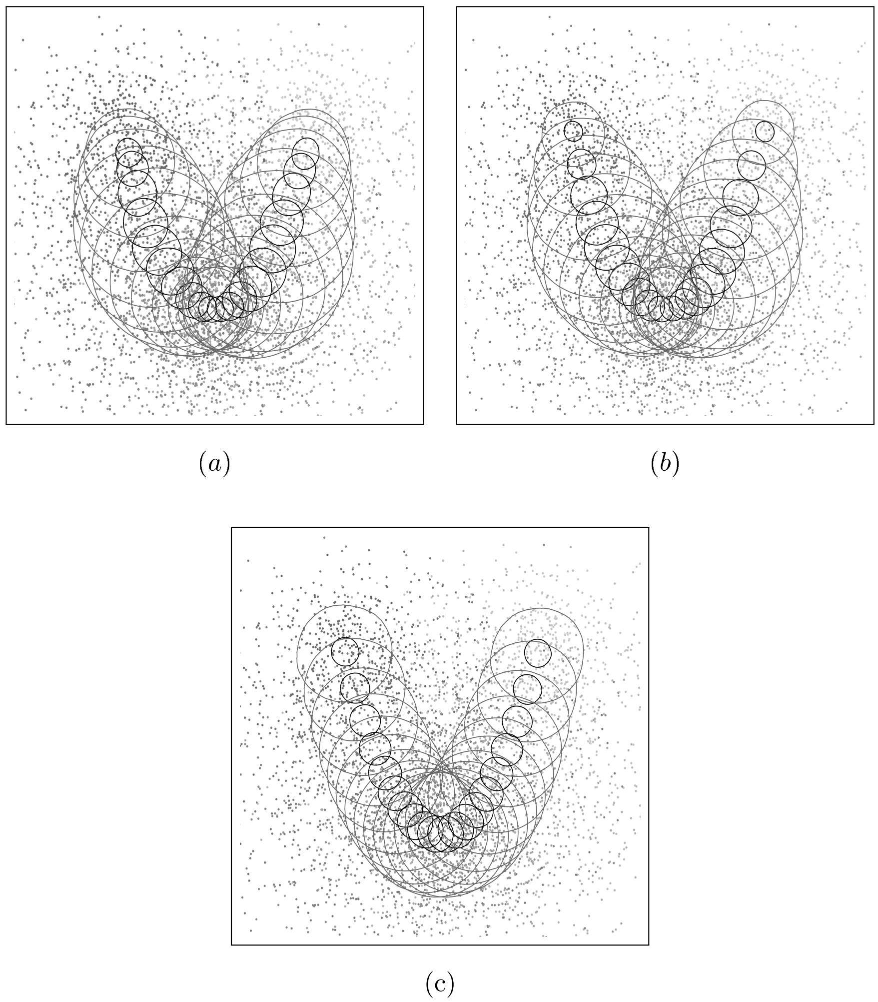


FIGURE 12.6
Spatial quantile regression: empirical contours obtained, for various values of τ and x_0, via (a) the (nonparametric) local constant method, (b) the (nonparametric) local linear method, and (c) the (parametric) linear regression model with regressors X and X^2, for the same $n = 4899$ observations as in Figure 12.5.

12.3.1.2 Linear spatial quantile regression

In the (multiple-output) linear regression case, the observations $(\mathbf{Y}_i, \mathbf{X}_i)$ satisfy an equation of the form

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{B}\mathbf{X}_i + \mathbf{E}_i, \quad i = 1, \ldots, n, \tag{12.12}$$

where $\mathbf{a}$ and $\mathbf{B}$ are $d \times 1$ and $d \times p$, respectively. The definitions in (12.11) straightforwardly generalize into

$$(\mathbf{a}_{\tau\mathbf{u}}, \mathbf{B}_{\tau\mathbf{u}}) := \operatorname{argmin}_{(\mathbf{a},\mathbf{B}) \in \mathbb{R}^d \times \mathbb{R}^{d\times p}} \mathrm{E}[\Phi_{\tau\mathbf{u}}(\mathbf{Y} - \mathbf{a} - \mathbf{B}\mathbf{X})] \tag{12.13}$$

and

$$(\mathbf{a}^{(n)}_{\tau\mathbf{u}}, \mathbf{B}^{(n)}_{\tau\mathbf{u}}) := \operatorname{argmin}_{(\mathbf{a},\mathbf{B}) \in \mathbb{R}^d \times \mathbb{R}^{d\times p}} \sum_{i=1}^{n} \Phi_{\tau\mathbf{u}}(\mathbf{Y}_i - \mathbf{a} - \mathbf{B}\mathbf{X}_i), \tag{12.14}$$

with $\mathbf{v} = \tau\mathbf{u}$ ranging over the open unit ball.

Explicit values of individual (fixed τ and $\mathbf{u}$) empirical spatial regression quantiles are easily obtained via convex optimization techniques, but the corresponding contours (viz., the envelopes of the $\tau\mathbf{u}$-quantile regression hyperplanes associated with fixed τ as $\mathbf{u}$ ranges over the unit sphere $\mathcal{S}^{d-1}$) can only be approximated by considering a large number of $\mathbf{u}$ values. A numerical illustration of the method, based on the same simulated data set as in Figure 12.5, is provided in Figure 12.6(c).

The spatial regression quantiles partially order the regression hyperplanes. On the other hand, they have no obvious probabilistic interpretation, and perform quite poorly for distributions with nonconvex level sets (see Figure 12.8(b)). They also fail to be fully affine-equivariant – a feature that can be taken care of via the transformation–retransformation technique advocated by Chakraborty and Chaudhuri (1996); see Chaudhuri (1996), Koltchinskii (1997), and Chakraborty (2003) for details, and Bai et al. (1990) for the special case of median regression.

12.3.1.3 Nonparametric spatial quantile regression

Taking expectations in (12.11) and (12.13) with respect to a Nadaraya–Watson estimate of the conditional distribution function of $\mathrm{P}^{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$, one can obtain, similarly as in Section 12.2.2.2, locally constant and locally linear (nonparametric) estimators of the conditional spatial quantiles. Such methods and their properties are investigated in Cheng and de Gooijer (2007) and Chaouch et al. (2009), to which where we refer the reader for details. An illustration is provided in Figure 12.6(a) and (b), for the same data as in Figure 12.5; the results look quite similar to those obtained from nonparametric directional quantile regression methods.

12.3.2 Elliptical quantiles

A concept of elliptical regression quantiles was proposed, very much in the same spirit as Koenker and Bassett's original definition, by Hlubinka and Šiman (2013, 2015) and Hallin and Šiman (2016). As in Section 12.2, we start with the pure location case ($p = 0$, no covariates) before turning to the general linear regression case.

12.3.2.1 Location case

The basic idea behind the concept is intuitively quite simple and appealing: instead of the inner envelope of a collection of directional Koenker–Bassett τ-quantile hyperplanes minimizing the expected value of a directional check function penalty (Section 12.2.2.1), consider an ellipsoid minimizing the same weighted L_1 objective function where, however, "above/below" the hyperplane is replaced with "outside/inside" the ellipsoid. This leads to the definition of a multivariate (location) elliptical τ-quantile as the ellipsoid

$$\mathcal{E}^{\mathrm{loc}}_{\tau} = \mathcal{E}^{\mathrm{loc}}_{\tau}(\mathbf{Y}) := \{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}'\mathbf{A}_{\tau}\mathbf{y} + \mathbf{y}'\mathbf{b}_{\tau} - c_{\tau} = 0\},$$

where $\mathbf{A}_\tau \in \mathbb{R}^{d\times d}$, $\mathbf{b}_\tau \in \mathbb{R}^{d\times 1}$, and $c_\tau \geqslant 0$ minimize, subject to $\mathbf{A}$ being symmetric and positive semidefinite with determinant 1 – a *shape matrix* in the sense of Paindaveine (2008) – the objective function

$$\Psi_\tau^{\mathrm{loc}}(\mathbf{A}, \mathbf{b}, c) := \mathrm{E}\rho_\tau(\mathbf{Y}'\mathbf{A}\mathbf{Y} + \mathbf{Y}'\mathbf{b} - c), \tag{12.15}$$

where ρ_τ, as usual, stands for the check function

$$z \mapsto \rho_\tau(z) := z(\tau - I(z < 0)) = \max\{(\tau - 1)z, \tau z\}.$$

Note that the argument of ρ_τ in (12.15) is positive or negative according to whether $\mathbf{Y}$ takes value inside or outside the ellipsoid with equation $\mathbf{y}'\mathbf{A}\mathbf{y} + \mathbf{y}'\mathbf{b} = c$.

The positive semidefiniteness of $\mathbf{A}_\tau$ and the condition on its determinant ensure that $\mathcal{E}_\tau^{\mathrm{loc}}$ is indeed an ellipsoid, centered at $\mathbf{s}_\tau := -\mathbf{A}_\tau^{-1}\mathbf{b}_\tau/2$, with equation $(\mathbf{y}-\mathbf{s}_\tau)'\mathbf{A}_\tau(\mathbf{y}-\mathbf{s}_\tau) = \kappa_\tau$, where $\kappa_\tau := c_\tau + \mathbf{b}_\tau'\mathbf{A}_\tau^{-1}\mathbf{b}_\tau/4$. The condition $\det(\mathbf{A}) = 1$ can be viewed as an identification constraint: for any $K > 0$, the triples $(\mathbf{A}, \mathbf{b}, c)$ and $(K\mathbf{A}, K\mathbf{b}, Kc)$ indeed define the same ellipsoid.

This definition certainly does not characterize an elliptical quantile as the solution of a linear programming problem; nor, as it stands, does it take the form of a convex optimization problem. The same concept, however, can be characterized as the unique solution of a convex optimization problem by relaxing the constraint $\det(\mathbf{A}) = 1$ into $(\det(\mathbf{A}))^{1/d} \geqslant 1$: unlike $\mathbf{A} \mapsto \det(\mathbf{A})$, the function $\mathbf{A} \mapsto (\det(\mathbf{A}))^{1/d}$ is concave on the cone of symmetric positive semidefinite matrices, and it can be shown that this convex optimization problem and the original nonconvex one share the same solution. That solution, moreover, is unique under absolutely continuous distributions with nonvanishing densities and finite moments of order 2.

12.3.2.2 Linear regression case

In the presence of covariates ($p \geqslant 1$), the traditional homoskedastic multiple-output linear regression model suggests, for an elliptical multiple-output regression quantile of order τ, a simple equation of the form

$$(\mathbf{y} - \boldsymbol{\beta}_\tau - \mathbf{B}_\tau\mathbf{x})'\mathbf{A}_\tau(\mathbf{y} - \boldsymbol{\beta}_\tau - \mathbf{B}_\tau\mathbf{x}) - \gamma_\tau = 0$$

with some $\mathbf{A}_\tau \in \mathbb{R}^{d\times d}$, $\boldsymbol{\beta}_\tau \in \mathbb{R}^{d\times 1}$, $\mathbf{B}_\tau \in \mathbb{R}^{d\times p}$, and $\gamma_\tau \geqslant 0$. The trouble is that the corresponding objective function

$$\mathrm{E}\rho_\tau\big((\mathbf{Y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{X})'\mathbf{A}(\mathbf{Y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{X}) - \gamma\big)$$

is not convex in $\boldsymbol{\beta}$ and $\mathbf{B}$, so that its minimization with respect to $\mathbf{A}$, $\boldsymbol{\beta}$, $\mathbf{B}$, and γ is not a *convex* optimization problem (see Hlubinka and Šiman, 2015).

In order to restore convexity, Hallin and Šiman (2016) consider instead the more general definition

$$\begin{aligned}\mathcal{E}_\tau^{\mathrm{reg}} := \big\{&(\mathbf{y}', \mathbf{x}')' \in \mathbb{R}^{d+p}: \\ &(\mathbf{y} - \boldsymbol{\beta}_\tau - \mathbf{B}_\tau\mathbf{x})'\mathbf{A}_\tau(\mathbf{y} - \boldsymbol{\beta}_\tau - \mathbf{B}_\tau\mathbf{x}) - (\gamma_\tau + \mathbf{c}_\tau'\mathbf{x} + \mathbf{x}'\mathbf{C}_\tau\mathbf{x}) = 0\big\}\end{aligned} \tag{12.16}$$

of an elliptical regression quantile $\mathcal{E}_\tau^{\mathrm{reg}} = \mathcal{E}_\tau^{\mathrm{reg}}(\mathbf{Y}, \mathbf{X})$, where a quadratic form of covariate-driven scale is allowed, and $\mathbf{A}_\tau, \boldsymbol{\beta}_\tau, \mathbf{B}_\tau, \gamma_\tau, \mathbf{c}_\tau$, and $\mathbf{C}_\tau$ jointly minimize

$$\Psi_\tau^{\mathrm{reg}} := \mathrm{E}\rho_\tau\big((\mathbf{Y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{X})'\mathbf{A}(\mathbf{Y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{X}) - (\gamma + \mathbf{c}'\mathbf{X} + \mathbf{X}'\mathbf{C}\mathbf{X})\big) \tag{12.17}$$

under the constraint that $\mathbf{C} \in \mathbb{R}^{p\times p}$ is symmetric and $\mathbf{A} \in \mathbb{R}^{d\times d}$ is symmetric positive semidefinite with $\det(\mathbf{A}) = 1$.

This minimization, again, does not take the form of a convex optimization problem. Let, therefore, $\mathbf{M} := (\mathbf{M}^1, \ \ldots, \ \mathbf{M}^6)$, with

$$\begin{aligned}
\mathbf{M}^1 &:= \mathbf{A} \in \mathbb{R}^{d\times d} \text{ symmetric positive semidefinite,} \\
\mathbf{M}^2 &:= \mathbf{B}'\mathbf{A}\mathbf{B} - \mathbf{C} \in \mathbb{R}^{p\times p} \text{ symmetric,} \\
\mathbf{M}^3 &:= -2\mathbf{B}'\mathbf{A} \in \mathbb{R}^{p\times d}, \\
\mathbf{M}^4 &:= -2\boldsymbol{\beta}'\mathbf{A} \in \mathbb{R}^{1\times d}, \\
\mathbf{M}^5 &:= 2\boldsymbol{\beta}'\mathbf{A}\mathbf{B} - \mathbf{c}' \in \mathbb{R}^{1\times p}, \\
\mathbf{M}^6 &:= \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta} - \gamma \in \mathbb{R}.
\end{aligned}$$

The correspondence between $\mathbf{M}$ and $(\mathbf{A}, \ \boldsymbol{\beta}, \ \mathbf{B}, \ \gamma, \ \mathbf{c}, \ \mathbf{C})$ is one-to-one, and $\mathbf{M}$ thus provides a reparametrization of the problem.

In this new parametrization, the elliptical regression quantile $\mathcal{E}_\tau^{\text{reg}}$ can be expressed as

$$\mathcal{E}_\tau^{\text{reg}} = \{(\mathbf{y}', \mathbf{x}')' \in \mathbb{R}^{d+p} : r(\mathbf{y}, \mathbf{x}, \mathbf{M}_\tau) = 0\},$$

where

$$\begin{aligned}
r(\mathbf{y}, \mathbf{x}, \mathbf{M}) &:= \mathbf{y}'\mathbf{M}^1\mathbf{y} + \mathbf{x}'\mathbf{M}^2\mathbf{x} + \mathbf{x}'\mathbf{M}^3\mathbf{y} + \mathbf{M}^4\mathbf{y} + \mathbf{M}^5\mathbf{x} + \mathbf{M}^6 \\
&= (\mathbf{y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{x})'\mathbf{A}(\mathbf{y} - \boldsymbol{\beta} - \mathbf{B}\mathbf{x}) - (\gamma + \mathbf{c}'\mathbf{x} + \mathbf{x}'\mathbf{C}\mathbf{x})
\end{aligned}$$

(r is thus positive outside, and negative inside, the ellipsoid with equation $r = 0$), and $\mathbf{M}_\tau := (\mathbf{M}_\tau^1, \ldots, \mathbf{M}_\tau^6)$ jointly minimize

$$\Psi_\tau^{\text{reg}} = \Psi_\tau^{\text{reg}}(\mathbf{M}) := \Psi_\tau^{\text{reg}}(\mathbf{M}^1, \ldots, \mathbf{M}^6) = \mathrm{E}\rho_\tau\big(r(\mathbf{Y}, \mathbf{X}, \mathbf{M})\big),$$

subject to $(\det(\mathbf{M}^1))^{1/d} \geqslant 1$; as in the location case, positive homogeneity of $\Psi_\tau^{\text{reg}}(\mathbf{M}^1, \ldots, \mathbf{M}^6)$ implies $\det(\mathbf{M}_\tau) = 1$. The considerable advantage of this parametrization in terms of $\mathbf{M}$ is that it leads to a *convex* optimization problem, hence to a *unique* minimum under the assumptions made (which include the existence of finite second-order moments).

The (Karush–)Kuhn–Tucker necessary and sufficient conditions characterizing the solution imply, in particular, that the probability content of $\mathcal{E}_\tau^{\text{reg}}$ is τ, and that

$$\mathrm{E}[(\mathbf{Y}', \mathbf{X}')'|r \geqslant 0] = \mathrm{E}[(\mathbf{Y}', \mathbf{X}')'|r < 0],$$

so that the probability mass centers of the interior and the exterior of $\mathcal{E}_\tau^{\text{reg}}$ coincide. It is easy to see, moreover, that the elliptical regression quantiles $\mathcal{E}_\tau^{\text{reg}}$ are both regression-equivariant and fully affine-equivariant.

In the sample case with n observations $(\mathbf{Y}_i', \mathbf{X}_i')'$, $i = 1, \ldots, n$, empirical versions $\mathcal{E}_{\tau;n}^{\text{reg}}$ of the elliptical regression quantiles $\mathcal{E}_\tau^{\text{reg}}$ are based on the empirical counterparts of (12.17). Classical results (such as van der Vaart and Wellner, 1996, Theorem 5.14) then guarantee basic convergence, as $n \to \infty$, of the vector

$$\mathbf{m}_{\tau;n} := \big(\mathrm{vec}(\mathbf{M}_{\tau;n}^1)', \mathrm{vec}(\mathbf{M}_{\tau;n}^2)', \mathrm{vec}(\mathbf{M}_{\tau;n}^3)', \mathbf{M}_{\tau;n}^4, \mathbf{M}_{\tau;n}^5, \mathbf{M}_{\tau;n}^6\big)'$$

of coefficients of the sample elliptical regression quantile to its (uniquely defined) population counterpart

$$\mathbf{m}_\tau := \big(\mathrm{vec}(\mathbf{M}_\tau^1)', \mathrm{vec}(\mathbf{M}_\tau^2)', \mathrm{vec}(\mathbf{M}_\tau^3)', \mathbf{M}_\tau^4, \mathbf{M}_\tau^5, \mathbf{M}_\tau^6\big)'.$$

Figure 12.7 illustrates the ability of elliptical regression quantiles to correctly estimate the trend and heteroskedasticity in linear regression models.

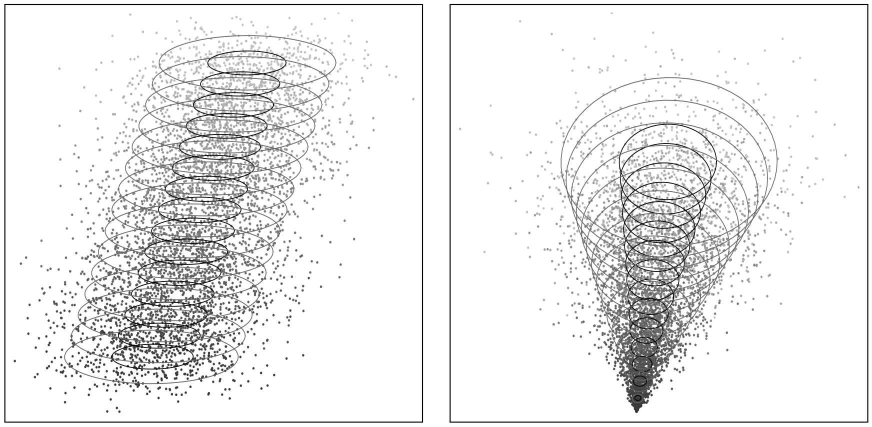


FIGURE 12.7
Elliptical linear quantile regression: the empirical contours estimated, for equidistant regressor values and two quantile levels, from (a) a homoskedastic, and (b) a heteroskedastic regression model with linear trend.

12.3.3 Depth-based quantiles

As mentioned in the introduction, depth contours (preferably indexed by their probability content) naturally provide a plausible concept of quantile contours. Many depth concepts are available in the literature, and we will restrict ourselves to two important cases: halfspace depth, the relation of which to directional quantiles has been outlined in Section 12.2; and the more recent concept of Monge–Kantorovich depth.

12.3.3.1 Halfspace depth quantiles

Halfspace depth contours, as we have seen, coincide with the directional quantile contours of Section 12.2. This, from many points of view, is a very appealing property. However, it also has some less attractive consequences, which mainly originate in the linear foundations of both concepts – namely, the very special role of hyperplanes in their definition. Among those disturbing consequences are the affine invariance (equivariance) and convexity of the depth/quantile contours. For $d \geqslant 2$, those features indeed do not resist any nonlinear transformation, even the continuous monotone increasing marginal ones. This is incompatible with one of the core properties of univariate quantiles: equivariance under *order-preserving transformations* – namely, the fact that the quantile of order τ of a continuous monotone increasing transformation $T(Y)$ of Y is the value $T(q_\tau^Y)$ of the same transformation computed at q_τ^Y, the quantile of order τ of Y.

Convexity, moreover, leads to quite unnatural quantile contours, for example, for distributions with non-convex level sets. Figure 12.8(a) exhibits some empirical halfspace depth contours for a sample from a "banana-shaped" distribution. As quantile contours, they clearly cannot account for the banana shape of the distribution, and the deepest point (playing, in the quantile terminology, the role of a median) is not really central to the sample.

Those drawbacks of halfspace depth, hence of the directional quantiles described in Section 12.2, also extend to most other concepts of statistical depth; spatial quantiles similarly do a pretty poor job (see Figure 12.8(b)). These problems were the main motivation behind

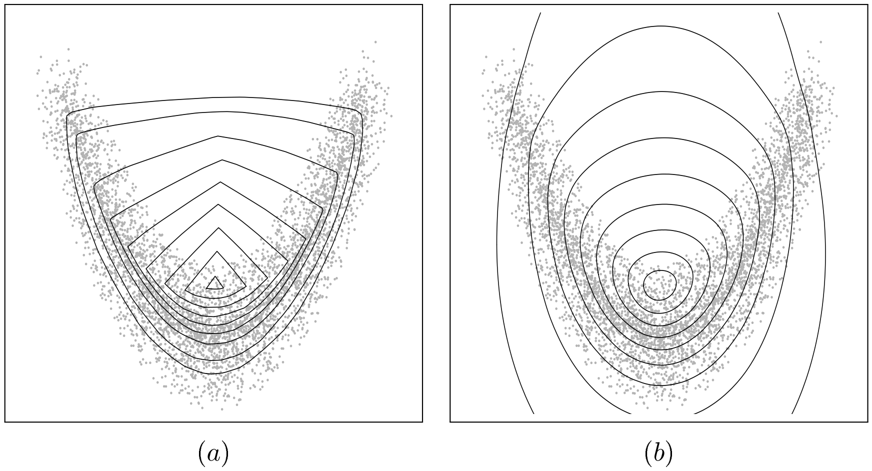


FIGURE 12.8
(a) Some empirical halfspace depth contours and (b) some spatial quantile contours obtained from 4899 simulated observations from a "banana-shaped" distribution.

the concept of Monge–Kantorovich depth, proposed by Chernozhukov et al. (2017) and based on measure transportation ideas, which we now briefly describe.

12.3.3.2 Monge–Kantorovich quantiles

The simplest and most intuitive formulation of the measure transportation problem is as follows. Let P_1 and P_2 be two probability measures over (for simplicity) $(\mathbb{R}^d, \mathcal{B}^d)$. Denote by $L : \mathbb{R}^{2d} \to [0, \infty]$ a Borel-measurable loss function such that $L(\mathbf{x}_1, \mathbf{x}_2)$ represents the cost of transporting $\mathbf{x}_1$ to $\mathbf{x}_2$. Monge's formulation of the optimal transportation problem is: find a measurable transport map $T_{\mathrm{P}_1;\mathrm{P}_2} : \mathbb{R}^d \to \mathbb{R}^d$ achieving the infimum

$$\inf_T \int_{\mathbb{R}^d} L(\mathbf{x}, T(\mathbf{x}))\mathrm{dP}_1, \quad T \text{ subject to } T * \mathrm{P}_1 = \mathrm{P}_2,$$

where $T * \mathrm{P}_1$ denotes the "push-forward of P_1 by T" (more classical statistical notation for this constraint would be $\mathrm{P}_1^{T\mathbf{X}} = \mathrm{P}_2$). A map $T_{\mathrm{P}_1;\mathrm{P}_2}$ that attains this infimum is called an "optimal transport map," or an "optimal transport" for short, mapping P_1 to P_2.

In the sequel, we restrict ourselves to the L^2 loss function $L(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|^2$. The results obtained by Kantorovich imply that, for that L^2 loss, if P_1 and P_2 are absolutely continuous with finite second-order moments, the solution exists, is (almost everywhere) unique, and the gradient of a convex (potential) function – a form of multivariate monotonicity. That type of result is further enhanced in a remarkable theorem by McCann (1995), itself extending a famous result by Brenier (1991), which implies that for any given (absolutely continuous) P_1 and P_2, there exists a P_1-essentially unique element in the class of gradients of convex functions mapping P_1 to P_2; under the existence of finite moments of order 2, moreover, that mapping coincides with the L^2-optimal transport of P_1 to P_2.

In dimension 1, halfspace depth contours are pairs of points of the form

$$\{F^{-1}(\tau), F^{-1}(1 - \tau)\}, \quad \tau \in (0, 1/2];$$

equivalently, letting $F_\pm := 2F - 1$, they are the $F_\pm$-inverse images of the one-dimensional spheres $\{2\tau - 1, 1 - 2\tau\}$, $\tau \in (0, 1/2]$, that is, the pairs

$$F_{\pm}^{-1}(\{-t, t\}), \quad t \in (0, 1).$$

The function $F_{\pm}$ maps $\mathbb{R}$ to the open one-dimensional unit ball $(-1, 1)$ and P to the uniform distribution over the unit ball; it is monotone increasing, hence the gradient (derivative) of a convex function. It thus follows from McCann's theorem that $F_{\pm}$ is the unique gradient of a convex function mapping P to the uniform distribution over the unit ball. Summing up, in dimension 1, halfspace depth contours are the images, by $F_{\pm}^{-1}$, where $F_{\pm}$ is the unique gradient of a convex function mapping P to the uniform distribution over the unit ball, of the spheres $\{\mathbf{t} |\ \|\mathbf{t}\| = \tau\}$ with probability content τ, $\tau \in (0, 1)$.

Turning to dimension d, define $F_{\pm}$ (from $\mathbb{R}^d$ to the open d-dimensional unit ball) as the unique gradient of a convex function mapping P to the uniform distribution over the unit ball;[2] that such an $F_{\pm}$ exists follows from McCann's theorem. The inverse $F_{\pm}^{-1}$ of $F_{\pm}$ qualifies as a quantile function (the Monge–Kantorovich quantile function), and the images, by $F_{\pm}^{-1}$, of the spheres $\{\mathbf{t} |\ \|\mathbf{t}\| = \tau\}$, with probability content $\tau \in (0, 1)$, as quantile contours (the Monge–Kantorovich quantile contours). Figure 12.9 (compare with Figure 12.8) shows that Monge–Kantorovich quantile contours, in contrast to the spatial and directional quantile ones, do pick up the nonconvex features of a distribution.

Each (absolutely continuous) distribution P on $\mathbb{R}^d$ is entirely characterized by the corresponding $F_{\pm}$, which induces a distribution-specific ordering of $\mathbb{R}^d$; that ordering is the combination of

(i) a center-outward ordering $\mathbf{y}_1 \preceq_{\mathrm{P}} \mathbf{y}_2$ if and only if $\|F_{\pm}(\mathbf{y}_1)\| \leqslant \|F_{\pm}(\mathbf{y}_2)\|$, and

(ii) an angular ordering, associated with the cosines

$$\cos_{\mathrm{P}}(\mathbf{y}_1, \mathbf{y}_2) := \big(F_{\pm}(\mathbf{y}_1)\big)'\big(F_{\pm}(\mathbf{y}_2)\big)/\|F_{\pm}(\mathbf{y}_1)\|\,\|F_{\pm}(\mathbf{y}_2)\|.$$

No moment conditions are required.

Unlike the directional quantile contours of Section 12.2, the Monge–Kantorovich ones are equivariant under order-preserving transformations – here, the class of transformations preserving, for some given P, (i) and (ii) above, that is, any T such that

$$\mathbf{y}_1 \preceq_{\mathrm{P}} \mathbf{y}_2 \quad \text{if and only if} \quad T(\mathbf{y}_1) \preceq_{T*\mathrm{P}} T(\mathbf{y}_2)$$

and

$$\cos_{\mathrm{P}}(\mathbf{y}_1, \mathbf{y}_2) = \cos_{T*\mathrm{P}}(T(\mathbf{y}_1), T(\mathbf{y}_2))$$

for any $\mathbf{y}_1$, $\mathbf{y}_2 \in \mathbb{R}^d$. Those transformations are of the form

$$T_{\mathrm{P,Q}} = \big(F_{\pm}^{\mathrm{Q}}\big)^{-1} \circ F_{\pm}^{\mathrm{P}},$$

where Q ranges over the family of absolutely continuous distributions over $\mathbb{R}^d$, $\big(F_{\pm}^{\mathrm{Q}}\big)^{-1}$stands for the corresponding Monge–Kantorovich quantile function, and $\big(F_{\pm}^{\mathrm{P}}\big)^{-1}$for the one associated with P. Equivariance trivially follows from the fact that

$$T_{\mathrm{P,Q}} * \mathrm{P} = \mathrm{Q}, \quad \text{hence } F_{\pm}^{T_{\mathrm{P,Q}}*\mathrm{P}} \circ T_{\mathrm{P,Q}} = F_{\pm}^{\mathrm{P}}.$$

Note that affine equivariance in general does not hold, since affine transformations, for general P, are no longer order-preserving.

[2]Here and in the sequel, "uniform over the unit ball" means the product measure of a uniform over the unit sphere $\mathcal{S}^{d-1}$ with a uniform over the unit interval of radial distances.

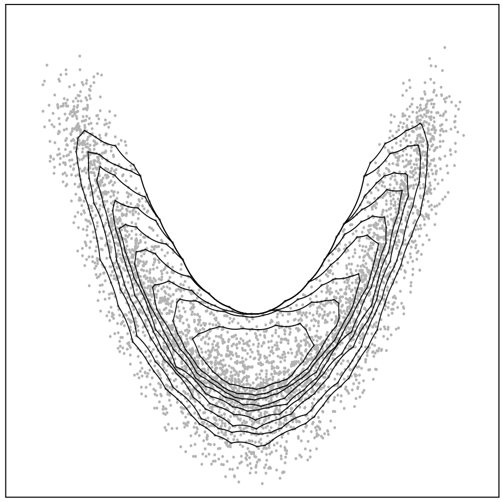


FIGURE 12.9
The same "banana-shaped" data as in Figure 12.8, with Monge–Kantorovich contours (for several quantile levels).

We refer the reader to Chernozhukov et al. (2017) for a description of the empirical versions of $F_{\pm}$ and their consistency properties.

All definitions above are about location ($p = 0$) only. Regression versions (and much more) are the subject of ongoing research, either along the lines of the local linear approach of Section 12.2.2.2, or along those developed by Carlier et al. (2016) who are considering another measure transportation-based approach where the reference uniform distribution is over the unit cube rather than the unit ball (see Section 12.4).

12.4 Some other concepts, and applications

Depth-based quantiles, as well as the elliptical ones, refer to center-outward orderings of $\mathbb{R}^d$: quantile regions are nested, and the reference structure is that of the unit ball. Other approaches are closer to the spirit of the coordinatewise definition of Section 12.2.1.1, where quantiles are indexed by d-tuples $(\tau_1, \ldots, \tau_d) \in (0, 1)^d$; the reference structure there is that of the unit cube. The prototype of that approach is the so-called *Rosenblatt transformation* (Rosenblatt, 1952). Cai (2010), extending ideas by Gilchrist (2000), recently proposed, in a Bayesian context, a quantile concept based on general mappings from the unit cube $(0, 1)^d$ to $\mathbb{R}^d$. Combining similar measure transportation ideas as in Chernozhukov et al. (2017) with the uniform distribution over the unit cube rather than the unit ball, Carlier et al. (2016) and Decurninge (2014) define multivariate quantile functions (associated with a distribution P over $\mathbb{R}^d$) and multiple-output quantile regression (in a linear regression setting) based on the inverse of the optimal transports mapping P to the uniform distribution over the unit cube $(0, 1)^d$.[3] The location version of the latter can be seen as a nonlinear version of the very

[3]Note that the Rosenblatt transformation, in general, is not the gradient of any convex function, hence does not belong to the class of optimal transports considered in this context; see, however, Carlier et al. (2010).

popular independent component analysis models. All those approaches crucially depend on the choice of a coordinate system (hence a unit cube). Yet another approach, where quantiles are constructed on the basis of some preexisting partial ordering $\preceq_0$ of $\mathbb{R}^d$, has been proposed by Belloni and Winkler (2011); not surprisingly, the result depends on the choice of $\preceq_0$.

The applications of multiple-output quantile regression methods are without number, in a virtually unlimited number of domains. An immediate byproduct is the detection of multivariate outliers, for example in growth charts or medical diagnoses. Growth charts so far have essentially been based on marginal quantile plots. The multiple-output quantile concepts described here allow for spotting multivariate outliers that do not outlie in any marginal direction, thus providing a much more powerful diagnostic tool; see McKeague et al. (2011) and the references therein, as well as Chakraborty (2003), Cheng and de Gooijer (2007), or Wei (2008) for related applications and empirical studies. Another obvious and so far largely unexplored application is the problem of multivariate value-at-risk assessment in financial and actuarial statistics.

12.5 Conclusion

Quantile regression methods, by aiming to reconstruct the collection of distributions of a d-dimensional response $\mathbf{Y}$ conditional on the values of a set of covariates $\mathbf{X} = \mathbf{x}$, irrespective of the field of application, address one of the central problems of statistics. The major obstacle to extending traditional single-output quantile regression methods to the multiple-output setting has been the lack of a canonical concept of multivariate quantile, itself related to the lack of a canonical ordering of the Euclidean space with dimension $d \geqslant 2$. Ordering $\mathbb{R}^d$ has remained an open problem and an active domain of research for many years, but recent contributions, and the introduction of measure transportation concepts, are bringing appealing solutions to that problem, hence appealing concepts of quantile regression. Many developments can be expected along those promising lines, although it is clear that much work still remains to be done.

The picture provided in this contribution is inevitably incomplete, as the subject is rapidly developing. Desirable as it is, a clear discussion of the advantages and disadvantages of the various approaches described here would be premature, due to the lack of computational and empirical experience with most of them, and the lack of methodological results about the most recent ones. Inferential statistics, moreover, have their limits, and empirical versions of quantile regression methods, at best, provide an alternative and legible version of the data under study: while (by far) more readable, a collection of empirical quantile contours indeed yields roughly the same complexity as the data themselves. Quantile regression, in that respect, is nothing more – but certainly nothing less – than a sophisticated, powerful, and most meaningful tool for data analysis.

Acknowledgments

Marc Hallin acknowledges the support of the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy). The research of Miroslav Šiman was supported by the Czech Science Foundation project GA17-07384S. We both thank Pavel Boček for technical support and Probal Chaudhuri for constructive remarks that helped improve this chapter.

Bibliography

A. Ahidar. *Surfaces quantiles: propriétés convergences et applications.* PhD thesis, Université de Toulouse, 2015.

A. Ahidar-Coutrix and P. Berthet. Convergence of multivariate quantile surfaces. Available from `https://arxiv.org/abs/1607.02604`, 2016.

Z. D. Bai, N. R. Chen, B. Q. Miao, and C. R. Rao. Asymptotic theory of least distances estimates in multivariate linear models. *Statistics*, 21:503–519, 1990.

A. Belloni and R. L. Winkler. On multivariate quantiles under partial orders. *Annals of Statistics*, 39:1125–1179, 2011.

P. Boček and M. Šiman. Directional quantile regression in Octave (and MATLAB). *Kybernetika*, 52:28–51, 2016.

Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44:375–417, 1991.

B. M. Brown. Statistical use of the spatial median. *Journal of the Royal Statistical Society, Series B*, 45:25–30, 1983.

Y. Cai. Multivariate quantile function models. *Statistica Sinica*, 20:481–496, 2010.

G. Carlier, A. Galichon, and F. Santambrogio. From Knothe's transport to Brenier's map. *SIAM Journal on Mathematical Analysis*, 41:2554–2576, 2010.

G. Carlier, V. Chernozhukov, and A. Galichon. Vector quantile regression. *Annals of Statistics*, 44:1165–1192, 2016.

A. Chakraborty and P. Chaudhuri. The spatial distribution in infinite-dimensional spaces and related quantiles and depths. *Annals of Statistics*, 42:1203–1231, 2014.

B. Chakraborty. On multivariate quantile regression. *Journal of Statistical Planning and Inference*, 110:109–132, 2003.

B. Chakraborty and P. Chaudhuri. On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proceedings of the American Mathematical Society*, 124:2539–2547, 1996.

M. Chaouch, A. Gannoun, and J. Saracco. Estimation de quantiles géométriques conditionnels et non conditionnels. *Journal de la Société Française de Statistique*, 150:1–27, 2009.

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

Y. Cheng and J. de Gooijer. On the uth geometric conditional quantile. *Journal of Statistical Planning and Inference*, 137:1914–1930, 2007.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45:223–256, 2017.

A. Decurninge. Multivariate quantiles and multivariate l-moments. arXiv:1409.6013, 2014.

W. G. Gilchrist. *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC, Boca Raton, FL, 2000.

J. C. Gower. Algorithm AS78: the mediancenter. *Applied Statistics*, 23:446–470, 1974.

J. B. S. Haldane. Note on the median of a multivariate distribution. *Biometrika*, 35:414–415, 1948.

M. Hallin and M. Šiman. Elliptical multiple-output quantile regression and convex optimization. *Statistics & Probability Letters*, 109:232–237, 2016.

M. Hallin, D. Paindaveine, and M. Šiman. Multivariate quantiles and multiple-output regression quantiles: From l_1 optimization to halfspace depth (with discussion). *Annals of Statistics*, 38:635–669, 2010.

M. Hallin, Z. Lu, D. Paindaveine, and M. Šiman. Local constant and local bilinear multiple-output quantile regression. *Bernoulli*, 21:1435–1466, 2015.

D. Hlubinka and M. Šiman. On elliptical quantiles in the quantile regression setup. *Journal of Multivariate Analysis*, 116:163–171, 2013.

D. Hlubinka and M. Šiman. On generalized elliptical quantiles in the nonlinear quantile regression setup. *TEST*, 24:249–264, 2015.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

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

L. Kong and I. Mizera. Quantile tomography: using quantiles with multivariate data. arXiv:0805.0056v1, 2008.

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

R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80:309–324, 1995.

I. W. McKeague, S. López-Pintado, M. Hallin, and M. Šiman. Analyzing growth trajectories. *Journal of Developmental Origins of Health and Disease*, 2:322–329, 2011.

I. Mizera. On depth and deep points: A calculus. *Annals of Statistics*, 30:1681–1736, 2002.

D. Paindaveine. A canonical definition of shape. *Statistics and Probability Letters*, 78: 2240–2247, 2008.

D. Paindaveine and M. Šiman. On directional multiple-output quantile regression. *Journal of Multivariate Analysis*, 102:193–212, 2011.

D. Paindaveine and M. Šiman. Computing multiple-output regression quantile regions. *Computational Statistics and Data Analysis*, 56:841–853, 2012a.

D. Paindaveine and M. Šiman. Computing multiple-output regression quantile regions from projection quantiles. *Computational Statistics*, 27:29–49, 2012b.

M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472, 1952.

P. J. Rousseeuw and M. Hubert. Regression depth. *Journal of the American Statistical Association*, 94:388–402, 1999.

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

R. Serfling. A depth function and a scale curve based on spatial quantiles. In Y. Dodge, editor, *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, pages 25–38. Birkhäuser, Basel, 2002b.

R. Serfling. Depth. In A. H. El-Shaarawi and W. Piegorsch, editors, *Encyclopedia of Environmetrics*, pages 636–641. Wiley, Chichester, 2nd edition, 2012.

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

J. W. Tukey. Mathematics and the picturing of data. In R. D. James, editor, *Proceedings of the International Congress of Mathematicians*, pages 523–531. Canadian Mathematical Congress, Vancouver, 1975.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence*. Springer, New York, 1996.

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

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