A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 5 (307), 2014

Grażyna Trzpiot[*]

SOME PROPERTIES OF SPATIAL QUANTILES

1. INTRODUCTION

Quantiles of univariate data are frequently used to construct popular descriptive statistics. For example, the median is a robust indicator of the central tendency of a population and the interquartile range is good for its dispersion. In addition, quantiles have been used in regression setup (called "*regression quantiles*") (see: Efron 1991; Koenker, Basset 1978) with a univariate response to get robust estimators of parameters in linear models (see: Chaudhuri 1992b; Koenker, Portnoy 1987).

From a practical point of view, quantiles are computed according to an order criterion. Because this order is not total on Rd, an extension of the classical quantile definition in cases when observations are in Rd can only be partial. It acts in this case as a quantile vector (called arithmetic), whose components are the marginal classical quantiles. This definition suffers from several weaknesses. In particular, it is not invariant by rotation and it does not take account of the possible existence of correlations between the different components of the vectors of observations (see: Chakraborty 2001).

In statistical literature, we can find some approaches to define quantiles for multivariate data being proposed. Brown and Hettmansperger (1987, 1989) introduced bivariate quantiles based on the definition of Oja's median (see: Oja 1983). Recently, Donoho and Gasko (1992), Liu, Parelius and Singh (1999) and Zuo and Serfling (2000) defined multivariate quantile using different depth functions; and Abdous and Theodorescu (1992), Chaudhuri (1996) and Koltchinskii (1997) defined them with a class of Mestimates (see: Serfling 1980).

The definition of multivariate quantiles proposed by Chaudhuri (1996) (called geometric) is equivariant under any homogeneous scale transformation of

[*] University of Economics in Katowice, Department of Demography and Economics Statistics.

the co-ordinates of the multivariate observations (Chaudhuri 1996). We will speak about spatial quantiles to refer to this definition.

We have a few steps in this paper. We start from the basic definition of univariate quantiles and in few steps we get to multivariate quantiles. Next, we will go to the sample view and discuss some estimators. The spatial approach of multivariate quantiles called spatial quantiles will be pointed out with a discussion of some estimators. At the end, we will mention the conditional spatial quantiles which can be developed from sample parameter to regression model (Trzpiot 2012).

2. UNIVARIATE TO MULTIVARIATE QUANTILES

2.1. Definition and properties

Let $Y \in R$ be an univariate random variable, and let F be its cumulative distribution function (*CDF*) The quantile function is defined as the inverse of the *CDF*. When F is a monotonically increasing function, its inverse can be defined without ambiguity, but it remains constant on all intervals on which the random variable does not take values. In a general way, the quantile function of Y is noted Q_F (.) and it is defined for $p \in (0,1)$ as (Figure 1):

$$Q_F(p) = F^{-1}(p) = \inf\{y : F(y) \geq p\}. \quad (1)$$

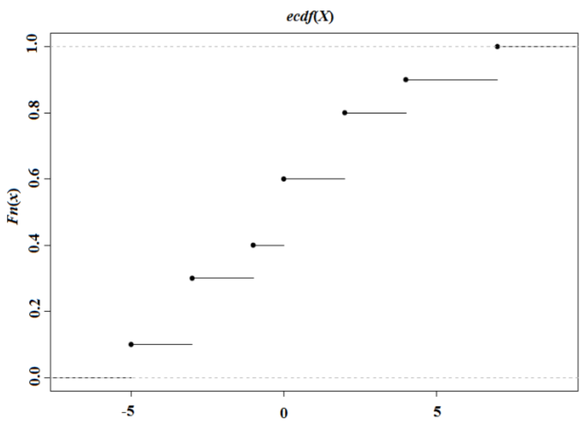

Figure 1. The quantile function

Source: own elaborations.

If Y is continuous, there is a one-to-one relationship between p and $Q_F(p)$. We should not use this co-ordinatewise for multivariate data, since it ignores all dependency patterns and is statistically inferior. Using Ferguson (1967) and Koenker and Basset (1978), the quantile can be defined as the solution of the following minimisation problem. Let $p \in (0,1)$ be a fixed probability. For $t \in R$, let $\phi(2p-1,\ t) = |t| + (2p-1)\,t$, the so-called loss function. The quantile function of Y is noted $Q_M(.)$ and it is defined as:

$$Q_M(p) = \arg\min\ E\{\phi(2p-1,\ Y-\theta)\},\quad \theta \in R. \tag{2}$$

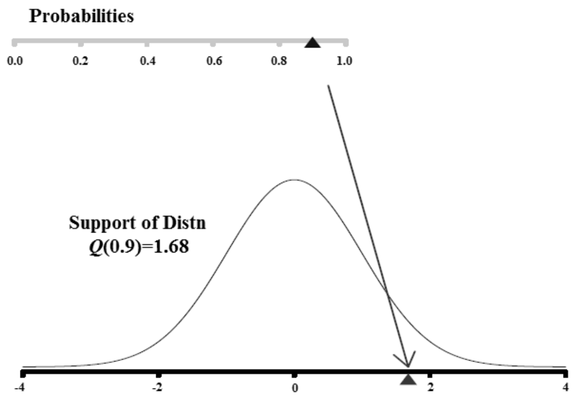

Figure 2. Univariate quantile mapping

Source: S. Chatterjee, *Quantiles and Data Depth: the Next Generation*, School of Statistics, University of Minnesota.

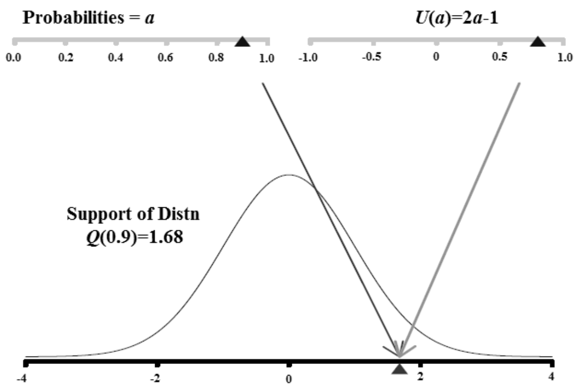

Figure 3. Univariate quantile mapping: formula

Source: S. Chatterjee, *Quantiles and Data Depth: the Next Generation*, School of Statistics, University of Minnesota.

It is easy to check (Figures: 2 and 3) that, for $u = 2p-1$, the quantile $Q_M(p)$ may also be represented as the solution y of the equation $E(S(y-Y)) = u$. That is $Q_M(p)=Q(u)$ with $u = 2p-1$.

For a fixed p, $Q_F(p) = Q_M(p) = Q(u)$ when $u = 2p-1$ (is a bijection). The function $Q^{-1}(.)$ is called the "*centred rank function*". The sign of $u = Q^{-1}(y)$ indicates the position of the point y compared to the median: if u is negative (resp. positive), y is on the left (resp. on the right) of the median. Moreover, using an alternative notation, the "*magnitude*" (for example, the absolute value in the univariate case) of $u = Q^{-1}(y)$ informs us about the order of the quantile: if u is close to -1 (resp. to +1), y is a quantile with order p close to 0 (resp. to 1).

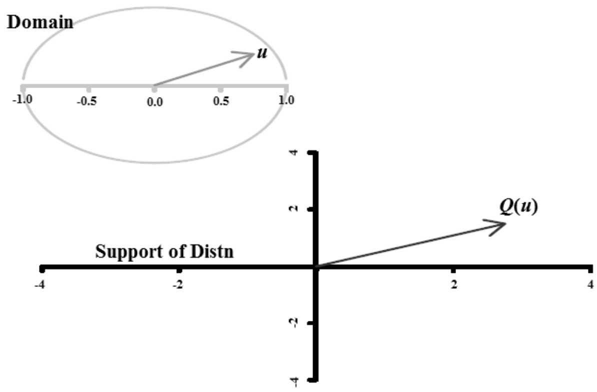

Figure 4. Bivariate quantiles – alternative notation

Source: S. Chatterjee, *Quantiles and Data Depth: the Next Generation*, School of Statistics, University of Minnesota.

We have introduced the characterisation $Q(u)$ for the quantile because it can be generalised in the multivariate framework. In practice, we will use this characterisation to calculate the estimator of the quantile.

2.2. Estimation

Let $Y_1,...,Y_n$ be n observations of y in R. A nonparametric estimator of the *CDF F* is given, for $y \in R$, by:

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i \le y\}} . \tag{3}$$

Thus, for $p \in (0,1)$, we can deduce an estimator $Q_{Fn}(p)$ of $Q_F(p)$ as follows:

$$Q_{F_n}(p) = F_n^{-1}(p) = \inf\left\{y : F_n(y) \geq p\right\}. \tag{4}$$

For $u = 2p-1$, the estimator $Q_n(u)$ of $Q(u)$ can be viewed as the solution y of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} S(y - Y_i) = u. \tag{5}$$

Using the characterisation given by the minimisation approach and for $u = 2p-1$, the quantile $Q_M(p)$ can be estimated by:

$$Q_{M,n}(u) = \arg\min_{\theta \in R} \sum_{i=1}^{n} \varphi(u, Y_i - \theta) = \arg\min_{\theta \in R} \sum_{i=1}^{n} |Y_i - \theta| + u(Y_i - \theta). \tag{6}$$

It is easy to check that, for $u = 2p-1$, the estimator $Q_{M,n}(u)$ of the quantile can be represented as the solution y of the equation (4). Thus, for $u = 2p-1$, these estimators of the quantile are equal: $Q_{Fn}(p) = Q_n(u) = Q_{M,n}(u)$.

3. SPATIAL QUANTILE

3.1. Definition and properties

When the random variable Y is a vector of R^d, the definition of a univariate quantile is not valid because it is based on the idea to order the observations. However, in R^d, the order is not total. From now on, the vectors are considered as a column and the superscript "T" is used to indicate the transpose of vectors or matrices. We suppose that $Y \in R^d$. In the statistical literature, multivariate quantiles have been studied by a certain number of authors, see: for example Abdous and Theodorescu (1992) and Chaudhuri (1996). We choose here to focus on the approach proposed by Chaudhuri.

According to Chaudhuri (1996), the definition of the spatial quantile is a generalisation of the univariate quantile definition introduced by Koenker and Basset (1978).

Suppose $Y \in R^d$ is a random variable. For every $u \in B^p = \{x : \|x\| < 1\}$ the u^{th} quantile $Q(u)$ is define as minimiser of:

$$\Psi_u(q) = E\left[\|Y - q\| + \langle u, Y - q \rangle\right]. \tag{7}$$

So, we consider the multivariate loss function. Now, we define generalised spatial quantiles.

Define $U = u / \|u\|$ for $u \neq 0$. Define $\beta = \|u\|$, thus $u = \beta U$. The projection of X in the direction of u we denote as $X_U U$ where $X_U = \langle X, U \rangle$. The orthogonal projection we denote as $X_{U'} = X - X_U U$.

For every $\lambda \in R$, the generalised spatial quantile minimise:

$$E\left[\|X_U - q_U\|\left[1 + \lambda(X_U - q_U)^{-2}\|X_{U\perp} - q_{U\perp}\|^2\right]^{1/2} + \beta(X_U - q_U)\right]. \tag{8}$$

For $\lambda = 0$ we get the projection quantile. It is computationally simple, has no limitation for sample size and dimension, works for infinite-dimensional observation and has good theoretical properties.

Sample generalised spatial quantiles are consistent and asymptotically Gaussian with an intractable dispersion parameter. The generalised bootstrap can be used for inference and obtaining all the statistical properties of these quantiles. Projection quantiles have a one-to-one relationship like univariate quantiles. Projection quantiles based confidence sets have exact coverage.

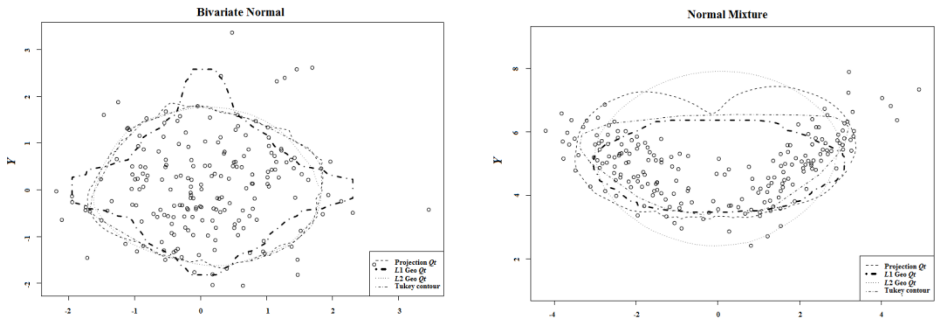

Figure 5. Example scatter plot

Source: S. Chatterjee, *Simultaneous Quantiles of Several Variables*, School of Statistics, University of Minnesota.

3.2. Estimation of spatial quantiles

Let F_n be an empirical nonparametric estimator of F obtained from the observations $Y_1,...,Y_n$ of $Y \in R^d$. We can define the estimator $Q_n(.)$ of thespatial quantile $Q(.)$ for all $\mathbf{u} \in B^d$, by:

$$Q_n(\mathbf{u}) = \arg\min_{\theta \in R^d} \sum_{i=1}^{n} \left(\varphi(\mathbf{u}, Y_i - \theta) - \varphi(\mathbf{u}, Y_i)\right). \tag{9}$$

The vector **u** gives us information about the estimator of the quantile $Q_n(u)$. To determine the order of the spatial quantile, we just have to calculate the norm of **u**: if $\|\mathbf{u}\| \approx 1$ (resp. 0), then $Q_n(\mathbf{u})$ is an extreme quantile (resp. central quantile, i.e.: close to the spatial median).

Because **u** is a vector of B^d, its direction indicates the position of the spatial quantile compared to the spatial median.

From the previous characterisations, it can be checked that, for $\mathbf{u} \in B^d$, the estimator $Q_n(\mathbf{u})$ of the spatial quantile $Q_n(\mathbf{u})$ can be seen as the solution y of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} S(y - Y_i) = \mathbf{u}. \tag{10}$$

The term $\|\mathbf{u}\|$ said "*extent of deviation*" must not be considered as the Euclidean distance between $Q(\mathbf{u})$ and the spatial median $\mathbf{M} = Q(0)$. Moreover, the distance between $Q(\mathbf{u})$ and M does not increase with $\|\mathbf{u}\|$.

Contrary to the univariate case where $u = 2p-1$, the "*magnitude*" $\|\mathbf{u}\|$ does not carry any probabilistic interpretation where $d \geq 2$. In particular, let us consider the region $\{Q_n(u) : \|\mathbf{u}\| \leq 0.5\}$. In the univariate case, it corresponds to the interquartile region with $\frac{1}{4} \leq p \leq \frac{3}{4}$. In the multivariate case, this region does not necessarily contain 50% of observations.

4. CONDITIONAL SPATIAL QUANTILE

4.1. Definition

Having a sample of observations $\{(X_1, Y_1), ..., (X_n, Y_n)\}$ from a vector (X, Y) with values in $R^s \times R^d$, we are interested in studying the relationship between X and Y. The conditional quantiles represent a mean to approach this problem.

In the univariate case (i.e.: $Y \in R$), when the functional form between X and Y is unknown, there is a large variety of methods allowing the estimation of the conditional quantiles. For example, we can quote the kernel estimation, the local constant kernel estimation and the double kernel estimation (see: Gannoun et al. (2002) for a description of these methods). On the other hand, few authors are interested in the conditional spatial quantile and their properties. Recently De Gooijer et al. (2006) have introduced the conditional spatial quantile based on the minimisation of the pseudonorm given by Abdous and Theodorescu (1992).

We present here an alternative formalisation of the conditional spatial quantile based on a generalisation of the notion of spatial quantile studied by Chaudhuri (1996). Chaudhuri indexes the spatial quantile by a vector **u** in B^d, which allows us to obtain not only the idea about the "*extreme*" and "*central*" observations, but also about their position in multivariate scatterplots.

We define the conditional spatial quantile of the variable **Y** given **X**=**x** as:

$$Q(\mathbf{u} \mid \mathbf{x}) = \arg\min_{\theta \in R^d} \int_{R^d} \{\varphi(\mathbf{u}, y - \theta) - \varphi(\mathbf{u}, y)\} F(dy \mid \mathbf{x}) . \tag{11}$$

Moreover, as in the previous section, the conditional spatial quantile can be seen as the solution y of the following equation:

$$E(S(y - \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) = \mathbf{u} . \tag{12}$$

4.2. Estimation of conditional spatial quantile

Let $F_n(.|\mathbf{x})$ be the nonparametric (Nadaraya-Watson) estimator of the conditional distribution function of **Y** given **X**=**x**, defined, for all $y \in R^d$, as:

$$F_n\left(y\middle|\mathbf{x}\right)=\frac{1}{n}\sum_{i=1}^{n}w_{n,i}1_{\{Y_i\le y\}}\ . \tag{13}$$

where $w_{n,i}=\dfrac{k\left(\left(x-X_i\right)/h_n\right)}{\sum_{i=1}^{n}k\left(\left(x-X_i\right)/h_n\right)}$ is a weight associated to Y_i, the kernel function, k is a density function and h_n (the window) is a real positive sequence such that $h_n\to 0$ as $n\to\infty$.

We can deduce using Equation (10), an estimator $Q_n(\mathbf{u}|\mathbf{x})$ of the conditional spatial quantile $Q(\mathbf{u}|\mathbf{x})$ as:

$$\begin{aligned}Q(\mathbf{u}|\mathbf{x})&=\arg\min_{\theta\in R^d}\int_{R^d}\{\varphi(\mathbf{u},y-\theta)-\varphi(\mathbf{u},y)\}F(dy|\mathbf{x})\\&=\arg\min_{\theta\in R^d}\sum_{i=1}^{n}w_{n,i}\{\varphi(\mathbf{u},y-\theta)-\varphi(\mathbf{u},y)\}\end{aligned}\ . \tag{14}$$

The estimator $Q_n(\mathbf{u}|\mathbf{x})$ of the quantile $Q(\mathbf{u}|\mathbf{x})$ can be viewed as the solution y of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}S\left(y-Y_i\right)w_{n,i}=\mathbf{u}\ . \tag{15}$$

5. CONCLUDING REMARKS

A leading multivariate extension of the univariate quantiles is the so-called "*spatial*" or "*geometric*" notion, for which sample versions are highly robust and conveniently satisfy a Bahadur–Kiefer representation.

New statistics based on spatial quantiles are presented for nonparametric estimations of multiple regression coefficients and for robust estimations of multivariate dispersion.

The important way to apply quantiles instead of different dispersion measure is the regression model.

Quantile regression is much better suited to analysing questions involving changes in the distribution of a dependent variable. Quantile regressions allow for separate effects of an explanatory variable on different points of the dependent variable distribution. Coefficient estimates are then frequently interpreted as being analogous to standard linear regression estimates, albeit for different points in the distribution of the dependent variable (Trzpiot 2008; 2009a,b,c; 2010; 2011a,b; 2012; 2013).

REFERENCES

Abdous B., Theodorescu R. (1992), *Note on the spatial quantile of a random vector*, "Statistics and Probability Letter", 13, pp. 333-336.

Barnett V. (1976), *The ordering of multivariate data (with comments)*, "Journal of Royal Statistical Society", Ser. A, 139, pp. 318-354.

Chakraborty B. (2001), *On affine equivariant multivariate quantiles*, T"he Institute of Statistical Mathematics", 53, pp. 380-403.

Chaudhuri P. (1992a), *Multivariate location estimation using extension of R-estimates through U-statistics type approach*, "Annals of Statistics", 20, pp. 897-916.

Chaudhuri P. (1992b), *Generalized regression quantiles: Forming a useful toolkit for robust linear regression*, (in:) Dodge Y. (ed.), *L1 Statistical Analysis and Related Methods*, Amsterdam: North-Holland, pp. 169-185.

Chaudhuri P. (1996), *On a geometric notation of quantiles for multivariate data*, "Journal of the American Statistical Association", 91, pp. 862-872.

Chaouch M., Gannoun A., Saracco J. (2008), *Conditional Spatial Quantile: Characterization and Nonparametric Estimation*, Cahier Du Gretha – 10.

Dabo-Niang S., Thiam (2010), *Robust quantile estimation and prediction for spatial processes*, "Statistics and Probability Letters", 80, pp. 1447-1458

Eddy W. F. (1985), *Ordering of Multivariate Data*, (in:) Billard L. (ed.), *Computer Science and Statistics:*
The Interface, Amesterdam: North-Holland, pp. 25-30.

Efron B. (1991), *Regression percentiles using asymmetric squared error loss*, "Statistica Sinica", 1,
pp. 93-125.

Ferguson T. (1967), *Mathematical Statistics: A Decision Theory Approach*, Academic Press: New York.

Koenker R., Basset G. (1978), *Regression Quantiles*, "Econometrica", 46, pp. 33-50.

Koenker R., Portnoy S. (1987), *L Estimation for linear models*, "Journal of the American statistical Association", 82, pp. 851-857.

Oja H. (1983), *Descriptive Statistics for Multivariate Trimming*, "Statistics and Probability Letters", 1, pp. 327-332.

Plackett R. L. (1976), *Comment on Ordering of multivariate data by V. Barnett*, "Journal of the Royal Statistical Society", Ser. A, 139, pp. 344-346.

Reiss R. D. (1989), *Approximation distributions of order statistics with applications to nonparametric statistics*, New York: Springer.

Serfling R. (1980), *Approximation theorem of mathematical statistics*, New York: John Wiley.

Serfling R. (2002), *Quantile functions for multivariate analysis: approaches and applications*, "Annals of Statistics", 25, pp. 435-477.

Trzpiot G. (2008), *The Implementation of Quantile Regression Methodology in VaR Estimation*, "Studies and Researches of Faculty of Economics and Management University of Szczecin".

Trzpiot G. (2009a), *Quantile Regression Model versus Factor Model Estimation*, "Financial Investments and Insurances", Vol 60.

Trzpiot G. (2009b), *Application weighted VaR in capital allocation*, "Polish Journal of Environmental Studies", Vol 18, 5B.

Trzpiot G. (2009c), *Estimation methods for quantile regression*, "Economics Studies", 53.
Trzpiot G. (2010), *Quantile Regression Model of Return Rate Relation – Volatility for Some Warsaw Stock Exchange Indexes*, "Finances, Financial Markets and Insurances. Capital Market", Vol 28, pp. 61-76.
Trzpiot G. (2011a), *Bayesian Quantile Regression*, „Studia Ekonomiczne", Zeszyty Naukowe nr 65, pp. 33-44.
Trzpiot G. (2011b), *Some tests for quantile regression models*, "Acta Universitatis Lodziensis Folia Economica", 255, pp. 125-135.
Trzpiot G. (2012), *Spatial quantile regression*, "Comparative Economic Research. Central and Eastern Europe", vol. 15, no 4, pp. 265-279.
Trzpiot G. (2013), *Properties of transformation quantile regression model*, "Acta Universitatis Lodziensis Folia Economica", 285, pp. 125-137
Zuo Y., Serfling R. (2000), *General notions of statistical depth function*, "Annals of Statistics", 28,
pp. 461-482.

ABSTRACT

Conditional quantiles are required in various economic, biomedical or industrial problems. Lack of objective basis for ordering multivariate observations is a major problem in extending the notion of quantiles or conditional quantiles (also called regression quantiles) in a multidimensional setting. We present characterisations of the spatial quantiles and the corresponding estimators. Nonparametric inference is very naturally quantile-based, and in recent years various notions of multivariate quantiles the spatial quantile function for whose sample version have been recalled.

WYBRANE WŁASNOŚCI PRZESTRZENNYCH KWANTYLI

ABSTRAKT

Warunkowe kwantyle są wykorzystywane w ekonomii, biomedycynie lub w przemyśle. Mamy problemy z wprowadzeniem relacji porządku w obserwacjach wielowymiarowych, co przenosi się również na uogólnienie definicji kwantyli oraz warunkowych kwantyli (regresji kwantylowej) w przestrzeni wielowymiarowej. Omówimy własności przestrzennych kwantyli oraz ich estymatory. Wnioskowanie nieparamertyczne jest wykorzystywane przy opisie kwantylowym. Przedstawimy różne notacje wielowymiarowych kwantyli oraz przestrzennych funkcji kwantylowych w zapisie dla próby badawczej.