Measuring Nonfundamentalness for Structural VARs

Marc Hallin
SBS-EM, ECARES, Université libre de Bruxelles

Miroslav Siman
Czech Academy of Sciences

January 2016

ECARES working paper 2016-03
Marc Hallin, ECARES, Université libre de Bruxelles, Belgium

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

Multiple-Output Quantile Regression
0.1 Multivariate quantiles, and the ordering of $\mathbb{R}^d$, $d \geq 2$

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

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

This, under absolute continuity with nonvanishing density (which, for simplicity, we henceforth throughout assume), yields, for $\tau \in (0, 1)$, what we can call the “traditional definition”
(a) **Regression quantiles: traditional definition.** The regression quantile of order $\tau$ of $Y$ (relative to the vector of covariates $X$ with values in $\mathbb{R}^p$) is the mapping $x \mapsto q_\tau(x) := F^{-1}(\tau) \quad x \in \mathbb{R}^p$ (1)

where $F(y) := P(Y \leq y | X = x)$.

The same concept, under finite moments of order one, also admits the $L_1$ characterization

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

$$E[\rho_\tau(Y - q) | X = x]$$ (2)

over $q \in \mathbb{R}$; the function $z \mapsto \rho_\tau(z) := (1 - \tau)|z| I[z < 0] + \tau z I[z \geq 0]$, 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 $\tau$ of $Y$ (relative to the vector of covariates $X$ with values in $\mathbb{R}^p$) is the hyperplane with equation

$$y = q_\tau(x) = \alpha_\tau + \beta_\tau'x$$

where $\alpha_\tau$ and $\beta_\tau$ are the minimizers, over $(a, b') \in \mathbb{R}^{p+1}$, of

$$E[\rho_\tau(Y - a - b'X)].$$ (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 $(Y_1, X_1), \ldots, (Y_n, X_n)$ an $n$-tuple of points in $\mathbb{R}^{p+1}$; the corresponding empirical regression quantile hyperplane of order $\tau$ is the hyperplane with equation

$$y = q_\tau^{(n)}(x) = \alpha_\tau^{(n)} + \beta_\tau^{(n)}'x$$

where $\alpha_\tau^{(n)}$ and $\beta_\tau^{(n)}$ are the minimizers, over $(a, b') \in \mathbb{R}^{p+1}$, of

$$\sum_{i=1}^n \rho_\tau(Y_i - a - b'X_i).$$ (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 $Y$, in most situations of practical interest, takes values in $\mathbb{R}^d$, with $d \geq 2$ (multiple-output case). An extension to $d \geq 2$ of the definitions above is thus extremely desirable. Unfortunately, they all exploit the canonical ordering of $\mathbb{R}$. Such an ordering no longer exists in $\mathbb{R}^d$, $d \geq 2$. 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—has attracted much interest in the literature, and many solutions have been proposed—among them, Möttönen and Oja (1995), Chaudhuri (1996), Koltchinskii (1997), Choi and Marden (1997), Cheng and de Gooijer (2007), Oja (2010), ... to quote only a few. The whole theory of statistical depth and also, in a sense, the theory of copulas are aiming at that 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 (2000, 2012).

### 0.2 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 $d \geq 2$, consists in trying to reduce the multivariate problem to a collection of univariate ones by considering univariate distributions associated with the $d$-dimensional ones. We start with the pure location case ($p = 0$, no covariates) and projection ideas.

#### 0.2.1 Projection methods

0.2.1.1 Marginal (coordinatewise) quantiles

If a coordinate system is adopted, $Y$ writes 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

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

The mapping $(\tau_1, \ldots, \tau_d) \mapsto q_{\tau_1, \ldots, \tau_d}$ actually is the inverse of the copula transform; in particular, $q_{1/2, \ldots, 1/2}$ yields the componentwise median. An empirical version of that mapping is readily obtained by considering the empirical marginal distributions of any observed $n$-tuple $Y_1, \ldots, Y_n$. 
This definition, however, does not provide an ordering of $\mathbb{R}^d$, but rather a $d$-tuple of (marginal) orderings. Moreover, $q_{\tau_1, \ldots, \tau_d}$ very crucially depends on the coordinate system adopted, and is not even rotation-equivariant.

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

0.2.1.2 Quantile biplots

Marginal quantiles actually are obtained by projecting $P$ (equivalently, $Y$) on $d$ mutually orthogonal straight lines (characterized by the canonical orthonormal basis $(u_1, \ldots, u_d)$) through some origin. If the influence of the arbitrary choice of a basis is to be removed, one also may like to look at projections onto all unit vectors $u \in S^{d-1}$ through some given origin. This approach is investigated in Kong and Mizera (2008) and, in some detail, in Section 2 of Ahidar (2015).

For each $(u, \tau) \in (0, 1)$, the univariate distribution of $u'Y$ yields a well-defined quantile of order $\tau$, $q_{\tau u}$, say. Define, for $\tau \in [1/2, 1)$, the directional quantile of order $\tau$ for direction $u$ as the point $q_{\tau u} := q_{\tau u}$. Substituting empirical quantiles $q_{\tau u}^{(n)}$ for the theoretical ones yields empirical counterparts $q_{\tau u}^{(n)}$. The collection for $u$ ranging over the unit sphere $S^{d-1}$ 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, this concept however 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 $q_{\tau u}^{(n)}$ is quite straightforward, the construction of empirical biplots, in principle, requires considering “infinitely many” directions $u$.

For all those reasons, the concept (which no longer appears in Kong and Mizera (2012)) will not be examined any further.

0.2.1.3 Directional quantile hyperplanes and contours

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

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

Denote by $H_{\tau u}$ the halfspace lying above $H_{\tau u}$. The intersection (for given $\tau$) $H(\tau) := \bigcap_{u \in S^{d-1}} H_{\tau u}$ of those halfspaces characterizes 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 \( \tau \) increases, and affine-equivariant.

The index \( \tau \) 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 \( \tau \) and the \( P_Y \)-probability of the quantile region \( \mathcal{H}(\tau) \).

Empirical versions \( H_{(\tau)}^{(n)} \), \( \mathcal{H}(\tau) \), and \( H^{(n)}(\tau) \), as usual, are obtained by replacing the distribution \( P_Y \) of \( Y \) with the empirical measure associated with some observed \( n \)-tuple \( Y_1, \ldots, Y_n \). However, the characterization of a given quantile contour \( H^{(n)}(\tau) \) involves an infinite number of directions \( u \), which of course is not implementable. In order to overcome this, one can compute the \( N \) hyperplanes \( H_{(\tau)}^{(n)} \) associated with a sample (random or systematic) of directions \( u_i \), \( i = 1, \ldots, N \): for absolutely continuous \( Y_1, \ldots, Y_n \), the resulting region \( \mathcal{H}(\tau) := \bigcap_{i=1}^N H_{(\tau)}^{(n)} \) is an approximation of the actual quantile region \( \mathcal{H}(\tau) \) (to which, under mild conditions, it converges as \( N \to \infty \))—a “biased” one, though, since, with probability one, \( \mathcal{H}(\tau) \) strictly includes \( \mathcal{H}(\tau) \) for all \( N \).

### 0.2.1.4 Relation to halfspace depth

Kong and Mizera 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 population).

Recall that the halfspace depth of a point \( y \) with respect to a probability distribution \( P_Y \) (Tukey 1975) is the minimum, over all hyperplanes running through \( y \), of the \( P_Y \)-probabilities of the halfspaces determined by those hyperplanes. The halfspace depth regions \( D(\delta) \) (respectively, the halfspace depth contours \( D(\delta) \)) are the collections of points with halfspace depth larger than or equal to \( \delta \) (respectively, with given halfspace depth \( \delta \)); those regions are convex and nested as depth decreases. The empirical depth of \( y \) with respect to the \( n \)-tuple \( Y_1, \ldots, Y_n \) is defined similarly with the empirical distribution of the \( Y_i \)'s playing the role of \( P_Y \). 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 \( H^{(n)}(\delta) \)'s facets. The definition adopted so far, which is related to the traditional univariate definition (1), does not readily provide a way to identify those directions, though. The directional Koenker-Bassett approach of Section 0.2.2, which extends the univariate \( L_1 \) definition (2), also leads to a numerical determination of the relevant directions.

In the presence of covariates \( p \geq 1 \), the connection with halfspace depth does not help much, as most existing regression depth concepts, inspired by
0.2.2 Directional Koenker-Bassett approach

0.2.2.1 Location case ($p = 0$)

Another directional approach is proposed in Hallin, Paindaveine and Šiman (2010). No projections there, and, rather than Kong and Mizera’s directional version of the traditional definition (1), a directional version of the $L_1$ definition (2) is adopted.

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

$$\begin{align*}
(a_{\tau u}, b'_{\tau u}) &= \arg\min_{(a, b') \in \mathbb{R}^d} E[\rho_\tau(Y_u - b'Y_u^\perp - a)].
\end{align*}$$

The empirical version $\Pi^{(n)}_{\tau u}$ of $\Pi_{\tau u}$, with equation $u'Y = b^{(n)}_{\tau u} + a^{(n)}_{\tau u}$, is obtained by replacing the distribution $P^Y$ of $Y$ with the empirical measure associated with an observed $n$-tuple $Y_1, \ldots, Y_n$:

$$\begin{align*}
(a^{(n)}_{\tau u}, b^{(n)}') &= \arg\min_{(a, b') \in \mathbb{R}^d} \sum_{i=1}^n \rho_\tau(Y_{i,u} - b'Y_{i,u}^\perp - a).
\end{align*}$$

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

Hallin et al. (2010) moreover show 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 0.2.1.3 is that, thanks to the “analytical” nature of...
the L₁ 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 0.2.1.3 may be the only feasible solution. The two points of view, moreover, can be reconciled—see Paindaveine and Šiman (2011).

Figure 0.2.2.1 shows (a) the empirical quantile contours of order τ = 0.2 obtained, from a dataset of n = 49 observations, based on 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 u values on S¹.

The multivariate quantile contours resulting from this directional Koenker-Bassett approach inherit from their relation to halfspace depth the nice geometric features—convexity, connectedness, nestedness, affine-equivariance—of the latter, while bringing to halfspace depth the nice analytical, computational, and probabilistic features of L₁ optimization—tractable asymptotics (Bahadur representation, root-n consistency, and asymptotic normality, etc.), L₁ characterization/optimality, implementable linear programming algorithms, optimization problems byproducts (duality and Lagrange multipliers); see Hallin et al. (2010) for explicit results and details.
**0.2.2.2 (Nonparametric) regression case \((p \geq 1)\)**

A linear regression extension \((p \geq 1)\) of the location concept just described is quite straightforward once the \(L_1\) approach is adopted. Definitions (3) and (4) indeed readily generalize: see Hallin et al. (2010) and Paindaveine and Šiman (2011). Hallin, Lu, Paindaveine and Šiman (2015) rather consider a fully general nonparametric regression setup, where the objective is a reconstruction of the *conditional* (on the value \(X = x\) of some regressor) quantile contours.

Denote by \((X'_1, Y'_1), \ldots, (X'_n, Y'_n)\)' an observed \(n\)-tuple of independent copies of \((X', Y')\)', where \(Y := (Y_1, \ldots, Y_d)'\) is a \(d\)-dimensional response and \(X := (X_1, \ldots, X_p)'\) a \(p\)-dimensional random vector of covariates.

The terminology “response” for \(Y\) and “covariate” for \(X\) clearly indicates that the objective here is an analysis of the \(d\)-dimensional distribution of \(Y\) conditional on \(X\), that is, a full study of the dependence of \(Y\) on \(X\). The relevant quantile hyperplanes, quantile/depth regions and contours of interest thus are the location quantile hyperplanes, quantile/depth regions and contours associated with the \(d\)-dimensional distributions of \(Y\) conditional on \(X\)—that is, the collection, for \(x\) ranging over \(\mathbb{R}^p\), of the hyperplanes, regions and contours associated with the distributions \(P_{Y|X=x}\) of \(Y\) conditional on \(X = x\). These contours indeed completely characterize (under absolutely continuous \(P_{Y|X=x}\)) the conditional distributions of \(Y\), hence the impact on the (multivariate) response \(Y\) of the regressors \(X\). When plotted against \(x\) (which is possible for \(d + p \leq 3\) only), those contours yield quantile regression “tubes”: see Figure 0.2.2.2.
Two consistent methods are provided in Hallin et al. (2015) for the estimation of such tubes: a local constant method, and a local bilinear one. Both estimators are based on a weighted (kernel-based) version of the location case, still leading to a parametrized linear programming problem with directional parameter $u$ ranging over the unit sphere $S^{d-1}$. Bahadur representations of the resulting estimators are established under appropriate technical conditions on the joint distributions of $(X', Y')'$, the kernel and the bandwidth defining the weights; those representations entail consistency and asymptotic normality.

Figure 0.2.2.3. Local constant empirical quantile regression tubes.

Local constant quantile contours yield, for given $\tau$ and a selected value $x_0$ of $x$, a “horizontal polygonal tube” (Figure 0.2.2.3) in $\mathbb{R}^{d+p}$, the interpretation of which is valid at $x_0$ only, and provides no information on the way the conditional distribution of $Y$ is varying in the neighborhood of $x_0$.

Local bilinear quantile contours are more informative, since they incorporate information on the derivatives with respect to $x$ of the coefficients of the conditional quantile hyperplanes; they also should 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 0.2.2.4, are no longer polygonal cylinders, but piecewise ruled quadrics.

Figure 0.2.2.5 shows the empirical contours ($\tau = 0.2$ and $0.4$) constructed via the local constant (a) and local bilinear (b) methods for a set of $n = 4899$ observations ($d = 2$, $p = 1$) simulated from the bivariate heteroscedastic
Figure 0.2.2.4. Local bilinear empirical quantile regression tubes.

Regression model

\[(Y_1, Y_2)' = (X, X^2)' + 0.5 (1 + 3|\sin(\pi X/2)|)(e_1, e_2)' \quad (5)\]

where \(X\) is uniform over \((-2, 2)\) and \((e_1, e_2)' \sim \mathcal{N}(0, I)\) is bivariate normal. The axes are those of the response space, and (unlike in Figure 0.2.2.4) 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.

\section*{0.3 Direct approaches}

All multiple-output quantile regression concepts presented so far were based on directional extensions of the usual single-output ones. More direct approaches are possible, though, along two main lines. The first one consists in substituting, in the traditional Koenker-Bassett (1978) definition, ellipsoids for hyperplanes, and the “above/below” indicators with “outside/inside” ones. The second approach, inspired by the relation (Section 0.2.1.4) between directional quantiles and halfspace depth, is based on recent measure transportation-related concepts (Chernozhukov, Galichon, Hallin and Henry 2015) of Monge-Kantorovich depth and quantiles.
Figure 0.2.2.5. The empirical contours ($\tau = 0.2$ and 0.4) obtained, via (a) the local constant method and (b) the local bilinear method, for $n = 4899$ observations from model (5), along (c) with their (exact) population counterparts; the dot at the center of the population contours is both the conditional mean and the conditionally deepest point.
0.3.1 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). As in Section 0.2, we start with the pure location case \((p = 0, \text{no covariates})\) before turning to the general regression case.

0.3.1.1 Location case

The basic idea behind the concept is intuitively quite simple and straightforward: instead of (Section 0.2.2.1) the inner envelope of a collection of directional Koenker-Bassett \(\tau\)-quantile hyperplanes minimizing the expected value of a directional check function penalty, rather 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 following definition of a multivariate (location) elliptical \(\tau\)-quantile as the ellipsoid

\[
\mathcal{E}^{\text{loc}}_{\tau} = \mathcal{E}^{\text{loc}}_{\tau}(Y) := \{ y \in \mathbb{R}^d : y' A_{\tau} y + y' b_{\tau} - c_{\tau} = 0 \},
\]

where \(A_{\tau} \in \mathbb{R}^{d \times d}, b_{\tau} \in \mathbb{R}^{d \times 1}\), and \(c_{\tau} \geq 0\) minimize, subject to \(A_{\tau}\) symmetric and positive semidefinite with determinant one (a shape matrix in the sense of Paindaveine (2008)); the objective function

\[
\Psi_{\tau}^{\text{loc}}(A, b, c) := E \rho_{\tau}(Y'AY + Y'b - c);
\]

as usual, \(\rho_{\tau}\) 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 \(\rho_{\tau}\) in (6) is positive or negative according as \(Y\) takes value inside or outside the ellipsoid with equation \(y' A_{\tau} y + y' b = c\).

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

This definition certainly does not characterize an elliptical quantile as the solution of a linear programming problem; nor does it, as it stands, 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 \(\det(A_{\tau}) = 1\) into \((\det(A_{\tau}))^{1/d} \geq 1\): unlike \(A \mapsto \det(A)\), the function \(A \mapsto (\det(A))^{1/d}\) is concave on the cone of symmetric positive semidefinite matrices, and it can be shown that this convex optimization problem and the original non-convex one share the same solution. That solution, moreover, is unique under absolutely continuous distributions with nonvanishing densities and finite moments of order two.
0.3.1.2 Linear regression case

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

\[
(y - \beta_{\tau} - B_{\tau} x)' A_{\tau} (y - \beta_{\tau} - B_{\tau} x) - \gamma_{\tau} = 0
\]

with some \( A_{\tau} \in \mathbb{R}^{d \times d} \), \( \beta_{\tau} \in \mathbb{R}^{d \times 1} \), \( B_{\tau} \in \mathbb{R}^{d \times p} \), and \( \gamma_{\tau} \geq 0 \). The trouble is that the corresponding objective function

\[
E_{\rho_{\tau}}((Y - \beta - B X)' A(Y - \beta - B X) - \gamma)
\]

is not convex in \( \beta \) and \( B \), so that its minimization with respect to \( A, \beta, B \), and \( \gamma \) 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

\[
E_{\tau}^{\text{reg}} := \left\{ (y', x')' \in \mathbb{R}^{d+p} : \quad (y - \beta_{\tau} - B_{\tau} x)' A_{\tau} (y - \beta_{\tau} - B_{\tau} x) - (\gamma_{\tau} + c' x + x' C_{\tau} x) = 0 \right\}
\]

of an elliptical regression quantile \( E_{\tau}^{\text{reg}} = E_{\tau}^{\text{reg}} (Y, X) \), where a quadratic form of covariate-driven scale is allowed, and \( A_{\tau}, \beta_{\tau}, B_{\tau}, \gamma_{\tau}, c, \) and \( C_{\tau} \) jointly minimize

\[
\Psi_{\tau}^{\text{reg}} := E_{\rho_{\tau}}((Y - \beta - B X)' A(Y - \beta - B X) - (\gamma + c' X + X' C X))
\]

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

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

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

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

In this new parametrization, the elliptical regression quantile \( E_{\tau}^{\text{reg}} \) can be expressed as

\[
E_{\tau}^{\text{reg}} = \{(y', x')' \in \mathbb{R}^{d+p} : r(y, x, M) = 0\}
\]

where

\[
r(y, x, M) := y'M^1 y + x'M^2 x + x'M^3 y + M^4 y + M^5 x + M^6 = (y - \beta - B x)' A (y - \beta - B x) - (\gamma + c' x + x' C x),
\]
(\(r\) is thus positive outside, and negative inside, the ellipsoid with equation \(r = 0\)) and \(\mathbf{M}_\tau := (\mathbf{M}_1^\tau, \ldots, \mathbf{M}_6^\tau)\) jointly minimize

\[ \Psi_{\tau}^{\text{reg}} = \Psi_{\tau}^{\text{reg}}(\mathbf{M}) := \Psi_{\tau}^{\text{reg}}(\mathbf{M}_1^\tau, \ldots, \mathbf{M}_6^\tau) = E_{\nu}(r(Y, X, \mathbf{M})) , \]

subject to \((\det(\mathbf{M}_1^\tau))^{1/d} \geq 1\); as in the location case, positive homogeneity of \(\Psi_{\tau}^{\text{reg}}(\mathbf{M}_1^\tau, \ldots, \mathbf{M}_6^\tau)\) 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 \(\tau\), and that \(E[(Y', X')' | r \geq 0] = E[(Y', X')' | r < 0]\), 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 \((Y'_i, X'_i)'\), \(i = 1, \ldots, n\), 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 (8). Classical results (such as Theorem 5.14 of van der Vaart and Wellner (1998)) then guarantee basic convergence, as \(n \to \infty\), of the vector

\[ \mathbf{m}_{\tau,n} := (\text{vec}(\mathbf{M}_1^\tau)'n, \text{vec}(\mathbf{M}_2^\tau)'n, \text{vec}(\mathbf{M}_3^\tau)'n, \mathbf{M}_4^\tau, \mathbf{M}_5^\tau, \mathbf{M}_6^\tau)'n \]

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

\[ \mathbf{m}_{\tau} := (\text{vec}(\mathbf{M}_1^\tau)', \text{vec}(\mathbf{M}_2^\tau)', \text{vec}(\mathbf{M}_3^\tau)', \mathbf{M}_4^\tau, \mathbf{M}_5^\tau, \mathbf{M}_6^\tau)' . \]

### 0.3.2 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 to two important cases: halfspace depth, the relation to directional quantiles of which has been outlined in Section 0.2, and the more recent concept of Monge-Kantorovich depth.

#### 0.3.2.1 Halfspace depth quantiles

Halfspace depth contours, as we have seen, coincide with the directional quantile contours of Section 0.2. 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—viz. 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 $d \geq 2$, those features indeed do not resist any nonlinear transformation, even the continuous monotone increasing marginal ones. This strongly violates one of the core properties of univariate quantiles: equivariance under \textit{order-preserving transformations}—namely, the fact that the quantile of order $\tau$ of a continuous monotone increasing transformation $T(Y)$ of $Y$ is the value $T(q^Y_\tau)$ of the same transformation computed at $q^Y_\tau$, the quantile of order $\tau$ of $Y$.

Convexity moreover leads to quite unnatural quantile contours, e.g., for distributions with non-convex level sets. Figure 0.3.2.1 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.

![Figure 0.3.2.1. Some empirical halfspace depth contours (4899 simulated i.i.d. observations from a “banana-shaped distribution”).](image)

Those drawbacks of halfspace depth, hence of the directional quantiles described in Section 0.2, were the main motivation behind the concept of Monge-Kantorovich depth proposed by Chernozhukov, Galichon, Hallin, and Henry (2015), based on measure transportation ideas, which we now briefly describe.
0.3.2.2 Monge-Kantorovich quantiles

The simplest and most intuitive formulation of the measure transportation problem is as follows. Let $P_1$ and $P_2$ denote two probability measures over (for simplicity) $(\mathbb{R}^d, \mathcal{B}^d)$. Let $L : \mathbb{R}^{2d} \to [0, \infty]$ be a Borel-measurable loss function such that $L(x_1, x_2)$ represents the cost of transporting $x_1$ to $x_2$. Monge’s formulation of the optimal transportation problem is: find a measurable transport map $T_{P_1; P_2} : \mathbb{R}^d \to \mathbb{R}^d$ achieving the infimum

$$
\inf_T \int_{\mathbb{R}^d} L(x, T(x)) dP_1 \quad \text{subject to} \quad T \ast P_1 = P_2
$$

where $T \ast P_1$ denotes the “push forward of $P_1$ by $T$”—more classical statistical notation for this constraint would be $P_1^T = P_2$. A map $T_{P_1; P_2}$ that attains this infimum is called an “optimal transport map”, in short, an “optimal transport” mapping $P_1$ to $P_2$.

In the sequel, we restrict to the $L^2$ loss function $L(x_1, x_2) = \|x_1 - x_2\|^2$. 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 (a.e.) unique, and the gradient of a convex (potential) function—a form of multivariate monotonicity. That type of result is further enhanced by a remarkable theorem by McCann (1995), itself generalizing a 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 two, that mapping moreover coincides with the $L^2$-optimal transport of $P_1$ to $P_2$.

In dimension one, halfspace depth contours are couples 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 couples

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

The function $F_\pm$ is mapping $\mathbb{R}$ 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 one, 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 $\{t | \|t\| = \tau\}$ with probability content $\tau$, $\tau \in (0, 1)$.

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$^1$—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 $\{t \mid \|t\| = \tau\}$, with probability content $\tau$, $\tau \in (0,1)$, as quantile contours—the Monge-Kantorovich quantile contours. Figure 0.3.2.2 (compare with Figure 0.3.2.2) shows that Monge-Kantorovich quantile contours, contrary to the directional quantile ones, do pick up the non-convex 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 $y_1 \preceq_P y_2$ iff $\|F_{\pm}(y_1)\| \leq \|F_{\pm}(y_2)\|$, and

(ii) an angular ordering, associated with the cosines

$$\cos_P(y_1, y_2) := \langle F_{\pm}(y_1)\rangle'\langle F_{\pm}(y_2)\rangle/\|F_{\pm}(y_1)\| \|F_{\pm}(y_2)\|.$$

No moment conditions are required.

---

$^1$Here and in the sequel, “uniform over the unit ball” means the product measure of a uniform over the unit sphere $S^{d-1}$ with a uniform over the unit interval of radial distances.
Unlike the directional quantile contours of Section 0.2, the Monge-Kantorovich ones are equivariant under order-preserving transformations—here, the class of transformations preserving, for some given P, (i) and (ii) above, i.e., any $T$ such that

$$y_1 \preceq_P y_2 \text{ iff } T(y_1) \preceq_{T \ast P} T(y_2)$$

and

$$\cos_P(y_1, y_2) = \cos_{T \ast P}(T(y_1), T(y_2))$$

for any $y_1, y_2 \in \mathbb{R}^d$. Those transformations are of the form

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

where $Q$ ranges over the family of absolutely continuous distributions over $\mathbb{R}^d$, and $(F_{Q}^{-1})^T$ stands for the corresponding Monge-Kantorovich quantile function (hence $(F_{P}^{-1})^T$ for the one associated with $P$). Equivariance trivially follows from the fact that $T_{P, Q} * P = Q$, hence $F_{Q}^{-1} \circ (T_{P, Q} * P) = F_{P}^{-1}$. Note that affine-equivariance in general does not hold, since affine transformations, for general $P$, are no longer order-preserving.

Empirical versions of $F_{\pm}$ and consistency results are available in Chernozhukov et al. (2015).

All definitions above are about location ($P = 0$) only; regression versions (and much more) are the subject of ongoing research.

## 0.4 Some other concepts, and applications

Depth-based quantiles, as well as the elliptical ones, refer to single-indexed concepts and a center-outward ordering 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 0.2.1.1, 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. (2015) with the uniform distribution over the unit-cube rather than the unit ball, Carlier, Chernozhukov and Galichon (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-in-the-parameter setting) based on the inverse of the optimal transports mapping $P$ to the uniform distribution over the unit cube $(0, 1)^d$.\(^2\)

\(^2\)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.
The location version of the latter can be seen as a nonlinear version of the very popular ICA (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, e.g. in growth charts or medical diagnoses. Growth charts so far have been, essentially, 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, López-Pintado, Hallin, and Siman (2011) and the references therein, as well as Wei (2008) for a related approach. Another obvious and so far largely unexplored application is the problem of multivariate value-at-risk assessment in financial and actuarial statistics.

### 0.5 Conclusion

Quantile regression methods, by aiming at a reconstruction of the collection of distributions of a response \( Y \), with values in \( \mathbb{R}^d \), conditional on a set of covariates \( X = x \), irrespective of the field of application, are addressing one of the most 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 an adequate concept of multivariate quantile— itself related to the lack of a canonical ordering of the Euclidean space with dimension \( d \geq 2 \). Ordering \( \mathbb{R}^d \) has remained an open problem and an active domain of research for many years, but recent contributions are bringing appealing solutions, hence appealing concepts of quantile regression.

Inferential statistics, however, have their limits, and empirical versions of quantile regression methods, at best, provide an alternative and legible version of the data: while (by far) more readable, a collection of empirical quantile contours indeed yields 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.

**Acknowledgements.** Marc Hallin acknowledges the support of the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy), a Crédit aux Chercheurs of the Fonds National de la Recherche Scientifique, and the Discovery grant DP150100210 of the Australian Research
Council. The research of Miroslav Šiman was supported by the Czech Science Foundation project GA14-07234S.
Bibliography


