Quantity Quantiles Linear Regression*

Paolo Radaelli Michele Zenga

Dipartimento di Metodi Quantitativi

per le Scienze Economiche ed Aziendali

Università degli Studi di Milano-Bicocca

Abstract

We show that the definition of the θth sample quantile as the solution to a minimization problem introduced by Koenker and Basset [1978] can be easily extended to obtain an analogous definition for the θth sample quantity quantile instead of the usual one. By means of this definition we introduce a linear regression model for quantity quantiles and analyze some properties of the residuals.

In section 4 we show a brief application of the methodology proposed.

Key words: quantile regression, quantity quantiles.

1 Introduction

The θth quantile ζ_θ of the real valued random variable (r.v.) Y with distribution function:

$$F_Y(y) = P(Y \leqslant y) \tag{1.1}$$

can be obtained:

*A short version of this paper has been presented at XLIII Scientific Meeting of the Società Italiana di Statistica (Radaelli and Zenga [2006]).

- by its quantile function:

$$B_Y(\theta) = \inf\{y \in \Re \mid F_Y(y) \geqslant \theta\} \qquad 0 < \theta < 1. \tag{1.2}$$

  for a fixed $\theta \in (0, 1)$;

- as the solution of a minimization problem which is the crucial feature in quantile regression.

For the latter, consider the problem of approximating the distribution of Y with a value $c \in \Re$. Given a loss function ℓ the optimal value c can be obtained by minimizing the expected loss:

$$r(c) = \mathbb{E}\left[\ell(Y - c)\right] = \mathbb{E}\left[\ell(U)\right] \tag{1.3}$$

which depends both on the distribution of Y and on the particular loss function adopted (see Peracchi [2001]). In the case the loss function specified is the quadratic one:

$$\ell(u) = u^2$$

the value of c that minimizes (1.3) is the mean value $\mu = \mathbb{E}(Y)$, while applying the absolute loss function:

$$\ell(u) = |u|$$

the value of c that minimizes (1.3) is the median $\zeta_{0.5}$ of Y.
The loss function applied in quantile regression, as introduced by Koenker and Basset [1978] (see also Koenker [2005]), is the *asymmetric absolute loss function*:

$$\ell_\theta(u) = \left[\theta\ I\{u > 0\} + (1 - \theta)\ I\{u \leqslant 0\}\right]|u| = \left[\theta - I\{u \leqslant 0\}\right]u \tag{1.4}$$

where $0 < \theta < 1$ and $I(A)$ denoting the indicator function of the event A. With the loss function (1.4) the expected loss (1.3) is:

$$\mathbb{E}\left[\ell_\theta(Y - c)\right] = (\theta - 1)\int_{-\infty}^{c}(y - c)\ dF_Y(y) + \theta\int_{c}^{\infty}(y - c)\ dF_Y(y) \tag{1.5}$$

which is minimized when c is the θth quantile ζ_θ of Y.
In this framework the θth quantile of Y can be defined as any solution to the minimization problem:

$$\min_{c \in \Re} \mathbb{E}\left[\ell_\theta\left(Y - c\right)\right]. \tag{1.6}$$

Considering now a random sample $Y_1, \ldots, Y_i, \ldots, Y_n$ from Y, the natural estimator for ζ_θ is the corresponding sample quantile $\widehat{\zeta_\theta}$ that can be definied as follows:

$$\sum_{i=1}^{n} I\left\{y_i < \widehat{\zeta_\theta}\right\} \leqslant n\theta \leqslant \sum_{i=1}^{n} I\left\{y_i \leqslant \widehat{\zeta_\theta}\right\}. \tag{1.7}$$

Thus, once the observations obtained from the sample are arranged in nondecreasing order:

$$y_{(1)} \leqslant y_{(2)} \leqslant \ldots \leqslant y_{(n)},$$

the θth sample quantile is unique, provided that $n\theta$ is not integer, and it is the value that fills the position $\lceil n\theta \rceil$[1]:

$$\widehat{\zeta_\theta} = y_{(\lceil n\theta \rceil)}.$$

In the case $n\theta$ is integer, the θth sample quantile is any value in the closed interval $\left[y_{(n\theta)}; y_{(n\theta+1)}\right]$.
Koenker and Basset [1978] have shown that, applying the asymmetric absolute loss function (1.4), the θth sample quantile can be obtained simply by replacing the distribution function F in (1.5) with the empirical distribution function:

$$F_n\left(y\right) = n^{-1} \sum_{i=1}^{n} I\left\{y_i \leqslant y\right\} \tag{1.8}$$

obtaining the minimization problem:

$$\min_{\zeta_\theta \in \Re} \frac{1}{n} \sum_{i=1}^{n} \ell_\theta\left(y_i - \zeta_\theta\right)$$
$$\equiv \min_{\zeta_\theta \in \mathbb{R}} \frac{1}{n}\left[\sum_{[i:y_i > \zeta_\theta]} \theta \left(y_i - \zeta_\theta\right) + \sum_{[i:y_i \leqslant \zeta_\theta]} \left(\theta - 1\right) \left(y_i - \zeta_\theta\right)\right]. \tag{1.9}$$

[1] $\lceil \cdot \rceil$ denotes the rounding up.

The point of departure of quantile regression is thus the definition of the θth sample quantile as the solution of a minimization problem instead of the usual procedure that implies the ordering of the sample observations. The optimization approach yields a natural generalization of the quantiles to the regression settings.
As pointed out in Hallock and Koenker [2001] the solution to problem (1.9) is an estimate of the unconditional θth quantile of Y. Suppose now to have p explanatory variables $X_1, \ldots, X_p$ and that the conditional quantile function of Y is linear and given by:

$$B_Y(\theta|\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_\theta \qquad 0 < \theta < 1. \tag{1.10}$$

In the quantile regression (linear) model the unknown parameters $\boldsymbol{\beta}_\theta$ in the conditional quantile function are estimated simply by replacing the scalar ζ_θ in (1.9) by the function $\mathbf{x}^T\boldsymbol{\beta}_\theta$. Thus the θth regression quantile is defined as any solution to the minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\beta}_\theta \in \Re^p} \frac{1}{n}\sum_{i=1}^{n} \ell_\theta\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_\theta\right) \\ &\equiv \min_{\boldsymbol{\beta}_\theta \in \Re^p} \quad \frac{1}{n}\left[\sum_{\left[i:y_i > \mathbf{x}_i^T\boldsymbol{\beta}_\theta\right]} \theta\ \left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_\theta\right) + \sum_{\left[i:y_i \leqslant \mathbf{x}_i^T\boldsymbol{\beta}_\theta\right]} (\theta - 1)\ \left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_\theta\right)\right] \end{aligned} \tag{1.11}$$

where $\mathbf{y} = [y_1, \ldots, y_n]$ is a vector of responses on the random variable Y and $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$ is the known $n \times p$ matrix of the regressors.
The case of the conditional median is obtained setting $\theta = 0.5$ in (1.11) and has been first introduced by Boscovich (see for example Hald [1998], Mineo [2003] and Stigler [1984]) and then investigated by De La Place [1966], Edgeworth (see [Bowley, 1972, pp.99-109]) and Bassett and Koenker [1978].
In the literature other asymmetrically weighted loss functions have been introduced: Newey and Powell [1987] proposed the use of the *asymmetric least squares loss function*:

$$\rho_\theta(u) = |\theta - I\{u \leqslant 0\}|\ \ u^2 \tag{1.12}$$

leading to the minimum problem:

$$\min_{\nu_\theta \in \Re} \frac{1}{n} \sum_{i=1}^{n} \rho_\theta \left(y_i - \nu_\theta \right)$$

$$\equiv \min_{\nu_\theta \in \Re} \quad \frac{1}{n} \left[\sum_{[i: y_i > \nu_\theta]} \theta \ \left(y_i - \nu_\theta \right)^2 + \sum_{[i: y_i \leqslant \nu_\theta]} \left(1 - \theta \right) \ \left(y_i - \nu_\theta \right)^2 \right]. \tag{1.13}$$

The solution to problem (1.13) has been defined by the authors the θth sample expectile[2] and the use of the asymmetric loss function (1.12) is then extended to the regression context leading to asymmetric least squares estimators (see also Efron [1991]).

2 Quantity quantiles

When one face with a nonnegative random variable, the income for instance, it is useful to consider not only the quantiles as defined in the previous section but the quantity quantiles as well.

From now on Y is supposed to be a nonnegative r.v. with density function f_Y and with finite and strictly positive mean value:

$$\mathbb{E}\left[Y \right] = \int_0^\infty y \ f_Y(y) \ dy = \mu.$$

The distribution function of Y is:

$$F_Y \left(y \right) = \int_0^y f_Y(t) \ dt \tag{2.1}$$

and the so-called *first incomplete moment* (or first-moment distribution) is given by:

$$Q_Y \left(y \right) = \int_0^y \frac{t}{\mu} \ f_Y(t) \ dt. \tag{2.2}$$

Thus, if Y represent the income, $F_Y \left(y \right)$ gives the fraction of the population with income no greater than y while $Q_Y \left(y \right)$ gives the share of the total income accruing

[2] The 0.5 expectile is the sample mean $\overline{y}$.

to the population with income no greater than y.
In this framework it is possibile to define the θth quantile:

$$\zeta_\theta = \inf\{y \in \Re \mid F_Y(y) \geqslant \theta\} \qquad 0 < \theta < 1 \tag{2.3}$$

and the θth *quantity quantile*:

$$\eta_\theta = \inf\{y \in \Re \mid Q_Y(y) \geqslant \theta\} \qquad 0 < \theta < 1. \tag{2.4}$$

It should be observed (e.g. Zenga [1984]) that:

$$F_Y(y) \geqslant Q_Y(y) \qquad \forall\, y > 0$$

and thus:

$$\zeta_\theta \leqslant \eta_\theta \qquad \forall\, 0 < \theta < 1.$$

This approach has been widely used in the study of income and wealth distribution and many concentration measures have been derived from the comparison of (2.1) and (2.2), see Zenga [1990] for a short review, and from the comparison of (2.3) and (2.4), see Zenga [1984], Zenga [1985], Zenga [1990], Berti and Rigo [1999] and Kleiber and Kotz [2003, pp.42-43].
It is interesting to observe that $\eta_{0.5}$ is the so called *dividing value* (e.g. Mortara [1933, pp.70-71]) defined as the value that balances the sum of the values lower than it and the sum of the values greater than it.
As the quantile can be defined as any solution to the minimization problem that makes use of the asymmetric absolute loss function (1.4), so the quantity quantile η_θ can be defined as any solution to a minimization problem that attaches a suitable weight to the loss function (1.4).
In particular let W be a r.v. with the same support as Y and density function:

$$q_W(y) = f_Y(y)\frac{y}{\mu}. \tag{2.5}$$

The distribution function of W is given by:

$$F_W(y) = \int_0^y q_W(t)\, dt = \int_0^y \frac{t}{\mu}\, f_Y(t)\, dt = Q_Y(y) \tag{2.6}$$

which is the first incomplete moment (2.2) of Y, therefore the θ-th quantile of the r.v. W is the θ-th quantity quantile of the r.v. Y.
Thus as Jones [1994] showed that expectiles of a given distribution F are precisely the ordinary quantiles of a distribution G related by an explicit formula to F, then we observe that quantity quantiles of a given distribution F are the ordinary quantiles of the distribution G obtained from F by applying the Lorenz function:

$$L(p) = \frac{1}{\mathbb{E}[Y]} \int_0^p F^{-1}(t)\, dt \qquad p \in [0,1]$$

to the distribution function F (see for example Berti and Rigo [1999]):

$$Q_Y(y) = L(F_Y(y)).$$

It is now straightforward to define the quantiles of the r.v. W in the same way as we showed in the previous section. Thus, according to (1.6), the θth quantile of W can be defined as the solution to the minimization problem:

$$\min_{c \in \Re} \mathbb{E}[\ell_\theta(W-c)]. \tag{2.7}$$

The expected loss in (2.7) can be rewritten as follows:

$$\begin{aligned}\mathbb{E}[\ell_\theta(W-c)] &= \int_{-\infty}^{\infty} \ell_\theta(w-c)\, f_W(w)\, dw = \int_{-\infty}^{\infty} \ell_\theta(y-c)\, \frac{y}{\mu}\, f_Y(y)\, dy \\ &= \mathbb{E}\left[\ell_\theta(Y-c)\, \frac{Y}{\mu}\right]\end{aligned}$$

i.e. the θ-th quantity quantile of the r.v. Y can be obtained by minimizing the expect loss in which the distances $|y-c|$ are weighted not only according to the asymmetric absolute loss function (1.4) but with the additional nonnegative weight y/μ.
For $\theta = 1/2$, (2.7) gives the dividing value and for $\theta \in (0,1)$ all the others quantity quantiles of the r.v. Y are obtained. Thus we observe that the relationship between quantity quantiles and the dividing value is that between the quantiles and the median and between the expectiles and the mean.
Consider now n sorted observations, obtained from a random sample $Y_1, \ldots, Y_i, \ldots, Y_n$ from Y:

$$0 \leqslant y_{(1)} \leqslant y_{(2)} \leqslant \ldots y_{(i)} \leqslant \ldots \leqslant y_{(n)} > 0$$

and let $T = \sum_{i=1}^{n} y_i > 0$ and $\overline{y}$ denote, respectively, the total amount and the arithmetic mean of the observations.

The natural estimator for η_θ is the corresponding sample quantity quantile:

$$\widehat{\eta}_\theta = \inf \left\{ y_{(i)} \;:\; \widehat{Q}(y_{(i)}) \geqslant \theta \right\}$$

where:

$$\widehat{Q}(b) = \frac{\displaystyle\sum_{j:\ y_j \leqslant b} y_j}{n\,\overline{y}} = \frac{\displaystyle\sum_{j:\ y_j \leqslant b} y_j}{T}.$$

Thus traditionally, in order to obtain the θ-th sample quantity quantile, one should sum the sorted values $y_{(i)}$ until at least the share θ of the total T is reached. Nevertheless by replacing the empirical distribution function (1.8) of Y with:

$$G_n(y) = n^{-1} \sum_{i=1}^{n} I\{y_i \leqslant y\} \frac{y_i}{\overline{y}} = \widehat{Q}(y)$$

the θ-th sample quantity quantile can be obtained as any solution to the minimization problem:

$$\begin{aligned} &\min_{b \in \Re} \sum_{i=1}^{n} \ell_\theta(y_i - b)\ \frac{y_i}{n\overline{y}} \\ &\equiv \min_{b \in \Re} \quad \left[\sum_{[i:y_i > b]} \theta\ (y_i - b) \frac{y_i}{n\overline{y}} + \sum_{[i:y_i \leqslant b]} (\theta - 1)\ (y_i - b) \frac{y_i}{n\overline{y}} \right]. \end{aligned} \tag{2.8}$$

Denoting with $D(b)$ the expression within square brackets in (2.8), it can be easily shown that its minimum is reached for $b = \widehat{\eta}_\theta$. In fact $D(b)$ can be rewritten as:

$$D(b) = \theta \sum_{[i:y_i > b]} \frac{y_i^2}{n\overline{y}} + (\theta - 1) \sum_{[i:y_i \leqslant b]} \frac{y_i^2}{n\overline{y}} + b\ \left[\widehat{Q}(b) - \theta\right]$$

In order to evaluate $D(b)$ when b varies we consider:

$$\begin{aligned} D(y_{(i)}) &= \theta \sum_{j=i+1}^{n} \frac{y_{(j)}^2}{n\overline{y}} + (\theta - 1) \sum_{j=1}^{i} \frac{y_{(j)}^2}{n\overline{y}} + y_{(i)}\ \left[\widehat{Q}(y_{(i)}) - \theta\right] \\ D(y_{(i+1)}) &= \theta \sum_{j=i+2}^{n} \frac{y_{(j)}^2}{n\overline{y}} + (\theta - 1) \sum_{j=1}^{i+1} \frac{y_{(j)}^2}{n\overline{y}} + y_{(i+1)}\ \left[\widehat{Q}(y_{(i+1)}) - \theta\right] \end{aligned}$$

and their difference:

$$D(y_{(i+1)}) - D(y_{(i)}) = \left[y_{(i+1)} - y_{(i)}\right] \left[\widehat{Q}(y_{(i)}) - \theta\right].$$

The difference $\left[y_{(i+1)} - y_{(i)}\right]$ is nonnegative, therefore:

$$D(y_{(i+1)}) - D(y_{(i)}) \begin{cases} < 0 & \text{if } \widehat{Q}(y_{(i)}) < \theta; \\ > 0 & \text{if } \widehat{Q}(y_{(i)}) > \theta. \end{cases}$$

Thus increasing b, the value of $D(b)$ decreases as long as $\widehat{Q}(b) < \theta$; the minimum is reached for b that satisfies $\widehat{Q}(b) = \theta$, i.e. $b = \widehat{\eta}_\theta$.

3 Linear regression for quantity quantiles

In section 1 we showed that the optimization problem (1.11) allows to obtain the regression coefficients $\boldsymbol{\beta}_\theta$ of the hyperplane for the θ-th conditional quantile of the dependent variable Y. Suppose now we are interested in a linear model for the θth conditional quantile of the r.v. W with density function (2.5) and distribution function (2.6). In particular, following the same setting, suppose to have p explanatory variables $X_1, \ldots, X_p$ and that the conditional quantile function of W is given by:

$$B_W(\theta|\mathbf{x}) = \mathbf{x}^T \boldsymbol{\gamma}_\theta \qquad 0 < \theta < 1. \tag{3.1}$$

the unknown parameters $\boldsymbol{\gamma}_\theta$ in the conditional quantile function (3.1) can be obtained simply by replacing b in (2.8) by the function $\mathbf{x}^T \boldsymbol{\gamma}_\theta$. Thus the θth regression quantity quantiles for the variable Y, i.e. the θth regression quantile for the variable W, is defined as any solution to the minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\gamma}_\theta \in \Re^p} \frac{1}{n} \sum_{i=1}^{n} \ell_\theta \left(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta\right) \frac{y_i}{\overline{y}} \\ &\equiv \min_{\boldsymbol{\gamma}_\theta \in \Re^p} \frac{1}{n} \left[\sum_{\left[i: y_i > \mathbf{x}_i^T \boldsymbol{\gamma}_\theta\right]} \theta \, \left(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta\right) \frac{y_i}{\overline{y}} + \sum_{\left[i: y_i \leqslant \mathbf{x}_i^T \boldsymbol{\gamma}_\theta\right]} (\theta - 1) \left(y_i - \mathbf{x}_i^T \boldsymbol{\gamma}_\theta\right) \frac{y_i}{\overline{y}} \right] \end{aligned} \tag{3.2}$$

where $\mathbf{y} = [y_1, \ldots, y_n]$ is a vector of responses on the random variable Y and $\mathbf{x}_i^T$ is ith row of the known $n \times p$ matrix of the regressors X.
Problem (3.2) can be viewed in the context of weighted linear quantile regression introduced by Koenker and Zhao [1994] for linear heteroscedastic model. In the case here considered we don't consider that the conditional densities of the response Y

are heterogeneous; we just use the additional nonnegative weights $y_i/\overline{y}$ in the minimization problem in order to obtain the quantity quantiles instead of the quantiles. Problem (3.2) can be reformulated, as in quantile regression (see for example Koenker and Basset [1978]), as the linear program:

$$\min_{(\boldsymbol{\gamma}_\theta, \mathbf{u}^+, \mathbf{u}^-) \in \Re^p \times \Re^{2n}} \theta \, \boldsymbol{\iota}^T \mathbf{u}^+ \frac{\mathbf{y}}{\overline{y}} \; + \; (1-\theta) \, \boldsymbol{\iota}^T \mathbf{u}^- \frac{\mathbf{y}}{\overline{y}} \tag{3.3}$$

subject to the constraints:

$$\begin{aligned} \mathbf{y} = X \, \boldsymbol{\gamma}_\theta + \mathbf{u}^+ - \mathbf{u}^- \\ \left(\mathbf{u}^+, \mathbf{u}^-\right) \; \in \Re^{2n}_+ \end{aligned} \tag{3.4}$$

where the vector of the residuals $\mathbf{u} = \mathbf{y} - X \, \boldsymbol{\gamma}_\theta$ is split into its positive $\mathbf{u}^+$ and negative $\mathbf{u}^-$ parts such that $\mathbf{u} = \mathbf{u}^+ - \mathbf{u}^-$.

In theorem 3.4 of Koenker and Basset [1978] the authors prove that, given a solution $\widehat{\boldsymbol{\beta}}_\theta$ to problem (1.11), if the matrix X contains an intercept (i.e. X contains a column of ones) then there will be roughly a proportion θ of negative residuals and a proportion $1-\theta$ of positive residuals; more precisely the following inequality holds:

$$\frac{n_-}{n} \leqslant \theta \leqslant 1 - \frac{n_+}{n} = \frac{n_- + n_0}{n} \tag{3.5}$$

where n_-, n_+ and n_0 denote the number of negative, positive and zero elements of the residual vector $\mathbf{u} = \mathbf{y} - X \, \widehat{\boldsymbol{\beta}}_\theta$.
An analogous theorem holds for quantity quantile regression: in particular, given a solution $\widehat{\boldsymbol{\gamma}}_\theta$ to problem (3.2) and the correspondent vector of residuals $\mathbf{u} = \mathbf{y} - X \, \widehat{\boldsymbol{\gamma}}_\theta$, consider the following partition of the set $\{1, \dots, n\}$:

- $Z = \{i \; : \; u_i = 0\}$ indexes of points with zero residual;
- $N = \{i \; : \; u_i < 0\}$ indexes of points with negative residual;
- $P = \{i \; : \; u_i > 0\}$ indexes of points with positive residual

then:

Theorem 3.1 *Let $T = n\overline{y} > 0$ denotes the total amount of the observations on Y. If X contains an intercept then:*

$$\sum_{i\in N}\frac{y_i}{T} \leqslant \theta \leqslant 1-\sum_{i\in P}\frac{y_i}{T} = \sum_{i\in N}\frac{y_i}{T}+\sum_{i\in Z}\frac{y_i}{T}. \tag{3.6}$$

Proof. *Rewrite (3.6) as:*

$$(\theta-1)\sum_{i\in Z}\frac{y_i}{T} \leqslant (1-\theta)\sum_{i\in N}\frac{y_i}{T}-\theta\sum_{i\in P}\frac{y_i}{T} \leqslant \theta\sum_{i\in Z}\frac{y_i}{T}. \tag{3.7}$$

and let:

$$\delta = \min_{i\in(P\cup N)}|u_i|$$

and:

$$0 < \varepsilon < \delta. \tag{3.8}$$

The difference:

$$(1-\theta)\sum_{i\in N}\frac{y_i}{T}-\theta\sum_{i\in P}\frac{y_i}{T}$$

can be positive, negative or null; then we consider separately two cases. Consider first the case:

$$(1-\theta)\sum_{i\in N}\frac{y_i}{T} \leqslant \theta\sum_{i\in P}\frac{y_i}{T}.$$

The term $\theta\sum_{i\in Z}\frac{y_i}{T}$ in (3.7), is nonnegative, so it must only be proved that:

$$(\theta-1)\sum_{i\in Z}\frac{y_i}{T} \leqslant (1-\theta)\sum_{i\in N}\frac{y_i}{T}-\theta\sum_{i\in P}\frac{y_i}{T}.$$

Let $\boldsymbol{\eta}_\theta = [\varepsilon\ 0\ \dots\ 0]^T + \widehat{\boldsymbol{\gamma}}_\theta$ and denote with $u_i(\widehat{\boldsymbol{\gamma}}_\theta)$ and $u_i(\boldsymbol{\eta}_\theta)$ the residuals of the n points from the hyperplanes with coefficients, respectively, $\widehat{\boldsymbol{\gamma}}_\theta$ and $\boldsymbol{\eta}_\theta$. The two hyperplanes differ only in the intercept thus we have:

$$u_i(\boldsymbol{\eta}_\theta) = y_i - x_i^T\boldsymbol{\eta}_\theta = y_i - x_i^T\widehat{\boldsymbol{\gamma}}_\theta - \varepsilon = u_i(\widehat{\boldsymbol{\gamma}}_\theta) - \varepsilon$$

and, given (3.8), we have:

$$u_i(\boldsymbol{\eta}_\theta) = \begin{cases} u_i(\widehat{\boldsymbol{\gamma}}_\theta) - \varepsilon > 0 & \textit{for } i \in P; \\ u_i(\widehat{\boldsymbol{\gamma}}_\theta) - \varepsilon < 0 & \textit{for } i \in N; \\ -\varepsilon < 0 & \textit{for } i \in Z. \end{cases}$$

Thus the loss associated to the hyperlane with regression coefficients $\boldsymbol{\eta}_\theta$ is given by:

$$\begin{aligned}
\sum_{i=1}^{n} \ell_\theta\left(y_i - \mathbf{x}_i^T \boldsymbol{\eta}_\theta\right)\frac{y_i}{T} &= \sum_{i=1}^{n} \ell_\theta\left(u_i(\boldsymbol{\eta}_\theta)\right)\frac{y_i}{T} \\
&= \sum_{i\in P}\left[\theta\left(u_i(\widehat{\boldsymbol{\gamma}}_\theta) - \varepsilon\right)\frac{y_i}{T}\right] + \sum_{i\in N}\left[(\theta-1)\left(u_i(\widehat{\boldsymbol{\gamma}}_\theta) - \varepsilon\right)\frac{y_i}{T}\right] + (1-\theta)\varepsilon\ \sum_{i\in Z}\frac{y_i}{T} \\
&= \sum_{i=1}^{n} \ell_\theta\left(u_i(\widehat{\boldsymbol{\gamma}}_\theta)\right)\frac{y_i}{T} - \theta\varepsilon\sum_{i\in P}\frac{y_i}{T} + (1-\theta)\varepsilon\sum_{i\in N}\frac{y_i}{T} + (1-\theta)\varepsilon\sum_{i\in Z}\frac{y_i}{T} \\
&= \sum_{i=1}^{n} \ell_\theta\left(u_i(\widehat{\boldsymbol{\gamma}}_\theta)\right)\frac{y_i}{T} + \varepsilon\left[(1-\theta)\sum_{i\in N}\frac{y_i}{T} - \theta\sum_{i\in P}\frac{y_i}{T} + (1-\theta)\sum_{i\in Z}\frac{y_i}{T}\right].
\end{aligned}$$

Note that if:

$$(1-\theta)\sum_{i\in N}\frac{y_i}{T} - \theta\sum_{i\in P}\frac{y_i}{T} + (1-\theta)\sum_{i\in Z}\frac{y_i}{T} < 0$$

that is:

$$\theta\sum_{i\in P}\frac{y_i}{T} - (1-\theta)\sum_{i\in N}\frac{y_i}{T} > (1-\theta)\sum_{i\in Z}\frac{y_i}{T}$$

we will have:

$$\sum_{i=1}^{n} \ell_\theta\left(u_i(\widehat{\boldsymbol{\gamma}}_\theta)\right)\frac{y_i}{T} > \sum_{i=1}^{n} \ell_\theta\left(u_i(\boldsymbol{\eta}_\theta)\right)\frac{y_i}{T}$$

and $\widehat{\boldsymbol{\gamma}}_\theta$ is not a solution to the minimum problem (3.2).
Hence the following inequality:

$$(\theta-1)\sum_{i\in Z}\frac{y_i}{T} \leqslant (1-\theta)\sum_{i\in N}\frac{y_i}{T} - \theta\sum_{i\in P}\frac{y_i}{T}$$

has to hold.
A similar argument can be used in the case:

$$(1-\theta)\sum_{i\in N}\frac{y_i}{T} > \theta\sum_{i\in P}\frac{y_i}{T}.$$

■

Relation (3.7) can be viewed as an extension of the property of quantity quantiles in the regression context: when the share $\sum_{i\in Z}\frac{y_i}{T}$ of the total amount T accruing to points lying on the regression hyperplane is negligible, then the points below the hyperplane (with negative residuals) absorb roughly a share θ of the total T and the points above the hyperplane (with positive residuals) absorb roughly a share $1-\theta$ of the total T.

Theorem 3.1 holds also when, instead of the weights $y_i;\ i=1,\dots,n$, in the minimum problem (3.2) one use the weights $\omega_i \quad i=1,\dots,n$ such that:

$$\omega_i \geqslant 0 \quad i=1,\dots,n \qquad \text{and} \qquad T=\sum_{i=1}^{n}\omega_i > 0.$$

In particular the relation (3.5) can be proved by setting $\omega_i = 1;\ i=1,\dots,n$.

4 Application to Banca d'Italia survey data

4.1 A brief data description

The data used in this application are supplied by the 2004 central Bank of Italy sample survey of household budgets (Banca d'Italia [2006]). The sampling design follows a two stages procedure: survey units are the municipalities in the first stage and households[3] in the second one. In the first stage municipalities are stratified by region and demographic size. Within each stratum, the municipalities in which interviews would be conducted were selected by including all municipalities with more than 40000 inhabitants while the remaining were randomly selected following a procedure that assigns a drawing probability proportional to size. In the second stage the individual households to be interviewed were then selected randomly. To each household is assigned an initial weight defined as the inverse of its probability of inclusion in the sample.

The 2004 survey covers 8012 households, drawn from registry office records in 344 municipalities, composed of 20581 individuals including 13341 income earners. Part of the sample is composed by households that were interviewed in previous surveys (panel households); the proportion of panel households is 44.98 per cent.

Information collected in the survey include demographic characteristics, housing,

[3]Household means a group of persons living together, whether or not they are related by kinship, who seek to satisfy their needs by pooling all or part of the income earned by group members.

health, education, employment and incomes, payment instruments and forms of saving, non durable and durable consumption, forms of insurance. From the items in the questionnaire, main economic aggregates such as net disposable income and net wealth are calculated.
In this application we consider three variables:

C: non-durable consumption expenditure;

Y: net disposable income;

S: household size.

Non-durable consumption expenditure and net disposable income are both expressed in euros and refer to the whole year 2004; they are obtained aggregating some items of the questionnaire (see appendix A and Banca d'Italia [2006]).
It should be observed that the household net disposable income Y could be negative considering that some items are subtracted in order to form it. This is the case of alimony and gifts paid, interests allowed on financial assets etc. In the 2004 survey this happens for four households and they are excluded from the present application; the households considered are thus $8012 - 4 = 8008$.
In all the computations that follow we considered the weights w_i ($i = 1, \ldots, 8012$) supplied by Banca d'Italia for each household; these weights, as stated before, are defined as the inverse of household's probability of inclusion in the sample. The sum of the weights is $\sum_{i=1}^{8012} w_i = 8012$. The total weight of the four households excluded, because of their negative net disposable income, is 2.489765 thus the total weight of the remaining households is $8012 - 2.489765 = 8009.510235$. In order to get a sum of weights equal to the number of households (8008) we adjusted the weights, with a negligible rectification, by defining the new ones:

$$p_i = w_i \frac{8008}{\sum_{i=1}^{8012} w_i - 2.489765} = w_i \frac{8008}{8009.510235} \qquad i = 1, \ldots, 8008. \tag{4.1}$$

Table 4.1 reports the distribution of the weights of the 8008 households by classes of non-durable consumption expenditure and size. The value in each cell is the sum of the weights p_i of the households with non-durable consumption expenditure in the correspondent row class and size given by the correspondent column label.

Average and median annual household non-durable consumption expenditure are, respectively, €20424.21 and €18000.00. The histogram of non-durable consumption expenditure (figure 1), suggests that its distribution could be modeled by a log-normal distribution.

Table 4.1: Distribution of the weights of the 8008 households by classes of non-durable consumption expenditure C and size S.

Non-durable	**Household size**									
consumption	1	2	3	4	5	6	7	8	9	**Total**
0 ⊣ 8400	431.387	114.912	55.170	22.478	13.379	1.173	1.506	3.848	0.364	644.218
8400 ⊣ 10800	326.212	166.605	64.246	55.124	20.431	3.700	0.000	1.752	0.000	638.070
10800 ⊣ 13200	301.377	304.304	115.752	96.261	18.669	2.596	1.158	0.000	0.000	840.116
13200 ⊣ 15600	254.127	332.313	210.959	167.822	56.692	11.104	0.471	0.399	0.000	1033.886
15600 ⊣ 17400	138.449	214.170	122.207	129.888	45.922	8.262	2.697	1.692	0.000	663.287
17400 ⊣ 19300	150.933	250.982	151.269	164.140	46.322	1.804	0.920	0.000	2.169	768.540
19300 ⊣ 21600	122.180	251.632	200.777	146.162	49.963	7.247	0.253	0.000	0.000	778.214
21600 ⊣ 24000	72.515	144.733	158.665	147.284	30.029	3.722	2.104	0.000	0.000	559.051
24000 ⊣ 26400	46.358	98.563	119.903	111.104	16.503	6.782	0.000	4.807	0.000	404.020
26400 ⊣ 28800	34.239	86.649	100.964	107.017	18.343	3.917	0.000	0.000	0.000	351.129
28800 ⊣ 31000	21.757	46.317	91.266	72.863	20.574	5.159	1.123	0.000	0.000	259.059
31000 ⊣ 33000	28.356	39.591	56.532	66.477	19.105	2.130	0.000	0.000	0.000	212.191
33000 ⊣ 36600	11.640	63.657	91.651	92.394	17.865	14.641	0.000	0.000	0.000	291.848
36600 ⊣ 40800	10.257	39.803	44.491	52.948	13.047	5.526	1.009	0.000	0.000	167.082
40800 ⊣ 45600	4.393	29.543	39.488	53.832	17.629	4.293	0.394	0.000	0.000	149.571
45600 ⊣ 48600	5.579	13.620	23.989	19.366	4.521	0.000	0.000	0.636	0.000	67.712
48600 ⊣ 54100	5.119	11.840	21.987	18.255	4.111	1.822	0.000	0.000	0.000	63.134
more than 54100	7.244	29.890	29.329	42.479	5.391	1.942	0.000	0.000	0.596	116.870
Total	1972.122	2239.125	1698.645	1565.894	418.496	85.819	11.636	13.133	3.129	8008.000

The average household consists of 2.58 members whereas the median household size is 2 (52.59% of the household sizes are no greater than 2).

Table 4.2 (see also figure 2) reports, for different values of θ, the quantiles $\widehat{\zeta}_\theta$ and the quantity quantiles $\widehat{\eta}_\theta$ of the household non-durable consumption expenditure and, in the last column, their ratios. For instance, for $\theta = 0.75$ we have:

$$\widehat{\zeta}_{0.75} = €25200 < €30000 = \widehat{\eta}_{0.75}$$

and this means that, after sorting the households by increasing values of their non-durable consumption expenditure, in order to reach a share equal to 75% of the households we must consider the ones with an expenditure no greater than

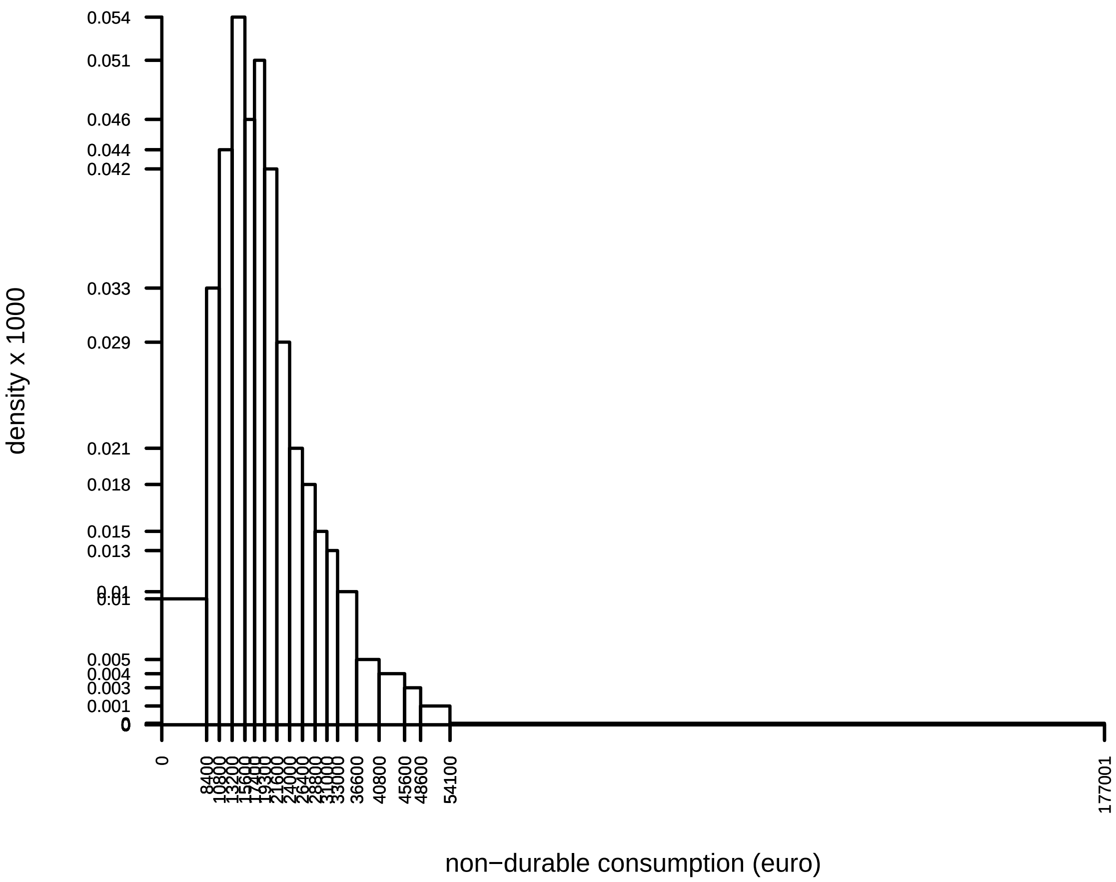


Figure 1: Histogram of non-durable consumption expenditure.

€25200 and that in order to reach a share equal to 75% of the total non-durable consumption expenditure we must consider the households with expenditure no greater than €30000.

Ratios between quantiles and quantity quantiles (last column in table 4.2) are almost constant; this strengthens the idea that the underlying distribution of non-durable consumption expenditure should be log-normal (see Zenga [1984]).

Table 4.3 reports the distribution of the weights of the 8008 households by classes of net disposable income and size. Average and median annual household net disposable income are, respectively, €29494.76 and €23846.72.

Table 4.4 reports the distribution of the weights of the 8008 households by classes of net disposable income and non-durable consumption expenditure.

Table 4.2: Quantiles $\widehat{\zeta}_\theta$ and quantity quantiles $\widehat{\eta}_\theta$ of household non-durable consumption expenditure (values in €).

θ	$\widehat{\zeta}_\theta$	$\widehat{\eta}_\theta$	$\widehat{\zeta}_\theta/\widehat{\eta}_\theta$
0%	1320	1320	1
5%	7440	10200	0.72941
10%	9600	12600	0.76190
15%	10800	14100	0.76596
20%	12000	15600	0.76923
25%	13200	16800	0.78571
30%	14400	18000	0.80000
35%	15000	19200	0.78125
40%	15920	20400	0.78039
45%	16800	21600	0.77778
50%	18000	22800	0.78947
55%	19200	24000	0.80000
60%	20400	26400	0.77273
65%	21600	27899.28	0.77421
70%	22800	30000	0.76000
75%	25200	32400	0.77778
80%	27600	35000	0.78857
85%	30000	38682.74	0.77554
90%	33600	44700	0.75168
95%	40800	56200	0.72598
100%	177000	177000	1

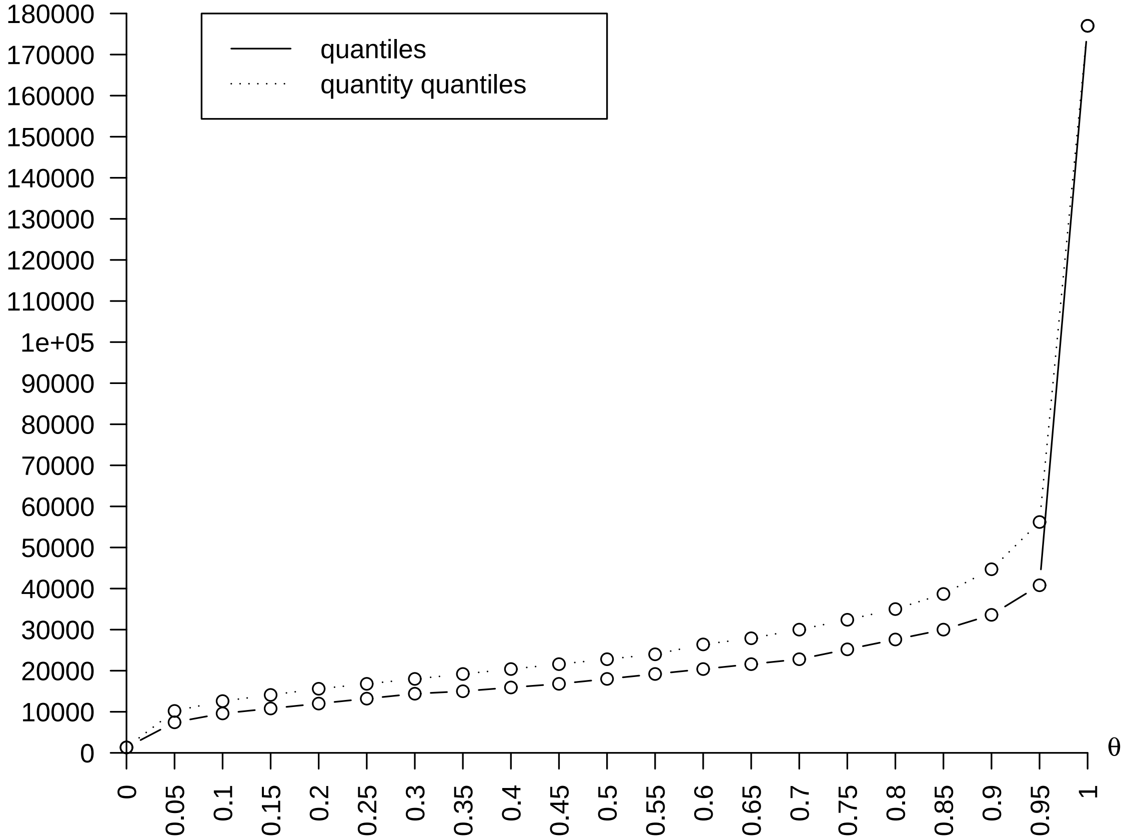


Figure 2: Quantiles and quantity quantiles of household non-durable consumption expenditure (values in €).

Table 4.3: Distribution of the weights of the 8008 households by net disposable income Y and size S.

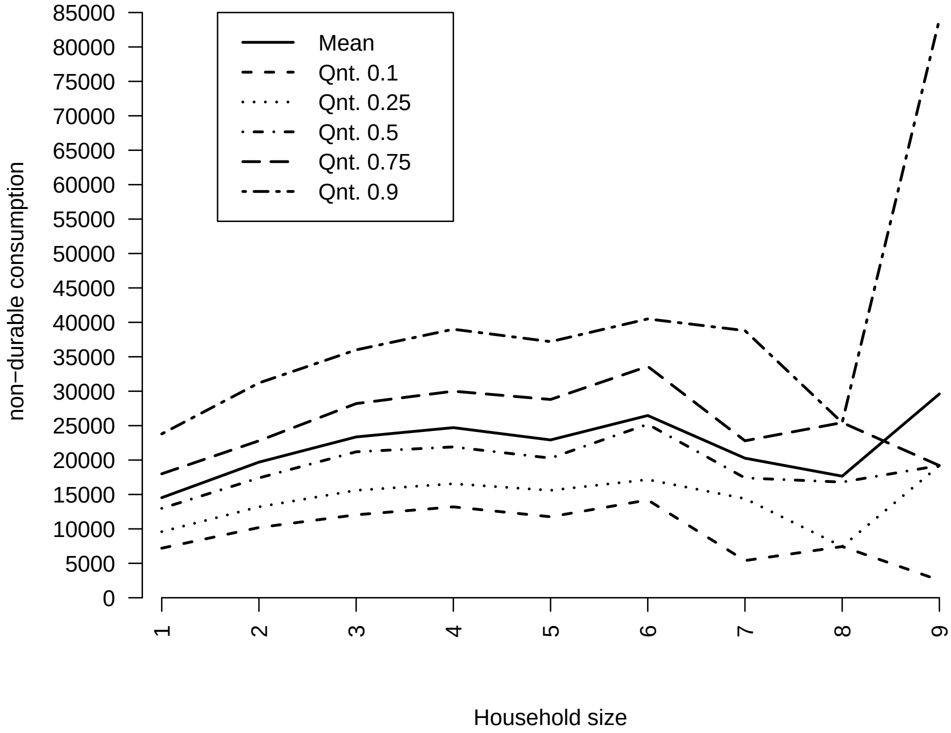
Net disposable income	**Household size**									**Total**
	1	2	3	4	5	6	7	8	9	
0 ⊣ 9000	315.258	84.514	60.422	40.093	16.994	1.173	3.362	1.752	0.364	523.932
9000 ⊣ 12900	390.130	155.057	69.489	54.595	20.919	2.718	1.158	3.848	0.000	697.914
12900 ⊣ 15800	325.539	224.612	99.905	86.845	27.773	5.720	0.000	0.000	0.000	770.393
15800 ⊣ 19200	276.038	279.327	129.299	148.986	54.938	10.547	0.282	0.000	2.169	901.587
19200 ⊣ 22600	224.898	304.206	155.575	134.147	37.291	11.638	0.000	0.399	0.000	868.154
22600 ⊣ 26700	148.946	304.581	153.712	136.994	48.894	8.534	0.253	0.000	0.000	801.913
26700 ⊣ 31100	92.348	218.438	200.664	128.954	40.607	7.611	3.594	2.140	0.000	694.357
31100 ⊣ 35400	48.068	176.894	162.464	154.473	30.287	5.009	0.000	0.000	0.000	577.195
35400 ⊣ 39300	35.650	131.125	121.022	104.542	21.442	3.688	0.000	0.000	0.000	417.469
39300 ⊣ 43100	28.596	83.176	157.800	109.928	21.114	1.619	0.515	0.000	0.000	402.748
43100 ⊣ 46800	13.428	47.386	82.097	89.118	12.948	5.115	0.461	1.705	0.000	252.258
46800 ⊣ 50400	15.757	54.352	42.852	63.208	12.103	2.460	0.000	1.692	0.000	192.423
50400 ⊣ 57700	22.276	54.348	100.202	99.706	33.956	7.479	0.494	0.000	0.000	318.462
57700 ⊣ 64600	11.347	34.265	57.786	75.557	12.188	3.718	0.000	0.000	0.000	194.861
64600 ⊣ 74800	10.367	30.017	36.073	47.154	12.153	4.616	0.394	0.000	0.000	140.775
74800 ⊣ 82600	2.455	16.978	22.714	32.118	4.446	0.823	0.000	0.000	0.000	79.534
82600 ⊣ 95500	0.732	8.313	18.897	17.936	2.559	1.551	0.000	0.000	0.000	49.988
more than 95500	10.289	31.537	27.672	41.536	7.884	1.801	1.123	1.598	0.596	124.037
Total	1972.122	2239.125	1698.645	1565.894	418.496	85.819	11.636	13.133	3.129	8008.000

Table 4.4: Distribution of the weights of the 8008 households by net disposable income Y and non-durable consumption expenditure C.

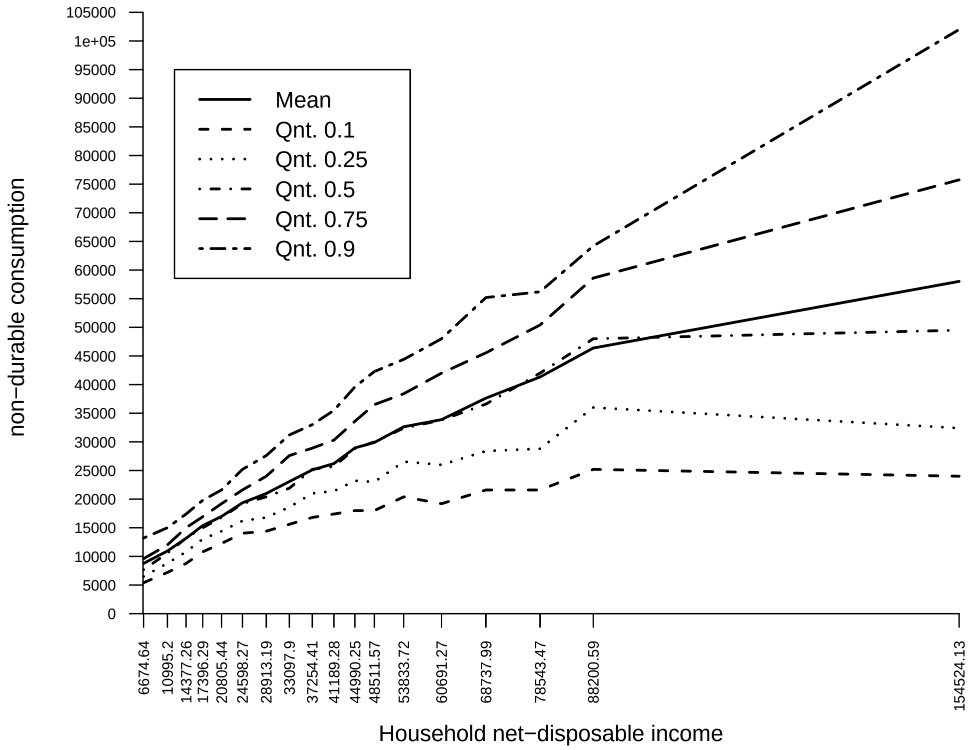
Net disposable income	Non-durable consumption expenditure																		Total
	0 ⊣ 8400	8400 ⊣ 10800	10800 ⊣ 13200	13200 ⊣ 15600	15600 ⊣ 17400	17400 ⊣ 19300	19300 ⊣ 21600	21600 ⊣ 24000	24000 ⊣ 26400	26400 ⊣ 28800	28800 ⊣ 31000	31000 ⊣ 33000	33000 ⊣ 36600	36600 ⊣ 40800	40800 ⊣ 45600	45600 ⊣ 48600	48600 ⊣ 54100	> 54100	
0 ⊣ 9000	349.821	86.393	40.074	16.862	6.332	8.867	4.989	5.211	0.731	0.639	0.266	0.293	0.580	2.093	0.000	0.000	0.000	0.780	523.932
9000 ⊣ 12900	168.985	256.213	156.236	70.620	13.578	7.440	6.635	2.560	9.798	0.000	0.723	1.087	3.508	0.532	0.000	0.000	0.000	0.000	697.914
12900 ⊣ 15800	69.825	132.166	240.984	195.133	61.541	38.335	14.573	10.065	5.780	0.497	0.000	0.000	0.761	0.734	0.000	0.000	0.000	0.000	770.393
15800 ⊣ 19200	30.501	71.676	181.638	278.825	150.584	93.024	41.914	23.786	8.595	3.016	5.764	2.400	6.128	2.708	0.532	0.000	0.000	0.497	901.587
19200 ⊣ 22600	5.819	43.415	96.518	185.105	145.812	194.157	128.830	36.014	18.794	10.629	2.038	0.000	1.023	0.000	0.000	0.000	0.000	0.000	868.154
22600 ⊣ 26700	12.637	19.904	31.782	100.390	131.042	139.895	171.831	95.599	59.043	20.003	8.381	6.250	1.141	0.897	0.000	0.000	0.000	3.118	801.913
26700 ⊣ 31100	1.489	11.775	23.838	103.748	59.230	93.454	131.440	101.651	65.990	46.513	23.881	17.707	5.840	4.242	0.999	0.000	0.000	2.561	694.357
31100 ⊣ 35400	4.601	6.615	17.151	30.060	42.566	77.131	107.279	88.524	46.485	61.372	28.945	20.065	31.715	11.068	2.826	0.000	0.000	0.791	577.195
35400 ⊣ 39300	0.539	1.040	21.574	10.679	13.901	24.838	60.070	61.206	56.818	62.164	39.427	24.171	26.932	6.629	5.827	0.654	0.678	0.322	417.469
39300 ⊣ 43100	0.000	1.143	10.638	15.486	13.519	31.454	49.741	62.385	36.381	38.689	52.721	24.286	34.379	18.592	9.446	2.593	0.000	1.295	402.748
43100 ⊣ 46800	0.000	1.494	2.581	7.357	7.730	17.676	13.635	23.380	27.000	27.399	25.986	30.246	34.700	12.647	14.791	2.625	1.938	1.074	252.258
46800 ⊣ 50400	0.000	0.632	6.511	6.606	5.436	11.112	12.630	16.824	7.738	16.871	18.961	14.966	28.594	19.376	19.714	5.828	0.623	0.000	192.423
50400 ⊣ 57700	0.000	1.706	1.828	8.420	2.072	9.572	15.619	12.405	27.690	30.587	27.253	35.422	57.568	33.167	28.270	14.276	7.655	4.950	318.462
57700 ⊣ 64600	0.000	3.624	8.415	2.578	0.000	9.966	6.209	9.289	12.374	11.500	12.221	19.024	24.314	24.921	17.665	15.474	11.267	6.021	194.861
64600 ⊣ 74800	0.000	0.274	0.347	0.328	3.614	6.997	4.333	3.210	12.084	6.305	4.917	10.315	17.730	14.210	24.397	7.671	9.562	14.481	140.775
74800 ⊣ 82600	0.000	0.000	0.000	0.000	4.951	0.438	3.071	1.137	4.760	8.146	0.682	0.168	5.789	8.628	11.868	5.349	11.798	12.750	79.534
82600 ⊣ 95500	0.000	0.000	0.000	0.000	0.963	0.650	2.812	0.000	1.383	2.649	0.000	1.560	4.061	2.268	5.562	5.902	6.968	15.209	49.988
more than 95500	0.000	0.000	0.000	1.689	0.416	3.533	2.603	5.807	2.577	4.152	6.893	4.231	7.085	4.370	7.674	7.340	12.644	53.022	124.037
Total	644.218	638.070	840.116	1033.886	663.287	768.540	778.214	559.051	404.020	351.129	259.059	212.191	291.848	167.082	149.571	67.712	63.134	116.870	8008.000

4.2 Regression

In this application we consider the household non-durable consumption expenditure C as the dependent variable and household size S and net disposable income Y as the explanatory variables. This section does not aim to be an in-depth analysis, we want only to show how linear regression for quantity quantiles work.
Figure 3 depicts the household non-durable consumption expenditure at various quantiles of the consumption distribution by household size, figure 3(a), and by net disposable income, figure 3(b) (abscissas are the mean net disposable incomes of each class).
Figure 3(a) shows that C generally increases as the household size rises from 1 to 6 members and then decreases (there is a scrape for household with 9 members but they are negligible because their overall weight with respect to the total is just $3.129/8008 < 0.04\%$). Moreover conditional quantiles of the non-durable consumption expenditure are nearly parallel throughout the household sizes between 1 and and 6 denoting, given that the relative weight of households with size greater than 6 is just $27.898/8008 < 0.35\%$, that there is no significant differential in household size effect across these quantiles. In figure 3(b) we observe that quantiles of C increase as the net disposable income rises. Nevertheless, there are some major differences in the increases at various quantiles in particular there is a steeper increase in consumption expenditure at higher quantiles of the distribution thus the spread between quantiles increases as the income rises. In this situation quantile regression allows to investigate changes in the consumption at different points of the distribution.
Figures 4 (a) and (b) follow the same approach as figure 3 but depict the quantity quantiles instead of quantiles. In both the figures the curve of the 0.50 quantity quantile, i.e. the dividing value (see section 2), is always over the curve representing the means.

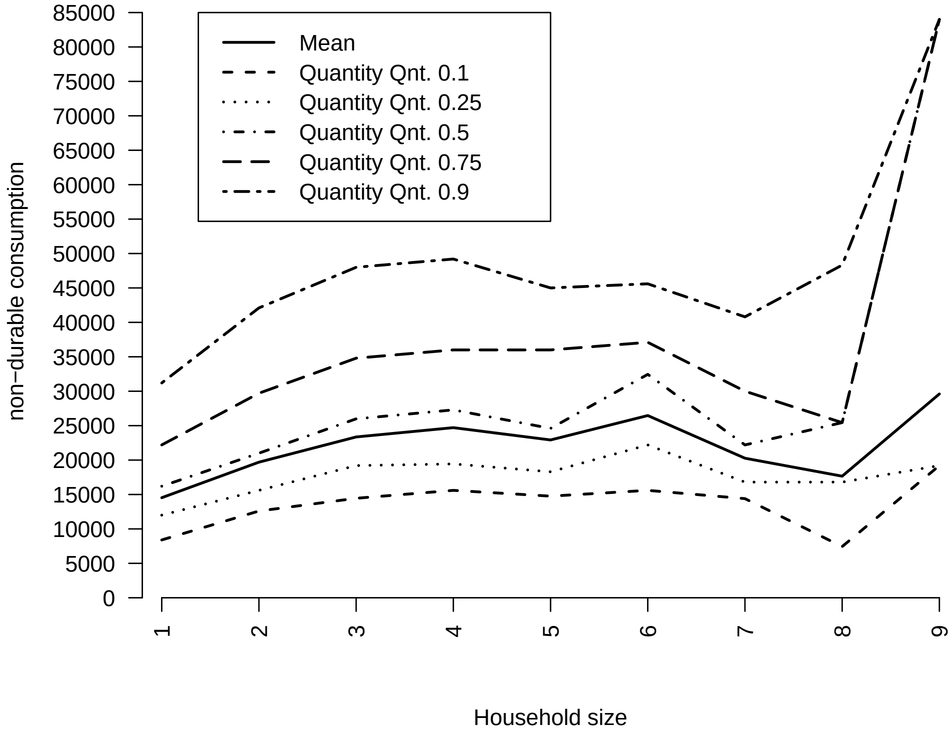


(a)

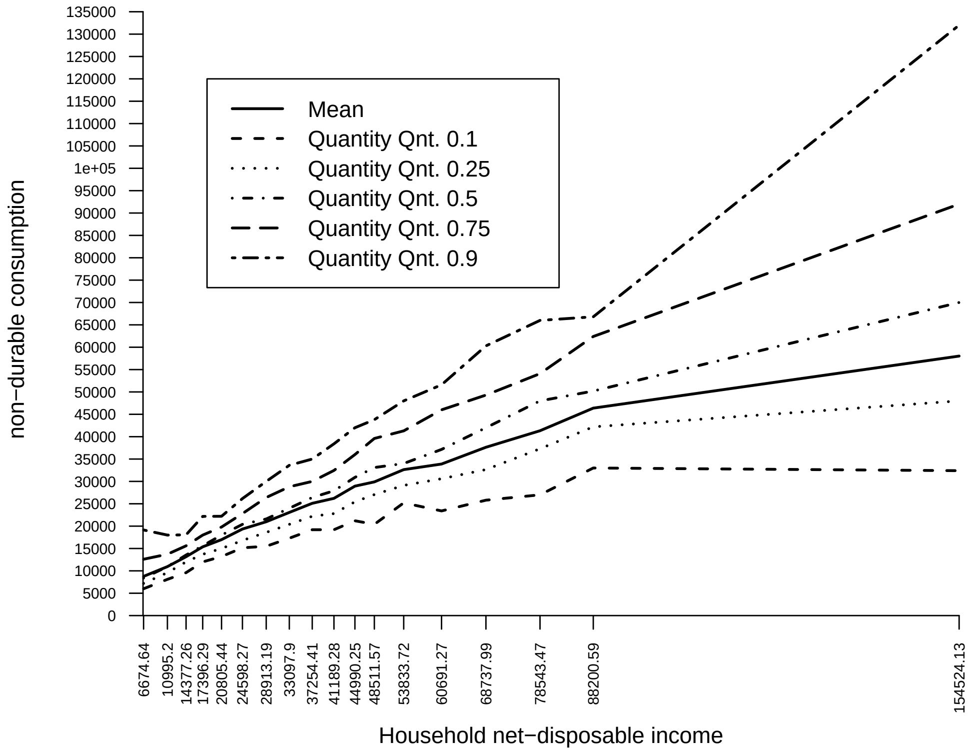


(b)

Figure 3: Non-durable consumption expenditure by quantiles (values in €).



(a)



(b)

Figure 4: Non-durable consumption expenditure by quantity quantiles (values in €).

The linear model we consider for non-durable consumption expenditure is:

$$c_i = \beta_0 + \beta_1 \log(s_i) + \beta_2 y_i + \varepsilon_i \quad i = 1, \ldots, 8008. \tag{4.2}$$

We take the logarithm of the household size S because from figure 3 (a) it seems to be more appropriate to describe the conditional mean of the non-durable consumption expenditure and the model might improve the appearance of the plot.
We start the analysis by improving least squares regression on model (4.2) assuming that the expected value of the error term conditional on the regressors is zero ($\mathbb{E}(\varepsilon_i|s_i, y_i) = 0$). The estimated model has $R^2 = 0.4468$ and all the regression coefficients are significant: household net disposable income matters much more than (log) household size when looking at the t-ratios (see table 4.5).

Table 4.5: Least squares regression analysis for model (4.2).

Predictor	Coefficient	Standard Error	t-value	p-value
Intercept	9317.4183	185.5	50.23	$< 2 \cdot 10^{-16}$
log (S)	3324.4177	180.2	18.44	$< 2 \cdot 10^{-16}$
Y	0.2855	0.003835	74.44	$< 2 \cdot 10^{-16}$

For quantile regression we rewrite model (4.2) by taking into account the dependence of the regression coefficients on $\theta \in (0; 1)$ (the order of the quantile we are interested in):

$$c_i(\theta) = \beta_0(\theta) + \beta_1(\theta) \log(s_i) + \beta_2(\theta) y_i + \varepsilon_i \quad i = 1, \ldots, 8008 \tag{4.3}$$

and assume that the θth quantile of the error term conditional on the regressors is zero ($Quant_\theta(\varepsilon_i|s_i, y_i) = 0$).
The minimum problem is thus, according to (1.11):

$$\min_{\beta_0(\theta),\beta_1(\theta),\beta_2(\theta)} \frac{1}{8008} \sum_{i=1}^{8008} \ell_\theta \left[c_i - \beta_0(\theta) - \beta_1(\theta) \log(s_i) - \beta_2(\theta) y_i\right] \; p_i \tag{4.4}$$

Results for estimation of quantile regression at the conditional quantiles of order 0.10, 0.25, 0.50, 0.75 and 0.90 are presented in table 4.6. The analysis has been carried out by the quantreg package (Koenker [2006]) under the software R; standard error are computed by the default method. Regression coefficients are all significant

Table 4.6: Quantile regression results for model (4.3).

θ	0.10		0.25		0.50		0.75		0.90	
Predictor	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std. Error	Coef.	Std.Err.
Intercept	5143.810	242.298	5460.279	136.530	5438.988	116.213	5819.889	169.055	6472.912	334.632
log(S)	2145.492	238.365	2018.661	53.023	1798.585	139.468	1630.773	122.622	1658.239	297.934
Y	0.199	0.016	0.316	0.009	0.452	0.009	0.570	0.010	0.678	0.019

(p-values of t-statistics are less than 10^{-6}).

Regression coefficients for $\log(S)$ decrease as the order of the conditional quantile increases meaning that for lower quantiles the rise of household size implies a larger growth of expenditure for non-durable consumptions than for higher ones.
On the other side non-durable consumption expenditure of households in the lower quantiles seems to be less reacting to a net disposable income growth. Moreover we observe that regressions coefficients obtained for median regression ($\theta = 0.50$) are considerably different from the ones obtained for least squares regression; this is due to the less sensitiveness (i.e. higher robustness) of median regression to the presence of extreme values of the dependent variable C with respect to least squares regression. A more detailed analysis should be interesting but this lies outside the aim of the present work.
Regression coefficients reported in table 4.6 have been obtained by ordinary quantile regression just considering the weights (4.1). Now we are interested in running quantile regression for the conditional quantity quantiles of the non-durable consumption expenditure (see figure 4). To this aim, according to (3.2), it is necessary to consider different weights of the residuals; the minimum problem is thus:

$$\min_{\beta_0(\theta),\beta_1(\theta),\beta_2(\theta)} \frac{1}{8008} \sum_{i=1}^{8008} \ell_\theta \left[c_i - \beta_0(\theta) - \beta_1(\theta)\log(s_i) - \beta_2(\theta) y_i\right] \; p_i \, \frac{c_i}{\overline{c}} \tag{4.5}$$

where $\overline{c} =$ €20424.21 denotes the average household non-durable consumption expenditure. In (4.5), with respect to (4.4), we consider the additional weights $\frac{c_i}{\overline{c}}$; this allows us to obtain regression coefficients for the *conditional quantity quantiles* of the dependent variable instead of the usual ones.
Results for estimation of quantile regression at the conditional quantity quantiles of order 0.10, 0.25, 0.50, 0.75 and 0.90 are presented in table 4.7.

Table 4.7: Quantity quantile regression results for model (4.3).

θ	0.10		0.25		0.50		0.75		0.90	
Predictor	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std. Error	Coef.	Std.Err.
Intercept	6431.436	455.571	6794.084	197.246	6376.898	329.441	6062.241	408.014	6506.396	607.906
log(S)	2476.242	267.191	2337.755	178.197	1854.700	183.141	1228.834	285.739	1210.259	509.912
Y	0.208	0.023	0.316	0.018	0.471	0.016	0.630	0.020	0.765	0.019

Figures 5, 6 and 7 compare the intercepts and the regression coefficients obtained for the quantile regression hyperplane (table 4.6) with the ones obtained for the quantity quantile regression one (table 4.7).

As to intercept (figure 5) we observe that for the quantity quantiles hyperplane it is always higher than the one obtained for quantile regression and that they get near when θ increases. This aspect reflect the results we have obtained in table 4.2 and figure 2 computing unconditional quantiles and quantity quantiles of the non-durable consumption expenditure.

With respect to the regression coefficients of the (log) household size (figure 6), we find that for both the hyperplanes the coefficients decrease as θ increases but the reduction is more remarkable for the quantity regression quantile hyperplane. In particular for lower quantile orders ($\theta < 0.5$) conditional quantity quantiles of C are more reacting to an increase of household size than quantiles and vice versa for higher quantile orders ($\theta > 0.5$).

As to the regression coefficients for net disposable income (figure 7) we notice a similar trend for both the hyperplanes; for $\theta > 0.5$ the regression coefficients for the quantity quantile hyperplane are slightly higher than the ones obtained for quantiles. This reflects in part what we can find by comparing figures 3(b) and 4(b).

5 Conclusions and further developments

In this paper we showed that the definition of the θth sample quantile as the solution to a minimization problem introduced by Koenker and Basset [1978] can be easily extended to the θth sample quantity quantile. Consequently the results obtained

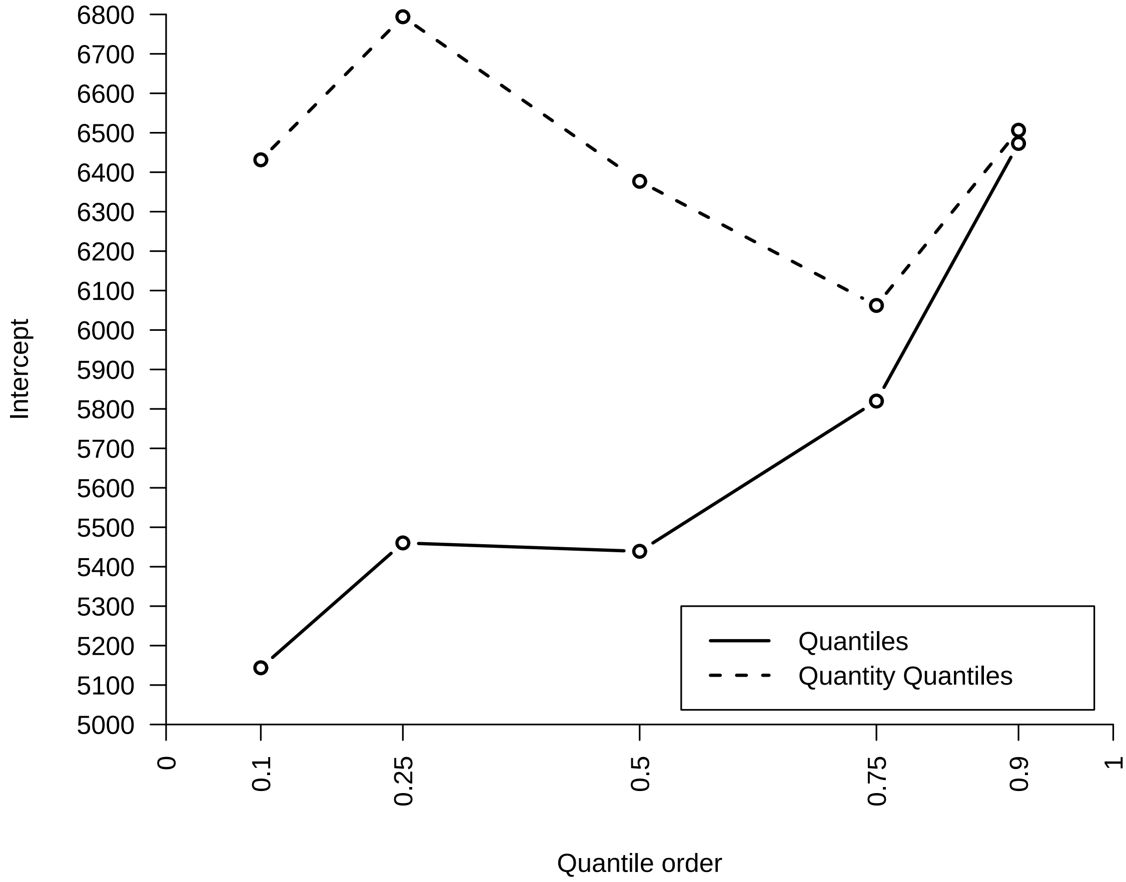


Figure 5: Comparison of the intercepts of the hyperplanes for conditional quantiles and quantity quantiles for different values of θ.

by Koenker and Basset [1978] can be extended in order to perform linear regression for quantity quantiles. Obviously quantity quantile regression shares with quantile regression many features and properties (see for example theorem 3.1) that should be investigated.

In section 4 we showed a possible application of the methodology here proposed by considering a linear regression model for household non-durable consumption expenditure with a brief comparison of the results obtained for both the regressions: quantiles and quantity quantiles. This subject needs a deeper analysis: in particular it is interesting to investigate which application fields can take advantage from the additional information that quantity quantile regression can offer.

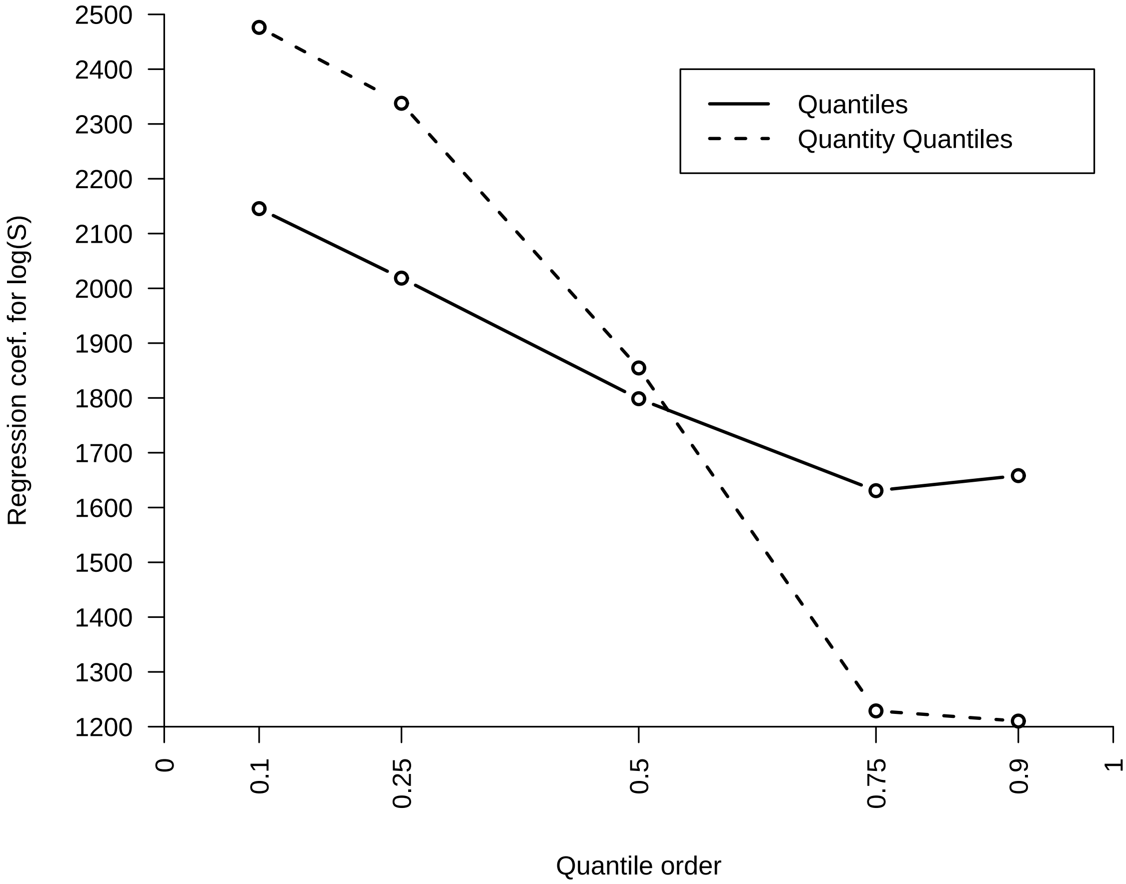


Figure 6: Comparison of the regression coefficients for log(S) of the hyperplanes for conditional quantiles and quantity quantiles for different values of θ.

A Appendix

Non-durable consumption is obtained by Banca d'Italia with the following expression:

$$C = (a + b) \cdot 12 + c + d + e/12$$

where:

a) monthly average spending of the household in 2004 on all consumer goods (both food and non-food consumption) excluding only:

- purchases of: precious objects, cars and household appliances and furniture;
- maintenance payments or other contributions to relatives or friends;

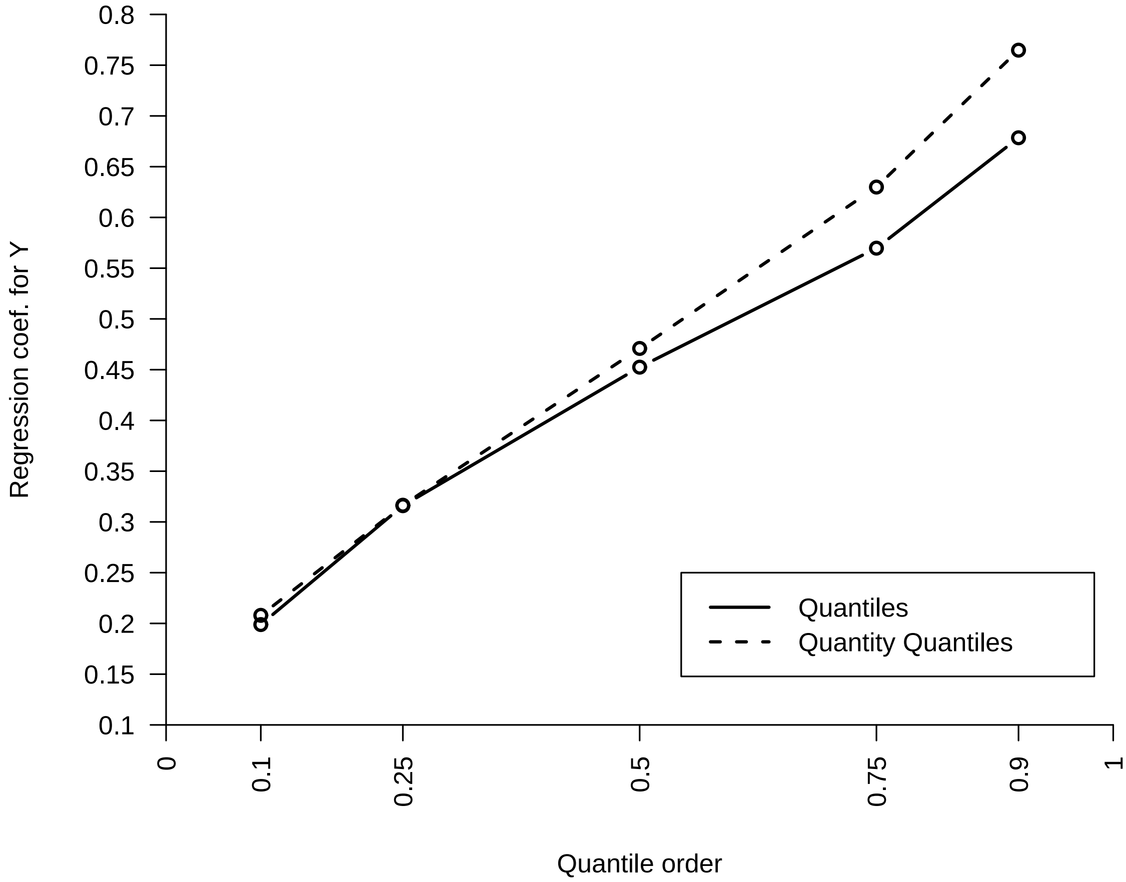


Figure 7: Comparison of the regression coefficients for Y of the hyperplanes for conditional quantiles and quantity quantiles for different values of θ.

- extraordinary maintenance of own dwelling;
- rent for the dwelling;
- mortgage payments;
- life insurance premiums;
- contributions to private pension funds.

b) monthly rent paid in 2004 for the dwelling, excluding condominium charges, heating and other sundry expenses;

c) monetary value of the fringe benefits such as lunch vouchers, trips, company cars etc. (excluding housing) get as employee in 2004;

d) monthly rent (excluding condominium charges, heating or other sundry expenses) of the household residence[4] that the household should gain assuming it wanted to rent it.

e) year's rent the household should gain assuming it wanted to rent other properties (dwelling other than the principal one) owned at the end of 2004.

Net disposable income Y is obtained adding up the following items:

I) Compensation of employees

- Net wages and salaries
- Fringe benefits

II) Pensions and net transfers

- Pensions and arrears
- Other transfers (economic assistance, scholarships, etc.)

III) Net income from self-employment

IV) Property income

- Income from buildings
- Income from financial assets.

Acknowledgements. This research was partially supported by Fondo d'Ateneo per la Ricerca anno 2005 - Università degli Studi di Milano-Bicocca.

[4]This item is applied only if the household residence is owned by the household or occupied under: redemption agreement, in usufruct or free of charge

References

Banca d'Italia (2006). I bilanci delle famiglie italiane nell'anno 2004. *Supplementi al Bollettino Statistico*, (7-17).

Bassett, G. and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Association*, 73(363):618–622.

Berti, P. and Rigo, P. (1999). Concentration Curve and Index, Zenga's. In Kotz, S., editor, *Encyclopedia of Statistical Sciences*, volume Update Volume 3, pages 101–104. John Wiley & Sons.

Bowley, A. (1972). *F.Y. Edgeworth's Contributions to Mathematical Statistics*. Reprints of Economic Classics. Augustus M. Kelley. First Edition 1928 (London: Royal Statistical Society, *Adelphi Terrace*, W.C. 2,1928).

De La Place, M. (1966). *Celestial Mechanics*, volume II. Chelsea Publishing Company, Inc. Translated, with a commentary, by N. Bodwditch.

Efron, B. (1991). Regression Percentiles Using Asymmetric Squared Error Loss. *Statistica Sinica*, 1:93–125.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Hallock, K. and Koenker, R. (2001). Quantile Regression. *Journal of Economic Perspectives*, 15(4):143–156.

Jones, M. (1994). Expectiles and M-quantiles are Quantiles. *Statistics & Probability Letters*, 20:149–153.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. (2006). The quantreg package. http://cran.r-project.org/src/contrib/Descriptions/quantreg.html.

Koenker, R. and Basset, G. J. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Zhao, Q. (1994). L-Estimation for Linear Heteroscedastic Models. *Journal of Nonparametric Statistics*, 3:223–235.

Mineo, A. (2003). Il Problema della Migliore Combinazione delle Osservazioni da Galileo a Subbotin. *Statistica & Applicazioni*, I(1):47–66.

Mortara, G. (1933). *Sommario di Statistica*. Università Bocconi di Milano.

Newey, W. and Powell, J. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55(4):819–847.

Peracchi, F. (2001). *Econometrics*. John Wiley & Sons.

Radaelli, P. and Zenga, M. (2006). Linear Regression for Quantity Quantiles. In Società Italiana di Statistica, editor, *Proceedings of the XLII Scientific Meeting*, volume Contributed Sessions, pages 605–608. Cleup. Università di Torino, 14-16 june 2006.

Stigler, S. (1984). Studies in the History of Probability and Statistics XL Boscovich, Simpson and a 1760 Manuscript Note on Fitting a Linear Relation. *Biometrika*, 71(3):615–620.

Zenga, M. (1984). Proposta per un Indice di Concentrazione Basato sui Rapporti tra Quantili di Popolazione e Quantili di Reddito. *Giornale degli Economisti e Annali di Economia*, pages 301–326.

Zenga, M. (1985). Un Secondo Indice di Concentrazione Basato sui Rapporti tra Quantili di Reddito e Quantili di Popolazione. *Rivista di Statistica Applicata*, (3):143–154.

Zenga, M. (1990). Concentration Curves And Concentration Indexes Derived From Them. In Dagum, C. and Zenga, M., editors, *Income and Wealth Distrbution,* *Inequality and Poverty*. Springer-Verlag. Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measuremente and Applications, held at the University of Pavia, Italy, September 28-30, 1989.