Correction to Censored Regression Quantiles by S. Portnoy, 98 (2003), 1001–1012

Tereza NEOCLEOUS, Karlien VANDEN BRANDEN, and Stephen PORTNOY

Portnoy (2003) presented an approach to the analysis of censored survival data based on a novel computational method for censored regression quantiles. A theorem on asymptotics was given, but in the course of thesis research, the first two authors found a major error in the proof. We have been unable to fix this proof as presented. However, a closely related "grid" algorithm is now our default method, and here we present correct results providing consistency at root-n rate for this grid algorithm. (Details of the proof are rather complicated and have been given in Neocleous 2005 and Vanden Branden 2005.)

Portnoy [2003, eqs. (5) and (6)] gave a general model for censored quantile regression, where the regression coefficients are $\{\beta(\tau)\}$ mapping $0 \le \tau \le 1$ to $\Re^p$ as a vector of regression coefficients such that $x_i'\beta(\tau)$ gives the τth conditional quantile of the response Y_i given the explanatory variables x_i. The censoring times and censoring indicators are denoted by $\{(C_i, \Delta_i): i = 1, \dots, n\}$. The distribution of $\{(Y_i, C_i, \Delta_i)\}$ remains as given by Portnoy (2003).

To define the "grid" algorithm, let $\epsilon > 0$ be given and define a grid of τ-values, $\epsilon \le t_1 < t_2 < \cdots < t_M \le 1 - \epsilon$. In addition, define the parameters along the grid,

$$\beta_k = \beta(t_k), \qquad k = 1, \dots, M, \qquad \text{and}$$
$$\beta = (\beta_1, \dots, \beta_M) \in \Re^{Mp}.$$

Following Portnoy (2003), we assume that the initial regression quantile, $\hat\beta_1$ at $\tau = t_1$, is below all censored points and can be computed by the usual (uncensored) regression quantile algorithm. We define $\hat\beta_k$ inductively as follows: Given $\hat\beta_\ell$ for $\ell \le k$ [and thus given weights $\hat w_i$ of the form $(\tau - t_\ell)/(1 - t_\ell)$], the regression quantile, $\hat\beta_{k+1}$ at t_{k+1} is obtained by minimizing the weighted and censored form of the objective function given by eq. (12) of Portnoy (2003). Weights for censored observations that are crossed in moving from t_k to t_{k+1} [i.e., for which $\tilde Y_i > x_i'\hat\beta(t_k)$ but $\tilde Y_i \le x_i'\hat\beta(t_{k+1})$] are defined with t_k replacing t_ℓ in $\hat w_i$. The algorithm stops at the last grid point t_M or ends at t_e when only censored observations remain above $x'\hat\beta(t_e)$. (See Debruyne and Hubert 2004; Vanden Branden 2005; Neocleous 2005 for some further details and applications of this algorithm.)

Because $\beta \in \Re^{Mp}$ and $M \to \infty$, we need asymptotic results for increasing dimension (sieve methods). An appropriate result providing consistency essentially at rate $n^{-1/3}$ was given by He and Shao (2000). Their hypotheses were verified by Vanden Branden (2005) under the following conditions:

(FY) The random response Y_i has density f_{Y_i} (conditional on x_i) such that there exist constants $a > 0$, $b < \infty$, and c such that

$$a \le f_{Y_i}(y) \le b, \qquad \left|f'_{Y_i}(y)\right| \le c \tag{1}$$

uniformly for $\epsilon \le F_{Y_i}(y) \le 1 - \epsilon$ and uniformly in $i = 1, 2, \dots, n$.

(X) There exists a constant D such that $\|x_i\| \le D$ uniformly in $i = 1, 2, \dots, n$. We furthermore assume that $x_i \in \Re^p$ with dimension p fixed.

(XX) $\frac{1}{n}\sum_{i=1}^n x_i x_i'$ is strictly positive definite.

(F) Let $\tilde F_{Y_i}(u)$ and $\tilde F_{C_i}(v)$ be the distribution functions for $I(Y_i \le C_i)Y_i$ and for $I(Y_i > C_i)C_i$ conditional on x_i and on Δ_i. The densities $\tilde f_{Y_i}$ and $\tilde f_{C_i}$ satisfy

$$a_1 \le \tilde f_{Y_i}(u) \le b_1, \qquad \left|\tilde f'_{Y_i}(u)\right| \le c_1$$

and

$$a_2 \le \tilde f_{C_i}(v) \le b_2, \qquad \left|\tilde f'_{C_i}(v)\right| \le c_2$$

uniformly for $\epsilon \le F_{Y_i}(y) \le 1 - \epsilon$ and uniformly in $i = 1, 2, \dots, n$ for some constants a_j, b_j, and c_j for $j = 1, 2$.

Under these conditions, we then have the following result.

Theorem 1. Let $\hat\beta$ be the censored regression quantile estimator on a grid $\epsilon \le t_1 < t_2 < \cdots < t_M \le 1 - \epsilon$, where $t_{k+1} - t_k = g_n$ satisfies $g_n = n^{-1/3}/a_n$ with $a_n \to \infty$. Then, under the foregoing conditions and for the norm in $\Re^{Mp}$,

$$\|\hat\beta - \beta\|^2 = \mathcal{O}_p\left(\frac{1}{g_n n}\right). \tag{2}$$

The rather complicated proof was given by Vanden Branden (2005). The most difficult part involves the Hessian matrix, D, for the associated gradient conditions (in $\Re^{Mp}$). Fortunately, the facts that the gradient conditions for $\hat\beta(t_k)$ depend on $\{\hat\beta(t_\ell): \ell \ne k\}$ only through the weights and that these weights are defined inductively mean that D is a triangular matrix where the $p \times p$ diagonal blocks dominate the matrix asymptotically. The $n^{-1/3}/a_n$ rate arises from the He–Shao result and the need for the estimated quantile functions to be sufficiently close to the true ones so that they are strictly monotonic along the grid.

Theorem 1 and a complicated chaining argument provide the following root-n rate of consistency at all fixed grid points (see Neocleous 2005 for details).

Theorem 2. Under the foregoing conditions, uniformly for t_k in the grid,

$$\|\hat\beta(t_k) - \beta(t_k)\| = \mathcal{O}_p\left(n^{-1/2}\right). \tag{3}$$

This research was carried out while T. Neocleous was a doctoral student in the Statistics Department, University of Illinois at Urbana–Champaign, K. Vanden Branden was a doctoral student in the Mathematics Department, Katholieke Universiteit Leuven, Belgium, and S. Portnoy was Professor, Statistics Department, University of Illinois at Urbana–Champaign, and Visiting "Francqui" Professor at Université Libre de Bruxelles and Katholieke Universiteit, Leuven, Belgium (E-mail: *sportnoy@uiuc.edu*).

Journal of the American Statistical Association
June 2006, Vol. 101, No. 474, Correction
DOI 10.1198/016214506000000087

[Received June 2005. Revised July 2005.]

REFERENCES

Debruyne, M., and Hubert, M. (2004), "Robust Regression Quantiles With Censored Data," in *Proceedings in Computational Statistics*, ed. J. Antoch, Heidelberg: Springer-Verlag, pp. 887–893.

He, X., and Shao, Q. (2000), "On Parameters of Increasing Dimension," *Journal of Multivariate Analysis*, 73, 120–135.

Neocleous, T. (2005), "A Partially Linear Model for Censored Regression Quantiles," unpublished doctoral thesis, University of Illinois at Urbana–Champaign, Dept. of Statistics.

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of American Statistical Association*, 98, 1001–1012.

Vanden Branden, K. (2005), "Robust Methods for High-Dimensional Data and a Theoretical Study of Depth-Related Estimators," unpublished doctoral thesis, Katholieke Universiteit Leuven, Dept. of Mathematics.