# A note on multivariate $M$-quantiles

Jens Breckling, Philip Kokic and Oliver Lübke[1]

**Abstract**

The extension of $M$-quantiles to a multivariate setting was originally introduced by Breckling and Chambers (1988). It turns out that in certain situations their definition does not produce intuitive results. We present an alternative definition based on a generalisation of the univariate estimating equations for $M$-quantiles that overcomes these shortcomings and includes the spatial median and the multivariate sample mean as special cases.

**Key words:** $M$-estimates; multivariate ordering; affine equivariance; robust statistics

## 1 Introduction

The idea of extending $M$-estimates and $M$-quantiles to a multivariate setting was originally proposed by Breckling and Chambers (1988). Their interest was to provide a robust technique for summarising the distribution of multidimensional data. Since they generalised standard $M$-estimates and univariate quantiles in an intuitive and simple manner their definition is preferable to the rather ad-hoc techniques for summarising multidimensional data presented in Tukey (1977). Furthermore, multivariate ($M$-) quantiles are an attractive alternative to quantiles based on an estimate of the multivariate density as they are simple to compute and do not suffer from the well-known 'curse of dimensionality' problem inherent in most nonparametric density estimation procedures (Scott 1992). A further use of multivariate $M$-quantiles is as a probability based ordering technique for multidimensional data, and hence they have applications in outlier detection and for performance measurement, see Kokic, Chambers, Breckling, and Beare (1997). A related technique for ordering multivariate data is the concept of depth contours as discussed by He and Wang (1997), which can also be extended to regression, see Rousseeuw and Hubert (1999).

In their definition Breckling and Chambers (1988) consider a sample $y_1, \ldots, y_n$ of observations, $y_i \in \mathbb{R}^k$, a given value $0 < p < 1$ and a vector $r \in \mathbb{R}^k$ of length 1, interpreted as a direction. Their aim was to define a vector $\theta = \theta_{p,r} \in \mathbb{R}^k$ that can be interpreted as the $k$-dimensional $M$-quantile of the sample of observations with respect to the direction $r$ and the probability $p$. The definition was based on a simple generalisation of the one-dimensional loss function for quantiles and

[1]Insiders Financial Solutions GmbH, Wilh.-Th.-Römheld-Str. 18, 55130 Mainz, Germany.

$M$-quantiles. Unfortunately their definition does not produce intuitive results in certain situations. For example, the estimated quantiles are often situated outside the convex hull of the data.

Chaudhuri (1996) presents a definition of multivariate geometric quantiles which, however, can be shown to coincide with a special case of the definition of Breckling and Chambers (1988) and hence is subject to the same problem.

The main purpose of this paper is to present an alternative definition based on a multivariate generalisation of the univariate estimating equations for quantiles and $M$-quantiles. Empirical evidence suggests that the new definition largely overcomes the shortcomings of the approach of Breckling and Chambers (1988). Furthermore, these $M$-quantiles are not restricted to lie on specific pre-specified surfaces and are thus free to represent and adapt to the true shape of the data. This is in contrast to the 'median ball' approach presented in Averous and Meste (1997).

# 2 The Breckling/Chambers approach

## 2.1 Definition

In the univariate case, the ordinary $p^{\text{th}}$ quantile can be defined as the scalar that minimizes the sum of the residuals weighted with an appropriate loss function. This idea can easily be extended to the definition of a univariate $M$-quantile by using a weighted form of a primitive of Huber's $M$-function as the loss function.

The main difficulty in a multivariate adaption of this concept is that there does not exist a natural ordering in $k$ dimensions, $k > 1$. Since distinguishing between the lower and upper quartiles in the one dimensional case, for example, corresponds to an indirect specification of direction, it immediately follows that the definition of any quantile other than the median requires such a specification.

So a directional unit vector $r$ is introduced with respect to which the $M$-quantile is defined. Without loss of generality one can assume $p \leq \frac{1}{2}$ because once a multivariate $M$-quantile $\theta = \theta_{p,r}$ is defined, one simply sets $\theta_{1-p,r} := \theta_{p,-r}$. Let $\alpha_i$ denote the angle between $y_i - \theta$ and $r$, so $\cos\alpha_i = (y_i - \theta)'r / \|y_i - \theta\|$, and let $\zeta := 1 - 2p$. Breckling and Chambers (1988) defined $\theta_{p,r}$ as the vector that minimizes the sum

$$\sum_{i=1}^{n}(1 - \zeta\cos\alpha_i)\cdot\rho(y_i - \theta), \text{ where } \rho(x) = \begin{cases} \frac{1}{2c}\|x\|^2 & \text{if } \|x\| < c, \\ \|x\| - \frac{c}{2} & \text{if } \|x\| \geq c. \end{cases} \tag{2.1}$$

The basic reasoning behind this definition is to introduce a weighting scheme for the residuals $y_i - \theta$ depending both on their length and the angle they form with $r$, and which produces the ordinary $M$-quantile in one dimension.[1] The parameter $c \geq 0$ determines where this weighting changes from an expectile type weighting to an ordinary quantile type weighting. If $c = 0$ pure quantile estimation results, while

[1]The definition of Breckling and Chambers (1988) is in fact slightly more general, allowing for a broader class of loss functions.

$c \to \infty$ corresponds to pure expectile estimation.[2] When $r$ moves around the whole $(k-1)$-dimensional unit sphere the resulting set of corresponding $M$-quantiles is a $(k-1)$-dimensional closed surface embedded within $k$-dimensional Euclidean space.

### 2.2 An example revealing shortcomings of the Breckling/Chambers definition

While the above definition yields acceptable results in many situations, in certain circumstances the result can be very different from what appears to be intuitive. A natural quality that a multivariate $M$-quantile should have is that it should lie within the convex hull of the sample. This, however, is not always the case as the following example shows. We consider a two-dimensional 'cigar shaped' data set, see figure 1. The sample was generated by adding normally distributed random terms with mean 0 and variance 0.01 to 200 equidistant points on the interval $[-1,1]$. To keep the interpretation of the results simple we compute the pure multivariate quantiles (i.e. we set $c=0$) for $p=0.05$. The directional vector $r$ is moved around the whole unit circle.

While the results seem reasonable for $r$ pointing more or less directly left or right, the results for $r$ pointing straight up or down are far from what we intuitively expect. To analyse this behaviour let us consider the case $r = \binom{0}{1}$. The sum that is to be minimized in order to obtain the quantile is given by (2.1). To minimize this sum one has to consider the two terms $1-\zeta\cos\alpha_i$ and $\|y_i-\theta\|$. The first of these gets smaller the closer the $\alpha_i$'s get to zero, that is, the more negative the $y$-component of $\theta$ is. Exactly the opposite is true for the second term: it gets *larger* when $\theta$ moves downwards. It appears that for small values of $p$ the first term dominates the second, pulling the resulting quantile down too far.

This can also be seen when looking at the problem in a different way, namely by finding the root of the derivative of (2.1) with respect to $\theta$. This yields

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i-\theta}{\|y_i-\theta\|}=\zeta r \tag{2.2}$$

as the estimating equation for $\theta$. For small $p$ the right hand side is almost equal to the vector $r$; more precisely for $r=\binom{0}{1}$ it is equal to $\binom{0}{\zeta}$. The left hand side is the average of the unit vectors pointing from $\theta$ to the sample elements. Obviously the only way this average can be equal to the right hand side is to move $\theta$ very far down.

# 3 Alternative approach

The above approach is based on finding the minimum of the sum of the residuals weighted with a certain *loss function*. We now introduce an approach that is based on finding the root of the sum of the residuals weighted with an appropriate *influence*

[2] The expression $c\to\infty$ is to be understood in the sense that $c$ is large enough for the interval $[0,c)$ to contain all values $\|y_i-\theta\|$.

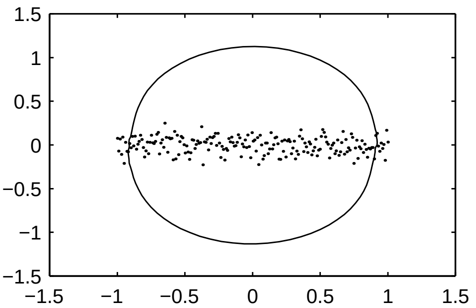


**Figure 1:** Quantiles of cigar shaped sample for $p = 0.05$ ($c = 0$), computed with the Breckling/Chambers approach.

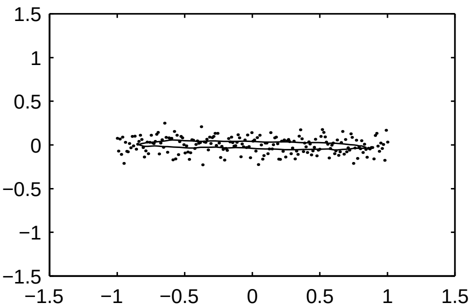


**Figure 2:** Quantiles of cigar shaped sample for $p = 0.05$ ($c = 0$), computed with the new approach.

*function.* This has two advantages: (1) The development of a multivariate $M$-quantile influence function is exactly analogous to how $M$-quantiles are motivated in the one-dimensional case; (2) Modelling the influence function instead of the loss function is a more general approach because for a given loss function there is always a corresponding influence function, while the opposite is not true. In fact, the specific influence function we shall define does *not* have a corresponding loss function.

Using the variables $r$, $\alpha_i$ and $\zeta$ as above, we define the $p^{\text{th}}$ sample $M$-quantile $\theta = \theta_{p,r}$, $0 < p \leq \frac{1}{2}$, as the solution of

$$\sum_{i=1}^{n} (y_i - \theta) w_i = 0, \tag{3.1}$$

where

$$w_i = \begin{cases} \frac{1-\zeta \cos \alpha_i}{c} & \text{if } \|y_i - \theta\| < c, \\ \frac{1-\zeta \cos \alpha_i}{\|y_i - \theta\|} & \text{if } \|y_i - \theta\| \geq c. \end{cases} \tag{3.2}$$

Again, in the univariate case this reduces to the standard definition of an $M$-quantile which in turn includes the pure quantile as a special case. It is easy to check that for $p = 0.5$ setting $c$ to zero yields the spatial median and setting it to infinity yields the multivariate sample mean (cf. footnote 2).

As equations (3.1) and (3.2) suggest, the numerical computation of $\theta$ can be achieved efficiently using iterative reweighting. Empirical evidence suggests that equation (3.1) does not have multiple roots: for all tested data sets, the algorithm converged very quickly and is very robust under (even extreme) changes of the starting values. However, the proof of this property remains an open problem.

We return to the example of the last section to demonstrate that the new definition leads to a more favourable outcome. Using the same parameters as before the results are as shown in figure 2. Clearly this is much closer to how we intuitively expect multivariate quantiles to behave than before. The overall shape of the quantile line adapts very well to the data and now all the quantiles lie within the convex hull of the data.

The main difference between the two definitions may best be seen when considering the case $c = 0$ again. In the Breckling/Chambers approach, to obtain the influence function we have to differentiate equation (2.1) with respect to $\theta$ and set this equal to zero, which yields

$$\sum_{i=1}^{n}\Big\{(1-\zeta\cos\alpha_i)\frac{y_i-\theta}{\|y_i-\theta\|}-\zeta\big(r-\frac{y_i-\theta}{\|y_i-\theta\|}\cos\alpha_i\big)\Big\}=0. \tag{3.3}$$

Comparing this to equations (3.1) and (3.2) shows that the first term in (3.3) corresponds exactly to the influence function of the new approach, so the two influence functions differ by the second term in (3.3). Thus it is this term that is responsible for the undesired effect discussed in section 2.2.

# 4 Properties

Clearly an $M$-quantile as defined in equations (3.1) and (3.2) is equivariant under rotation and translation of the data set. Also it is straightforward to see that it is equivariant under re-scaling of the data set as a whole. The $M$-quantiles are not, however, equivariant under arbitrary affine transformations which is a property required by some authors (see, for example, Chakraborty 2000). However, as already pointed out by Averous and Meste (1997), affine equivariance runs counter to certain robustness requirements and it is impossible to fulfil if the standard spatial median is to be included as a special case since the latter is not affine equivariant.

Our definition of a multivariate $M$-quantile implies that it is always located within the convex hull of the data, which can be seen as follows: The weights $w_i$ are strictly positive, so in order to make the sum in (3.1) zero there have to be sample elements on both sides of any hyperplane orthogonal to one of the coordinate axes and passing through $\theta$. Because of the equivariance under rotation property mentioned above, this is also true for an *arbitrary* hyperplane passing through $\theta$. Thus $\theta$ must lie within the convex hull of the data.

In the univariate case, a quantile is described and defined via the geometric property that the proportion of sample elements lying above the $p^{\text{th}}$ quantile is equal to $1-p$. The analogous behaviour in the multivariate setting would be the following: the proportion of sample elements lying beyond a hyperplane orthogonal to $r$ and passing through $\theta$ be equal to $1-p$. This corresponds to first projecting all the sample elements on the straight line given by some pivotal point and the directional vector $r$, and then taking the standard univariate quantile along that direction. A natural choice for the pivotal point would be the spatial median. This approach, however, has two distinct drawbacks. Firstly, by reducing the problem to the univariate case one gives away much of the information implied in the spatial distribution of the data. Secondly, this definition would again lead to quantiles situated outside the convex hull of the data, see figure 3. In turn this means that the definition presented in this paper does not allow this simple geometric interpretation. But can we expect

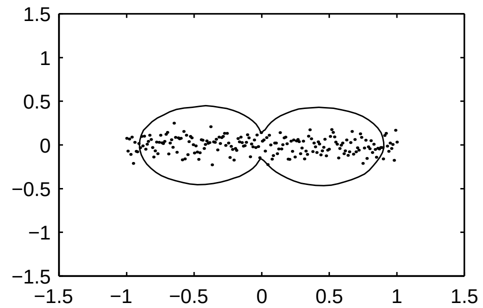


**Figure 3:** Quantiles of cigar shaped sample for $p = 0.05$ ($c = 0$), computed with the "projection" approach.

this kind of "neat" geometric behaviour in the first place? To discuss this point, let us go back to the univariate setting again.

A one-dimensional ($M$-)quantile can be defined in terms of an estimating equation where each residual is assigned a weight depending on its sign. This weighting in principal corresponds to counting the number of residuals left and right of the quantile. When adapting this definition to the multivariate setting the main thing to notice is that distinguishing between positive and negative residuals is equivalent to a specification of direction (one out of two). In dimensions higher than 1, the assignment of weights becomes more complex, because the angle between a chosen directional vector $r$ and a residual $y_i - \theta$ is a continuous variable ranging from $\pi$ to $-\pi$. Thus there arises an ambiguity concerning the definition of the weighting factors, and it becomes clear that the geometric interpretation of a multivariate quantile will be a lot more obscure than in the univariate case and that it will depend on the specific choice of weighting scheme. In Breckling, Kokic, and Lübke (2000) we present a generalisation of our approach that takes this ambiguity into account. A whole family of weighting functions is introduced that include the specification (3.2) as a special case.

We conclude by considering a further fundamental property one might impose on a multivariate $M$-quantile: if $p_1 < p_2$, then the closed surface corresponding to $p_1$ should encompass the surface corresponding to $p_2$, thus implying a probability based ordering. Empirical evidence suggests that this condition is indeed fulfilled by our definition. For examples supporting this statement and a more detailed discussion on the background of the introduced definition we refer the reader to Breckling, Kokic, and Lübke (2000).

# References


Averous, J. and M. Meste (1997). Median Balls: An Extension of the Interquantile Intervals to Multivariate Distributions. *Journal of Multivariate Analysis 63*, 222–241.

Breckling, J. and R. Chambers (1988). *M*-quantiles. *Biometrika 75*(4), 761–771.

Breckling, J., P. Kokic, and O. Lübke (2000). A New Definition of Multivariate *M*-quantiles Based on a Generalisation of the Univariate Estimating Equations. Working paper 1, Insiders Financial Solutions GmbH, Mainz.

Chakraborty, B. (2000). On Affine Equivariant Multivariate Quantiles. *Annals of the Institute of Statistical Mathematics*. To appear.

Chaudhuri, P. (1996). On a Geometric Notion of Quantiles for Multivariate Data. *Journal of the American Statistical Association 91*, 862–872.

He, X. and G. Wang (1997). Convergence of depth contours for multivariate datasets. *Annals of Statistics 52*(2), 495–504.

Kokic, P., R. Chambers, J. Breckling, and S. Beare (1997). A Measure of Production Performance. *Journal of Business & Economic Statistics 15*(4), 445–451.

Rousseeuw, P. J. and M. Hubert (1999). Regression depth. *Journal of the American Statistical Association 94*(446), 388–402.

Scott, D. W. (1992). *Multivariate Density Estimation. Theory, Practice, and Visualization.* New York: John Wiley & Sons, Inc.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Reading, Massachusetts: Addison-Wesley.