Computational Statistics & Data Analysis 6 (1988) 353–359
North-Holland

A property of the observations fit by the extreme regression quantiles

Gilbert W. BASSETT, Jr.
Department of Economics, University of Illinois, Box 4348, Chicago, IL 60680, USA

Abstract: The extreme regression quantile estimates have recently been proposed as a computationally fast method for detecting discrepant points and constructing high breakdown estimates for the linear model. Further support for this proposal comes from the following property. The convex hull of the observations which exactly fit an extreme regression quantile estimate must contain the mean values of the design variables. Hence, the observations identified by an extreme regression quantile cannot all be clustered on one side of the overall mean values of the design variables.

Keywords: Regression quantiles, Extreme values, Breakdown point.

1. Introduction

The extreme regression quantile statistics were proposed as a tool for detecting outliers and for constructing high breakdown estimates in a recent paper by Steve Portnoy [4]. The observations fit exactly by the extreme regression quantiles, θ near zero and one, correspond to linear model versions of the smallest and largest order statistics and, hence, might be expected to identify outlying values of the dependent variable; see Koenker and Bassett [2]. The regression quantiles, as M-estimates, may be also expected to be influenced by and tend to go through extreme design points. Portnoy proposes an iterative procedure for peeling such observations away from the data set. He shows that an estimate based on this method achieves high breakdown without being as computationally difficult as other high breakdown procedures; see Rousseeuw [5] and Siegel [6].

This note establishes another property for the extreme regression quantiles. The mean values of the explanatory variables must fall inside the convex hull of the observations which exactly fit the extreme regression quantiles. For example, in Figure 1 there are two lines which fit two of the observations and which have all residuals nonnegative. One of these lines corresponds to the smallest extreme regression quantile estimate. The property established below says that line I must be the extreme regression quantile because it is the one which fits observations on opposite sides of the design mean $\bar{z}$. Conversely, the observations at the design

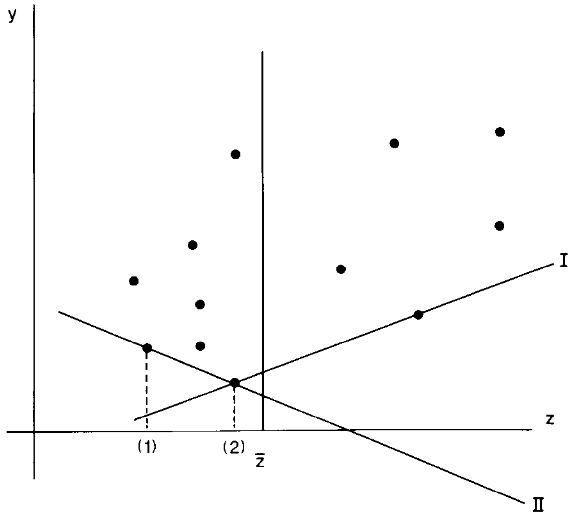

Fig. 1. Extreme regression quantile estimates.

points labeled (1) and (2) can never (for any values taken by the dependent variable) correspond to an extreme regression quantile estimate because they fall on the same side of $\bar{z}$. The general property offers some support to Portnoy's idea by showing that the design points fit by the extreme regression quantiles cannot all cluster to one "side" of the overall mean values of the design variables.

The properties of the extreme regression quantiles are presented and explained further in the next section. The proofs are in Section 3. The Appendix discusses some situations which for simplicity not covered by the theorem in Section 2.

2. The results

Let y denote a vector in R^n. The design for the linear model is an $(n \times p)$ matrix X where rank $X = p$, and the first column is a vector of ones; the model has an intercept. The ith observation on the $(p-1)$ nonconstant explanatory variables is $z_i = (z_{i1}, \ldots, z_{i(p-1)})$ so that $x_i = (1, z_i)$. The mean vector of the design variables is $\bar{x} = (1, \bar{z}) = n^{-1}\Sigma_{i=1}^n x_i$.

The θth regression quantile is denoted by $B^*(\theta)$ and is the solution set to

$$\min_{b \in R^p} \rho(b:\theta) \tag{1}$$

where θ is a parameter in (0, 1) and

$$\rho(b:\theta) = \sum_{i=1}^{n} c(y_i - x_i b : \theta)$$

and $c(e:\theta) = \theta\,|\,e\,|$ or $(1-\theta)\,|\,e\,|$ as $e > 0$ or $e < 0$.

The smallest regression quantile is denoted by $\underline{B}^*$ and is defined by

$$\underline{B}^* = \{\beta^* \mid \beta^* \in B^*(\theta) \text{ some (small } \theta),$$
$$\text{and } \operatorname{sgn}(y_i - x_i\beta^*) \geqslant 0,\ i = 1, \ldots, n\}.$$

We will say $\underline{B}^*$ is unique if it contains a single element. When $p = 1$, $\underline{B}^*$ is equal to the smallest order statistic (and is always unique).

The largest regression quantile $\overline{B}^*$ is defined in a similar fashion; that is,

$$\overline{B}^* = \{\beta^* \mid \beta^* \in B^*(\theta), \text{ some } \theta \text{ (close to one)},$$
$$\text{and } \operatorname{sgn}(y_i - x_i\beta^*) \leqslant 0,\ i = 1, \ldots, n\}.$$

All the results below hold for both $\underline{B}^*$ and $\overline{B}^*$ and will be stated solely for the smallest regression quantile.

Two simplifying restrictions will be used in stating results. The first takes $\underline{B}^*$ to be unique. As explained in the appendix, there are only very special situations under which $\underline{B}^*$ is not unique and, in case this does occur, it can always be eliminated by a slight perturbation of the data. When $\underline{B}^*$ is unique it will correspond to an estimate which fits p of the observations exactly and the remaining observations will have nonnegative residuals. The second simplifying restriction is a restriction on $y \in R^n$ which insures that the residuals from the $\underline{B}^*$ fit are not only nonnegative but positive. We will then be able to unambiguously identify $\underline{B}^*$ with the p observations which it fits exactly. The events not covered by this restriction are best treated separately and they can be in practice ignored if y is the value of a continuous random vector because they then have probability zero. (The first restriction corresponds to there being no multiple optimal solutions to the linear programming regression quantile problem (see Koenker and d'Orey [3]) and the latter corresponds to the absence of degeneracy.)

Let h denote p distinct indices from $(1, \ldots, n)$ and let $\bar{h}$ denote the complementary $n - p$ indices. $X(h)$, $X(\bar{h})$, $y(h)$ and $y(\bar{h})$ will denote submatrices or subvectors with rows or components in the indicated h or $\bar{h}$ index set.

Let H denote all the h's such that rank $X(h) = p$. Corresponding to each h there is a $b(h) = X(h)^{-1}y(h)$ where $b(h)$ is an estimate which exactly fits the observations $i \in h$. Let S denote the set of all $b(h)$, $h \in H$.

If the solution set $B^*(\theta)$ is a single point, $\beta^*(\theta)$, then $\operatorname{sgn}(y_i - x_i\beta^*(\theta)) = 0$ for *at least* p of the observations; that is, $\beta^*(\theta) \in S$. If the solution set is not a single point it will be the convex hull of at most $p + 1$ points each of which is in S. (But this fails to hold at $\theta = 0$ where the solution set is unbounded and contains any b which makes all the residuals non-negative. This is the reason for taking $0 < \theta < 1$.) The uniqueness assumption for $\underline{B}^*$ means that $B^*(\theta)$ is unique when θ increases slightly away from zero.

The simplifying restriction on y which insures a one-to-one mapping between an observation subset h and an estimate $b(h)$ is to take $y \in \overline{R}^n$ where

$$\overline{R}^n = \{y \in R^n \mid y_i - x_i b(h) \neq 0,\ \text{all } i \in \bar{h},\ \text{all } h \in H\}$$

(the dependence of $\bar{R}^n$ on X will be hopefully clear from the context). When y is in $\bar{R}^n$ there will be no more than p points, $(y_i, z_i) \in R^p$ which fall on any nonvertical hyperplane in R^p. The assumption $y \in \bar{R}^n$ is sometimes referred to as y being in general position.

Finally, given vectors $v_i \in R^n$, $i = 1, \ldots, T$ let the convex hull be denoted by

$$\operatorname{conv}(v_i \mid i = 1, \ldots, T) = \left\{ \sum_{i=1}^{T} \lambda_i v_i \,\middle|\, \sum_{i=1}^{T} \lambda_i = 1,\ 0 \leqslant \lambda_i \leqslant 1 \right\}.$$

For example, if $v_i \in R$ then $\operatorname{conv}(v_1, v_2)$ would be the interval with endpoints v_1 and v_2. If $v_i \in R^2$ then $\operatorname{conv}(v_1, v_2, v_3)$ would be a triangle with corners at v_1, v_2, and v_3 (or a line segment if the v_i are not linearly independent).

Theorem. *Suppose $y \in \bar{R}^n$ and $\underline{B}^*$ is unique. Let h^* denote the p indices such that $\underline{B}^* = b(h^*)$. Then*

$$\bar{z} \in \operatorname{conv}(z_i \mid i \in h^*)$$

Conversely, let h *be any set of indices such that $\bar{z} \notin \operatorname{conv}(z_i \mid i \in h)$. Then $b(h)$ will not be in $\underline{B}^*$ for any $y \in \bar{R}^n$.*

To see what the theorem says for $p = 3$ consider Figure 2. Each point represents a design point $(z_{i1}, z_{i2}) \in R^2$, $i = 1, \ldots, n$, and the mean of the design variables is (0, 0). Consider the design points labeled 1, 2, 3. Visualize the data (y_i, z_i) in R^3 and suppose the plane which contains these three observations lies

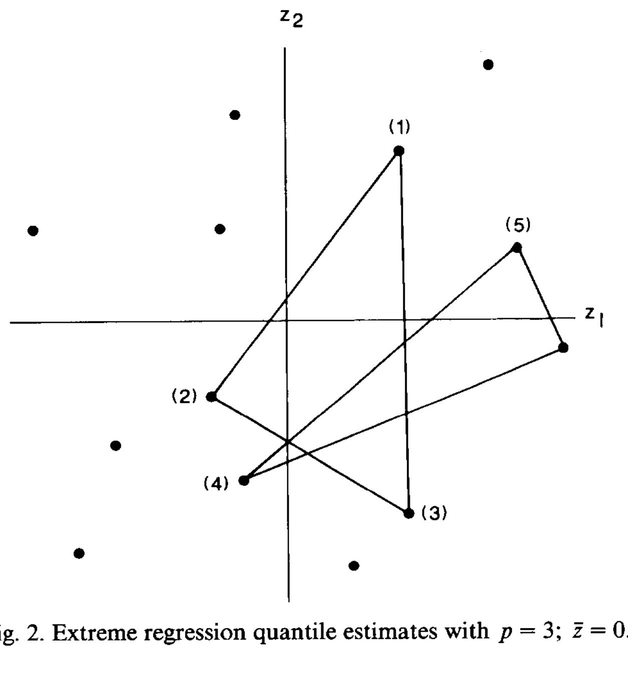

Fig. 2. Extreme regression quantile estimates with $p = 3$; $\bar{z} = 0$.

below all the remaining (y_i, z_i) points. The theorem says that this must be the extreme regression quantile hyperplane because the mean (0, 0) is in the convex hull of observations 1, 2, and 3.

Conversely, consider the observations labeled 4, 5, 6. According to the theorem there is no configuration of $y \in \overline{R}^n$ which will ever make the (4, 5, 6) hyperplane in R^3 equal to the extreme regression quantile estimate. The problem with (4, 5, 6) is that it does not contain the mean vector (0, 0) and hence can never yield the extreme regression quantile estimate.[1]

3. Proofs

Let $\rho'(b : w : \theta)$ be the directional derivative of $\rho(b : \theta)$ at b in the direction w, for $0 < \theta < 1$. We have

$$\rho'(b : w : \theta) = \sum_{i=1}^{n} \left[-\theta + 0.5 - 0.5\,\mathrm{sgn}^*(y_i - x_i b : -x_i w) \right] x_i w$$

where $\mathrm{sgn}^*(e : u) = \mathrm{sgn}(e)$ if $e \neq 0$ and $\mathrm{sgn}(u)$ otherwise. Write $w = X(h)^{-1}v$, $v \in R^p$, $b = b(h) = X(h)^{-1}y(h)$ so that

$$\rho'\left(b(h) : X(h)^{-1}v : \theta\right) = -\theta n \sum_{i=1}^{p} \lambda_i(h) v_i + \sum_{i=1}^{p} \left[0.5 v_i + 0.5\,|v_i|\right] + s(v)$$

where

$$s(v) = \sum_{i \in \bar{h}} \left[0.5 - 0.5\,\mathrm{sgn}^*\left(y_i - x_i b(h) : -x_i X(h)^{-1}v\right)\right] x_i X(h)^{-1} v \qquad (3.1)$$

and

$$\lambda(h) = \left(\lambda_1(h), \ldots, \lambda_p(h)\right) = \bar{x} X(h)^{-1}. \qquad (3.2)$$

Now suppose $b(h) \in \underline{B}^*$ and y is in $\overline{R}^n$. We then have $s(v) = 0$. Further, since ρ is convex it must be the case that $\rho'(b(h) : X(h)^{-1}v : \theta) \geqslant 0$ for all $|v| = 1$ and, as can be verified, this holds if and only if $\lambda_i(h) \geqslant 0$ for all $i \in h$. That is,

$$\text{if } y \in \overline{R}^n \text{ and } b(h) \in \underline{B}^* \text{ then } \lambda_i(h) \geqslant 0, \text{ all } i \in h. \qquad (3.3)$$

[1] The case of a nonunique extreme regression quantile estimate corresponds to $\bar{z}$ being on the boundary of the convex hull. Also, a general geometric property which is not instantly obvious follows from the fact that there is always some smallest regression quantile estimate. Namely, for any $y \in \overline{R}^n$ there will always exist a hyperplane which goes through p observations, where all remaining observations are above the hyperplane *and* where the mean vector $\bar{z}$ is in the convex hull of the p observations (one might have thought (incorrectly) that some configuration of data could leave $\bar{z}$ outside the convex hull of all the "lower" observations).

Also, it can be verified that $\lambda_i(h) > 0$ for all $i \in h$ if and only if the directional derivative is positive at all $|v| = 1$ which by convexity means $\underline{B}^*$ is unique, so

$$\text{if } y \in \overline{R}^n \text{ and } b(h) \text{ is unique then } \lambda_i(h) > 0, \text{ all } i \in h. \tag{3.4}$$

Now look at $\lambda(h)$:

$$\bar{x}X(h)^{-1} = \lambda(h) \quad \text{or} \quad \bar{x} = \lambda(h)X(h),$$

which in detail says

$$1 = \sum_{i=1}^{p} \lambda_i(h) \tag{3.5}$$

because the model has an intercept and

$$\bar{z} = \sum_{i \in n} \lambda_i(h) z_i \tag{3.6}$$

Now (3.3) says that when $y \in \overline{R}^n$ and $b(h) \in \underline{B}^*$ there will exist $\lambda_i(h) \geqslant 0$ satisfying (3.5) and (3.6); that is, $\bar{z}$ is a convex combination of the z_i, $i \in h$; this is the first part of the theorem. (Actually, if $\underline{B}^*$ is unique then $\bar{z}$ must be in the interior of the convex hull of the z_i, $i \in h$ because uniqueness implies $\lambda_i(h) > 0$; see Appendix.)

To establish the converse suppose $y \overline{\in} R^n$ and $b(h) \in \underline{B}^*$ for some h such that

$$\bar{z} \in \sum_{i \in h} \lambda_i(h) z_i.$$

The latter means that (3.6) and (3.5) will hold only if for at least one i we have $\lambda_i(h) < 0$; suppose this holds at $i = 1$. Then from (3.1) the directional derivative at $v_i = e_i = (-1, 0, \ldots, 0)$ is $\theta n \lambda_1(h) < 0$, so that ρ is decreasing at $b(h)$ and this contradicts $b(h) \in \underline{B}^*$.

Appendix

Consider the set of θ where $B^*(\theta)$ is not unique. (When $p = 1$ and $y \in \overline{R}^n$ this set consists of the points(i/n), $i = 1, \ldots, n$.) Suppose it were true that this set was finite for any $y \in \overline{R}^n$ and any design X. There would then exist an open interval $(0, \theta')$ where $B^*(\theta)$ was constant and unique and hence the smallest regression quantile would always be unique.

The conjecture about the finite number of θ values where B^* is not unique is, however, false; see the example in Bassett and Koenker [1, p. 410]. For another example, consider the data in Figure A1. In this case there is an interval $(0, \theta')$ where $B^*(\theta)$ is a nonsingleton set for each $\theta \in (0, \theta')$. (The lines I and II and all convex combinations of these lines are in $B^*(\theta)$ for small θ values.) The next theorem shows that this occurs only when the design points take special values and hence can be eliminated by a slight perturbation of design points; Portnoy [4] refers to this as dithering with the data.

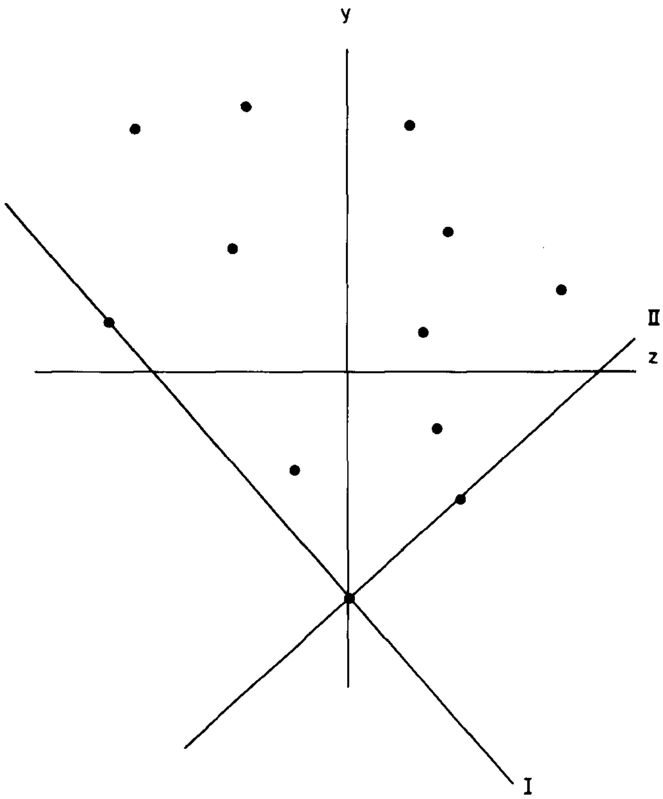

Fig. A1. Nonunique extreme regression quantiles; $\bar{z} = 0$.

Theorem A1. *Suppose $\bar{z}$ is not on the boundary of* $\operatorname{conv}(z_i \mid i \in h)$ *for any $h \in H$. Then $\underline{B}^*$ is unique for all $y \in \overline{R}^n$.*

Proof: Under the stated conditions we have, on using results from the proof in Section 3, that $\rho'(\beta^* : w) \geqslant 0$, all w, actually implies $\rho'(\beta^* : w) > 0$, all w, so that, by convexity of ρ, if β^* solves the regression quantile problem then it is the unique solution. This completes the proof.

Since nonuniqueness requires $\bar{z}$ to be on the boundary of a convex hull of observations, a dithering of the design values will move $\bar{z}$ away from the boundary and will insure a unique extreme regression quantile.

References

[1] G.W. Bassett and R.W. Koenker, An empirical quantile function for linear modles with i.i.d. errors, *Journal of the American Statistical Association* **77** (1982) 407–425.

[2] R. Koenker and G. Bassett, Regression quantiles, *Econometrica* **46** (1) (1978) 33–50.

[3] R. Koenker and d'Orey, "Computing regression quantiles, *Applied Statistics* (forthcoming).

[4] S. Portnoy, Using regression fractiles to identify outliers, in: Y. Dodge (Ed.), *Statistical Data Analysis based on the L_1-Norm and Related Methods* (North-Holland, Amsterdam, 1987).

[5] P. Rousseew, Least median of squares regression, *Journal of the American Statistical Association* (388) (1984) 871–880.

[6] A.F. Siegel, Robust regression using repeated medians, *Biometrika* **49** (1) (1982) 242–244.