Fitzenberger, Bernd

Working Paper

A note on estimating censored quantile regressions

Diskussionspapier, No. 14

Provided in Cooperation with:
Department of Economics, University of Konstanz

Forschungsschwerpunkt
"Internationale Arbeitsmarktforschung"

Center for International Labor Economics
(CILE)

Fakultät für Wirtschaftswissenschaften und Statistik
Universität Konstanz

Bernd Fitzenberger

A Note on Estimating Censored Quantile Regressions

Postfach 5560 D 139
78434 Konstanz
Deutschland / Germany

Diskussionspapier
14 – 1994

A Note on Estimating Censored Quantile Regressions

Bernd Fitzenberger

Diskussionspapier

Nr. 14

Juni 1994

A Note on Estimating Censored Quantile Regressions

Bernd Fitzenberger, Universität Konstanz[1]

May 1994

Abstract

This note is concerned with estimating censored quantile regressions (CQR). As its major contribution, a new algorithm, called BRCENS, is developed as an adaption of the Barrodale–Roberts algorithm for the standard quantile regression problem. In a subsequent simulation study, BRCENS performs well in comparison with the iterative linear programming algorithm (ILPA) suggested recently by Buchinsky. In the theoretical analysis, this note generalizes the asymptotic theory for estimating CQR to the case with observation specific censoring points and with fairly arbitrary non–stationarity and dependency in the data. Building on the interpolation property of the coefficient estimate, the ILPA is shown to suffer from some theoretical inconsistencies.

Keywords: Censored Quantile Regression, Consistency, Asymptotic Normality, Interpolation Property, Algorithms.

[1]This research benefitted from discussions with Thomas E. MaCurdy, Peter Winker, and Frank Wolak. All errors are my sole responsibility. Address: Bernd Fitzenberger, Fakultät für Wirtschaftswissenschaften und Statistik, Universität Konstanz, Postfach 5560 <D 139>, 78434 Konstanz, Germany. E-mail: fitzenbe@sonne.wiwi.uni-konstanz.de

1 Introduction

While censored quantile regressions (CQR) have been recognized as an interesting robust estimation approach for the censored regression problem, their application in applied research seems to have been limited by the lack of an efficient algorithm. Also censored quantile regressions have mostly been considered in the case with a common censoring point for all observations. In this note, the consistency and asymptotic normality of the CQR estimator is established in a time series setup without the assumption of fixed censoring points and with fairly arbitrary forms of heteroskedasticity and autocorrelation of the data. Building on a theoretical characterization of the censored quantile regression estimate, the new algorithm BRCENS for estimating censored quantile regressions is developed and it is contrasted with the iterative linear programming algorithm (ILPA) suggested recently by Buchinsky (1994). In light of the theoretical results, ILPA is shown to exhibit some theoretical inconsistencies. The limited set of simulation results is quite encouraging for recommending the use of BRCENS.

Barrodale and Roberts (1973) and (1974) developed an efficient algorithm (BRA) for the problem of least absolute deviation (median) regression. Quantile regressions were introduced by Koenker and Bassett (1978). A simple modification of BRA allows for estimating general quantile regressions, as indicated in Koenker and d'Orey (1987). Powell (1984) and (1986) extended the concept of quantile regressions to the censored regression problem with fixed censoring points for all observations. He established that CQR estimation leads to a consistent and asymptotically normal estimator even with a restricted form of heteroskedasticity. Womersley (1986) provided a theoretical characterization of the estimate in the censored median regression problem and he proposed a finite direct descent method to calculate the estimator. Dielman (1992) showed that simplex based algorithms, which are modified versions of the BRA, appear to be most efficient for median regression problems. Buchinsky (1994) suggested a new algorithm for the CQR problem, ILPA, and used it to analyze the U.S. wage structure, to my knowledge the first empirical application of CQR. The algorithm BRCENS, developed in this note, was first applied in Fitzenberger et al. (1994) to study the German wage structure.

This paper proceeds as follows. Section 2 establishes the asymptotic theory of the censored quantile regression estimator in a time series setup without the assumption of fixed censoring points and with fairly arbitrary forms of heteroskedasticity and autocorrelation of the data. In addition, a theoretical characterization of the estimate is provided. Section 3 develops the new algorithm BRCENS and describes ILPA. Two theoretical inconsistencies of IPLA are shown and a modified algorithm is suggested, called modified iterative linear programming algorithm, MILPA. Section 4 describes a small simulation study which contrasts the relative performance of BRCENS, ILPA, and MILPA. Section 5 concludes. The appendix provides the proof of the asymptotic results in section 2 and a description of the formal implementation of BRCENS.

2 Theoretical Results

This section discusses theoretical aspects of estimating CQRs. In section 2.1 describes the CQR estimation problem and consistency and asymptotic normality of the censored quantile regression estimator (CQRE) are established for a more general setup than previously considered in the literature. Section 2.2 presents a characterization of the CQRE with respect to the interpolation of data points by the estimated regression hyperplane.

2.1 Asymptotic Theory

This section describes the method of estimating CQRs and a new asymptotic result is established which allows for fairly arbitrary forms of heteroskedasticity and autocorrelation in the data and for observation specific censoring points.

Before describing the method of CQRs, the following notation is introduced. Within a linear regression context, let the dependent variable be the $T \times 1$ vector, $y = (y_1, ..., y_T)$, the design matrix be the $T \times k$ matrix $X = (x_1, ..., x_T)'$, with $x_t = (x_{t1}, ..., x_{tk})$, the $T \times 1$ vector of observation specific censoring values be $yc = (yc_1, ..., yc_T)$, the $T \times 1$ vector of disturbances be $\epsilon = (\epsilon_1, ..., \epsilon_T)'$ and the $k \times 1$ parameter vector be β_0, for $t = 1, 2, ..., T$ and $T = 1, 2, ...$. Let the θ weighted sign function $sgn_\theta(\epsilon_t)$ be defined as

$$sgn_\theta(\epsilon_t) \equiv \theta I(\epsilon_t > 0) - (1-\theta) I(\epsilon_t < 0)$$

with $\theta \in (0,1)$ and I(.) denotes the indicator function.

Quantile Regressions were introduced by Koenker and Bassett (1978). Powell (1984) and (1986) proposed the method of censored least absolute deviation (median) regressions and CQRs, respectively, as a robust estimation approach for the censored regression model. The CQR estimation problem is to minimize for a given θ over β the piecewise linear function defined by

$$(1) \qquad \widehat{\beta_T} \in \underset{\beta}{\operatorname{argmin}} \ \frac{1}{T} \sum_{t=1}^{T} sgn_\theta(y_t - min[x_t'\beta, yc_t])\,(y_t - min[x_t'\beta, yc_t])$$

With the assumption of a common censoring point, $yc_1 = yc_2 = ... = yc_T$, for all observations, Powell established that unlike the standard Tobit maximum likelihood estimator, the CQRE provides an estimator which is consistent and asymptotically normally distributed without an assumption that the errors are normally distributed or homoskedastic. The following theorem extends upon Powell's results in two respects. First, observation specific censoring points are allowed for and second, it is established that the CQRE is consistent and asymptotically normal for fairly arbitrary non-stationarity and dependency in the data.

Theorem 1: (Asymptotic Theory of CQRE)
Suppose for some $\theta \in (0,1)$

(i) $y_t^* = x_t'\beta_0 + \epsilon_t$ and $y_t = min[y_t^*, yc_t]$.

(ii) (a) $\{x_t', \epsilon_t, yc_t\}_{t=1,...,T}$ is a strong mixing sequence of size $-2r/(r-2)$ where $r > 2$.

(b) The $\{\epsilon_t\}$'s have some distribution on $(-\infty \times ... \times \infty)$ and let their joint distribution function be absolutely continuous on the cartesian product of the intervals $B_{t,d} = \{\epsilon_t : -d \leq \epsilon_t \leq d\}$ for all t=1,...,T and some $d > 0$.

(c) Let $F_t = \sigma(..., x_{t-1}, x_t, ..., yc_{t-1}, yc_t)$ be the σ–algebra generated by the sequence of x_t's and yc_t's until t, and let the distribution of ϵ_t conditional on F_t have a density $f_t(\epsilon_t) \equiv f_t(\epsilon_t \mid F_t)$ for $-d \leq \epsilon_t \leq d$, where $f_t(\epsilon_t)$ is a.s. Lipschitz continuous in ϵ_t uniformly in t with a Lipschitz constant $0 < L_0 < \infty$. There exist f_L and f_U such that $0 < f_L < f_t(\epsilon_t) < f_U < \infty$ a.s. for $-d \leq \epsilon_t \leq d$ and all t=1,...,T.

(iii) (a) $E\{sgn_\theta(\epsilon_t) \mid F_t\} = 0$ a.s. for all t=1,...,T.

(b) $J_T = Var\{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t \, I(x_t'\beta_0 < yc_t) \, sgn_\theta(\epsilon_t)\}$ is uniformly positive definite in T.

(iv) (a) $E \mid x_{ti} \mid^{r'} < \delta$ for all t=1,...,T, i=1,...,k and some δ, r', and η such that $0 < \delta < \infty$, $r' = max(3+\eta, r)$ and $\eta > 0$.

(b) $M_T = E\{\frac{1}{T} \sum_{t=1}^{T} I(x_t'\beta_0 < yc_t - \zeta) \, x_t x_t'\}$ is uniformly positive definite in T for $\zeta > 0$.

(c) $L_T = E\{\frac{1}{T} \sum_{t=1}^{T} f_t(0) \, I(x_t'\beta_0 < yc_t) \, x_t x_t'\}$ is uniformly positive definite in T.

(d) $G_t(z, \beta, \rho) = E\{I(\mid x_t'\beta - yc_t \mid \leq \| x_t \| z) \| x_t \|^\rho\}$ is o(z) for z near zero, β near β_0, and ρ=0,1,2, uniformly in t, i.e. $G_t(z, \beta, \rho) \leq K_1 z$ if $0 \leq z < \xi_0$ and $\| \beta - \beta_0 \| < \xi_0$ for some K_1 and ξ_0.

(v) B is a compact set in $\Re^k$ and $\beta_0 \in$B.

Then a.) $\widehat{\beta_T} \in \underset{\beta \in B}{\text{argmin}} \; \frac{1}{T} \sum_{t=1}^{T} sgn_\theta(y_t - min[x_t'\beta, yc_t]) \, (y_t - min[x_t'\beta, yc_t]) \rightarrow \beta_0$ a.s. for $T \rightarrow \infty$.

b.) $D_T^{-1/2} \sqrt{T} \, (\widehat{\beta_T} - \beta_0) \overset{D}{\rightarrow} N(0, I_k)$, where $D_T = L_T^{-1} J_T L_T^{-1}$ and I_k is the $k \times k$ identity matrix.

The proof of theorem 1 can be found in Appendix 1. The result is related to recent work on the asymptotic behavior of standard quantile regression estimators with non–stationary error terms in Portnoy (1991), Weiss (1991), and Fitzenberger (1993). The treatment here is closely related to Powell (1984), (1986) and Fitzenberger (1993).

The CQRE is not necessarily unique. Theorem 1 establishes for every sequence of minimizers given by equation 1, $\widehat{\beta_T}$, that the consistency and asymptotic normality results hold.

The following short discussion is concerned with the role of certain regularity conditions. Condition (iii)(a) states that the conditional θ–quantile of ϵ_t is zero, which in addition to condition (iv)(b) guaranties consistency of the CQRE. In analogy to Fitzenberger (1993) for the case of standard quantile regressions, this assumption allows only for a form of heteroskedasticity, which leaves the θ–quantile unchanged. The approach differs from Portnoy (1991), who uses a less restrictive assumption and identifies an asymptotic bias of the coefficient estimator.[2] Strong mixing of the data and condition (iii) basically imply asymptotic normality of $\sum_{t=1}^{T} x_t I(x_t'\beta_0 < yc_t) sgn_\theta(\epsilon_t)$. The extension of Huber's (1967) Lemma 3 to the setup in this note allows for asymptotically interchanging $T^{-1/2}\sum_{t=1}^{T} x_t I(x_t'\beta_0 < yc_t) sgn_\theta(\epsilon_t)$ and $T^{-1/2}\sum_{t=1}^{T} E\{x_t I(x_t'\widehat{\beta_T} < yc_t) sgn_\theta(\epsilon_t - x_t'(\widehat{\beta_T} - \beta_0))\}$. The latter expression is differentiable in the coefficient estimate and its Jacobian at β_0 equals the matrix $-L_T$. In order to prove consistency, it is necessary that the regression is "asymptotically identified", condition (iv)(b), by the set of uncensored points. For asymptotic normality, the conditional quantile function $min[x_t'\beta, yc_t]$ of y_t must be well behaved when β is close to β_0. Thus, condition (iv)(d) rules out sequences of values (x_t, yc_t) with $x_t'\beta_0 = yc_t$ with positive frequency. Conditions (iv)(b) and (d) are adapted versions of conditions R.1 and R.2 in Powell (1984) to the setup in this note.

2.2 Interpolation Property

This section provides a characterization of the CQRE analogous to the characterization of the standard quantile regression estimator in Koenker and Bassett (1978), theorem 3.1. This characterization provides the basis for the algorithm BRCENS developed in section 3.1.

Since the minimization problem described in equation 1 does not necessarily have a unique solution, let the set of optimizers $\widehat{B_T}$ of the distance function $Q_T(\beta)$ be defined by

$$\widehat{B_T} \equiv \underset{\beta}{\text{argmin}}\; Q_T(\beta) \equiv \sum_{t=1}^{T} sgn_\theta(y_t - min[x_t'\beta, yc_t])\,(y_t - min[x_t'\beta, yc_t]) \,.$$

The following Theorem, characterizing the elements of $\widehat{B_T}$, can be proved. Most of the following result for censored median regressions can be found in Womersley (1985), Theorem 1.

Theorem 2: (Interpolation Property of CQRE's) If the design matrix X has rank k', every element in the set of minimizers, $\widehat{B_T}$, is a convex combinations of solutions in $\widehat{B_T}$ interpolating k' data points, i.e., $\forall \widehat{\beta_T} \in \widehat{B_T}\ \exists\ j$ solutions $\beta_{T,1}, ..., \beta_{T,j} \in \widehat{B_T}$ and associated scalars $\lambda_1, ..., \lambda_j$ with $0 \leq \lambda_i \leq 1$, $i = 1, ..., j$,

[2] In his analysis of standard quantile regressions with non–stationary, dependent errors, Portnoy assumes the errors to be "close to m–dependent" and, put into the notation of this note, he only requires that $\sum_{t=1}^{T} x_t sgn_\theta(\epsilon_t)$ grows at a rate between $T^{1/2}$ and $T^{3/4}$. With $\sqrt{T}$ asymptotic normality, this introduces an asymptotic bias.

and $\sum_{i=1}^{j} \lambda_i = 1$, such that $\widehat{\beta_T}$ is a convex combination of the $\beta_{T,1}, ..., \beta_{T,j}$, i.e., $\widehat{\beta_T} = \sum_{i=1}^{j} \lambda_i \beta_{T,i}$, and each $\beta_{T,i}$ interpolates at least k' data points such that for $i = 1, ..., j \; \exists \; k'$ observations $\{(y_{t,i,1}, x_{t,i,1}), ..., (y_{t,i,k'}, x_{t,i,k'})\}$ with

(IP) $y_{t,i,l} = x'_{t,i,l}\beta_{T,i}$ for $l = 1, ..., k'$ and the rank of $(x_{t,i,1}, ..., x_{t,i,k'})'$ equals k'.

Note that theorem 2 does not imply that $\widehat{B_T}$ is a convex set as in the case of standard quantile regressions.[3] Womersley (1986), Theorem 1, establishes for the case of censored median regressions that there exists a global optimum with the interpolation property (IP). Womersley's approach can be easily adapted to the CQR problem.

In the following, a heuristic argument is given for theorem 2, which forms the basis for developing the new algorithm BRCENS in section 3.1. The argument focuses on the directional derivative of the distance funtion, which plays a central role in section 3.

Since the distance function $Q_T(\beta)$ is piecewise linear and bounded from below, there is at least one global minimum. If the global minimum is unique, it must be located at a kink of the distance function. In general, all global minima can be represented as convex combinations of minima at kinks, but the set of minima, $\widehat{B_T}$, is not a convex set itself.

The directional derivative of $Q_T(\beta)$ for some $w \in \Re^k$ is given by

$$
\begin{aligned}
H'_T(\beta, w) &= \lim_{a \downarrow 0} \frac{Q_T(\beta + aw) - Q_T(\beta)}{a} \\
&= \sum_{t=1}^{T} [I(x'_t\beta < yc_t)\{-sgn_\theta(y_t - x'_t\beta) - I(x'_t\beta = y_t) sgn_\theta(-x'_t w)\} \\
&\quad + I(x'_t\beta = yc_t)\{(1-\theta) I(y_t < yc_t, x'_t w < 0) \\
&\qquad\qquad - \theta I(y_t = yc_t, x'_t w < 0)\}] x'_t w
\end{aligned} \tag{2}
$$

This expression allows to identify the kinks of $Q_T(\beta)$. A vector β represents a kink, if the directional derivative $H'_T(\beta, w)$ changes for some w in an arbitrarily small neighborhood of β. At that point, there must be observations with $x'_t\beta = y_t$ or $x'_t\beta = yc_t$. Analogous to standard quantile regressions, where kinks in the objective imply that there must be observations with $x'_t\beta = y_t$,[4] the kinks in the set of minimizers must interpolate k' observations $\{(y_{t,i,1}, yc_{t,i,1}, x_{t,i,1}), ..., (y_{t,i,k'}, yc_{t,i,k'}, x_{t,i,k'})\}$ whereby the rank of $(x_{t,i,1}, ..., x_{t,i,k'})'$ equals k'. If not, there existed a kink and a vector $w \in \Re^k$ and $w \neq 0$ such that $x'_t w = 0$ for all data points interpolated at that kink. Then either (i) $H'_T(\beta, w) < 0$ or $H'_T(\beta, w) > 0$ or (ii) $H'_T(\beta, w) = 0$. If (i) is true, β cannot be an optimum, and if (ii) is true, one can move in direction w without increasing or decreasing the objective function, while all of the data points

[3] See Koenker and Bassett (1978), theorem 3.1.

[4] See Koenker and Bassett (1978).

interpolated at β keep being interpolated. This can be continued until a point is reached, where an additional data point is interpolated, which proves the claim.

At this stage, it remains to be shown that a solution β cannot be optimal if it interpolates a data point at the censoring point, while the observation itself is not censored and the vector of regressors at that point is not in the space spanned by the other interpolated data points. The result can be established easily considering the directional derivatives

$$\begin{aligned} H_T'(\beta - aw, w) &= \textstyle\sum_{t=1}^{T}[I(x_t'\beta < yc_t)\{-sgn_\theta(y_t - x_t'\beta) - I(x_t'\beta = y_t)sgn_\theta(x_t'w)\} \\ &\quad +I(x_t'\beta = yc_t)\{(1-\theta)I(y_t < yc_t, x_t'w > 0) \\ &\qquad -\theta I(y_t = yc_t, x_t'w > 0)\}]x_t'w \end{aligned}$$

and

$$\begin{aligned} H_T'(\beta + aw, w) &= \textstyle\sum_{t=1}^{T}[I(x_t'\beta < yc_t)\{-sgn_\theta(y_t - x_t'\beta) - I(x_t'\beta = y_t)sgn_\theta(-x_t'w)\} \\ &\quad +I(x_t'\beta = yc_t)\{(1-\theta)I(y_t < yc_t, x_t'w < 0) \\ &\qquad -\theta I(y_t = yc_t, x_t'w < 0)\}]x_t'w \end{aligned}$$

where w is chosen that for all other interpolated points, $x_t'w = 0$, and $a > 0$, small.[5] For β to be optimal, it is necessary that

$$H_T'(\beta + aw, w) - H_T'(\beta - aw, w) \geq 0$$

but in the case considered here.

$$H_T'(\beta + aw, w) - H_T'(\beta - aw, w) = \sum_{t=1}^{T} I(x_t'\beta = yc_t)\{I(x_t'w < 0) - I(x_t'w > 0)\}\theta x_t'w$$

which proves to be strictly negative, thus implying that β cannot be optimal.

3 Algorithms for CQR Estimation

This section presents two algorithms for the problem of CQR estimation. The lack of an efficient algorithm was a major obstacle for the use of CQR in applied research. In section 3.1, a new algorithm, called BRCENS, is developed based on the standard Barrodale–Roberts–Algorithm. Section 3.2 discusses the algorithm suggested recently by Buchinsky (1994) and shows that it is theoretically inconsistent. A third algorithm developed by Womersley (1986) is based on a finite direct descent method. It is not considered in this note and to my knowledge it has not yet been used in applied econometric research.[6]

[5] a is chosen such that (i) $y_t \neq x_t'\beta \Longrightarrow y_t \neq x_t'(\beta \pm aw)$, (ii) $yc_t \neq x_t'\beta \Longrightarrow yc_t \neq x_t'(\beta \pm aw)$, (iii) $yc_t > (<) x_t'\beta \Longrightarrow yc_t > (<) x_t'(\beta \pm aw)$, and (iv) $y_t > (<) x_t'\beta \Longrightarrow y_t > (<) x_t'(\beta \pm aw)$.

[6] I am planing to include Womersley's algorithm in the simulation study in section 4. According to Dielman (1992), simplex approaches which are extensions of the Barrodale–Roberts–Algorithm appear to be most efficient for computing least absolute deviation regressions, whereas evidence concerning direct descent methods appears somewhat mixed. This evidence by itself justifies the development of the new algorithm BRCENS, but I am hoping to receive the program used in Womerley's example, Womersley (1986), section 5, in order to compare the algorithms in practice.

3.1 BRCENS

This section describes the main ideas used to implement a new algorithm to estimate CQR's called BRCENS.[7] I develop BRCENS by adapting the Barrodale–Roberts–Algorithm (BRA) to CQR's.[8]

Since the quantile regression problem has a linear programming structure, it can be solved by a simplex algorithm. Barrodale and Roberts noted that with the special structure of the least absolute deviation (LAD) problem a standard simplex algorithm was quite inefficient. The BRA uses a condensed form of the simplex tableau with only $T+1$ rows and $k+1$ columns, where T is the number of observations and k the number of regressors. The core of the tableau consists of the representation of the nonbasic variables, i.e. interpolated data points or coefficients being zero, in terms of the basis variables, i.e. data points not necessarily being interpolated or coefficients not necessarily being zero.[9] BRA significantly reduces the number of simplex transformations by only considering transformations towards solutions where k' observations are interpolated, where k' is the rank of the design matrix. In addition, before performing a simplex transformation BRA bypasses several kinks, i.e., solutions where at least k' observations are interpolated, until the "marginal cost" reduction becomes non–positive, i.e., until the objective cannot be reduced further. After a simplex transformation, the "marginal cost" reductions are recalculated. BRA stops when all "marginal cost" reductions are non–positive. If all of them are strictly negative, a unique global optimum is reached.

In light of the interpolation property of CQRE presented in section 2.2, I keep the same simplex setup when adapting the BRA. In what follows, the heuristics of BRCENS are described. A short formal description of the simplex setup and my adaption of the "marginal cost" reductions are given in the appendix. In the first stage, BRCENS starts with the coefficient vector at zero. By subsequent simplex iterations, as many coefficients are brought into the basis as the rank k' of the design matrix allows for. If $k' = k$, the latter being the number of regressors, then all coefficients are brought into the basis. The first stage ends with k' interpolated data points as nonbasic variables, each one represented by a column of the tableau. The set of nonbasic variables – data points and coefficients – define the current solution. In the second stage, BRCENS continues by interchanging data points in the basis with nonbasic data points. The algorithm stops, when all "marginal cost" reductions of all nonbasic variables are non–positive indicating that an interchange of a nonbasic variable with any data point in the basis would not lead to a reduction of the objective.

My modification of BRA concerns the calculation of the "marginal cost" reductions. Both for stage 1 and 2, "marginal cost" reductions have to be calculated for each nonbasic variable. The "marginal cost" reduction for a nonbasic variable is the marginal change in the objective when the variable changes from being zero.

[7] Barrodale–Roberts–Algorithm for Censored Quantile Regressions.

[8] The BRA is suggested in Barrodale and Roberts (1973). A FORTRAN source code is provided in Barrodale and Roberts (1974). Koenker and D'Orey (1987) adapt the algorithm to the quantile regression case.

[9] For further details, cf. Barrodale–Roberts (1973)

Since there are k nonbasic variables, this change defines a one–dimensional search direction starting from the current solution to look for a reduction of the objective. Thus, the "marginal cost" reduction is the negative of the directional derivative in the search direction, presented in equation 2, times some positive scalar multiple. The simplex approach allows for an efficient calculation of the "marginal cost" reductions. BRCENS has to take account of the possibility that, first, for both nonbasic and basis variables the current solution could interpolate the censoring points, which implies that there is only an effect on the objective, if the search direction is such that the predicted value on the regression hyperplane moves below the censoring point, and that, second, for a data point in the basis the predicted value could lie above the censoring point, thus not implying any contribution to the "marginal cost" reduction when moving into the search direction.

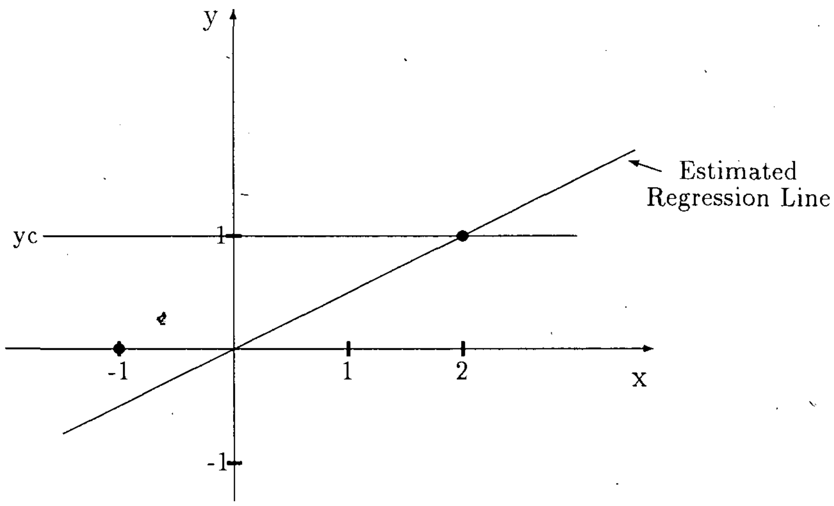
Analogous to BRA, BRCENS considers to move the nonbasic variable with the highest "marginal cost" reduction into the basis. This defines the search direction to find a new solution and the algorithm involves to move into that direction until the "marginal cost" reduction becomes non–positive. At this new point, the objective exhibits a local minimum along the search direction.

While moving along the search direction, the "marginal cost" reduction can only change at points where the regression hyperplane interpolates y_t or yc_t for some observation in the basis not interpolated before. Then the "marginal cost" reduction is updated accordingly until it becomes non–positive. One of the data points in the basis for which y_t is interpolated here now leaves the basis and is replaced by the nonbasic variable defining the search direction. The arguments in section 2.2 guarantee that the data point leaving the basis never has to be a point where the new solution interpolates the censoring point and the observation is uncensored at the same time. The interchange of variables is performed by a standard simplex step, which updates the representation of the nonbasic variables in terms of the basis variables and now the "marginal cost" reductions have to be recalculated for the nonbasic variables. The algorithm terminates when all "marginal cost" reductions become non–positive. If they are all strictly negative, a strict local minimum of the objective is guaranteed, analogous to the standard quantile regression case. In contrast to the latter case, however, due to the nonconvexity of the CQR problem, this is not necessarily a global minimum.

3.2 ILPA and MILPA

This section is concerned with the iterative linear programming algorithm (ILPA) used in Buchinsky (1994) who estimates CQR's in order to study the structure of wages in the United States. Further an extension of ILPA, the modified iterative linear programming algorithm (MILPA), is suggested. ILPA is based on iterating a sequence of standard quantile regressions. Unfortunately, this simple algorithm is based on two ideas, which prove to be inconsistent from a theoretical perspective. Nevertheless, the algorithm has a lot of intuitive appeal, and MILPA is suggested to provide a remedy against one of the problems. Before going into details, it should be emphasized that the issue is not caused by the fact, that Buchinsky treats only

Figure 1: Example for which Censored Median Regression Interpolates a Censoring Point[a]



(a) See main text for further explanations.

the case of a fixed censoring point for all observations.

To begin with, a description of the ILPA follows.[10] ILPA starts with some initial coefficient estimate $\hat{\beta}_0$ and a counter $j = 1$. The following iterative steps are continued until convergence is achieved:

Step 1: For the j^{th} iteration, determine the set M_j of observations with $x_t'\hat{\beta}_j < yc_t$. If $j = 1$ or $M_j \neq M_{j-1}$ then continue with step 2, otherwise terminate and take $\widehat{\beta_T} = \hat{\beta}_{j-1}$ as the CQRE.

Step 2: Calculate $\hat{\beta}_j$ as the standard quantile regression estimate for the set of observations M_j by means of the BRA.[11] Set $j := j + 1$ and repeat step 1.

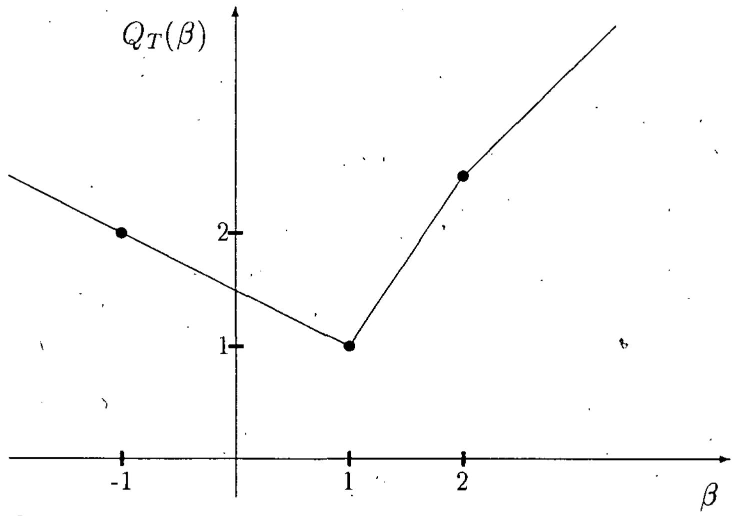
Buchinsky states that ILPA is not guaranteed to converge, but he claims that once convergence is achieved the coefficient estimate represents a local minimum of the problem.

ILPA is based on the following two ideas. First, for an optimal solution $\widehat{\beta_T}$ of the CQR problem, the set of observations for which the predicted value lies on or above the censoring point, i.e. $x_t'\widehat{\beta_T} \geq yc_t$, could have been excluded from the estimation. My analysis in section 2.2 establishes that this claim is only true for observations with $x_t'\widehat{\beta_T} > yc_t$, since the directional derivative, defined in equation 2, depends also on data points with $x_t'\widehat{\beta_T} = yc_t$, albeit in a "one-sided" way. The following simple example provides a case where the estimated regression hyperplane interpolates a censored data point. For a censored median regression, $\theta = 0.5$, with

[10]Cf. Buchinsky (1994), p. 412.

[11]For $\theta \neq 0.5$, by means of the modification of BRA in Koenker and d'Orey (1987).

Figure 2: $Q_T(\beta)$ in Example for which a Fixpoint of ILPA is not a Local Minimizer[a]



(a) See main text for further explanations.

only one regressor, $k = 1$, take a sample of just two observations, $T = 2$, with $(y_1, yc_1, x_{11}) = (0, 1, -1)$ and $(y_2, yc_2, x_{12}) = (1, 1, 2)$. The data are depicted in figure 1 and the estimator in this case is $\widehat{\beta_T} = 1/2$. The estimated censored median regression interpolates the censored data point 2, which ILPA could not take account of. In fact, ILPA does not converge in this case, whatever the starting estimate.

The second idea which ILPA is based on is the claim that, when convergence is reached, the coefficient estimate represents a local minimum of the CQR problem. This represents a second theoretical inconsistency which is illustrated by the following counterexample. For a censored median regression, $\theta = 0.5$, with only one regressor, $k = 1$, take a sample of just two observations, $T = 2$, with $(y_1, yc_1, x_{11}) = (-1, 1, -1)$ and $(y_2, yc_2, x_{12}) = (-1/2, 1, 1/2)$. If ILPA starts with $\hat{\beta}_0 = -1$, convergence is achieved after two iterations with $\hat{\beta}_0 = \hat{\beta}_1 = \hat{\beta}_2$, thus showing that $\hat{\beta}_0 = -1$ is a fixed point, which is not a local minimum of the problem. Figure 2 depicts the distance function, $Q_T(\beta)$, for the example.

ILPA can be slightly modified such that it allows for the case that the regression interpolates a censored data point. I call this modification the modified iterative linear programming algorithm (MILPA), which coincides with ILPA except that in step 1 "$<$" is replaced by "$\leq$" when defining the set M_j. A problem with MILPA is, of course, that it takes account of the directional derivative in "both directions" from a censoring point. Despite the issues discussed here, ILPA and MILPA have some intuitive appeal, especially when all observations have a common censoring point. It is likely that they converge properly in samples with a moderate degree of censoring where the optimal regression lies below the censoring points for all observations, since in that case, there are no observations interpolated at the censoring points, i.e., such points do not influence the directional derivative.

This intuition is confirmed by the results of the subsequent simulation study, which compares ILPA and MILPA with the algorithm BRCENS developed in section 3.1.

4 Simulation Study

This section presents the results of a small scale simulation study comparing the performance of the three algorithms BRCENS, ILPA, and MILPA described in section 3. For various data generating processes, I analyze first, how often each algorithm converges, second, how often it reaches the true optimum of the problem, and third, conditional on convergence, how is the relative performance for each pair of the three algorithms.

For the simulation study, 200 random samples are drawn for each of the following data generating processes (DGP). Each random sample consists of $T = 100$ observations. The estimation problem is to estimate a censored median regression as a function of one regressor allowing for an intercept, i.e. $k = 2$. The following DGP's, (A) to (L), differ by the assumptions on how the samples of $(y_t, yc_t, x_{t2})_{t=1,\ldots,100}$ are generated.

(A) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t = Const$ and $y_t = min(yc_t, \epsilon_t)$.

(B) $x_{t2} = -10.0 + 0.2t$, $\epsilon_t \sim N(0,1)$, $yc_t = Const$ and $y_t = min(yc_t, \epsilon_t)$.

(C) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t = Const$ and $y_t = min(yc_t, 0.5+0.5x_{t2}+\epsilon_t)$.

(D) $x_{t2} = -10.0+0.2t$, $\epsilon_t \sim N(0,1)$, $yc_t = Const$ and $y_t = min(yc_t, 0.5+0.5x_{t2}+\epsilon_t)$.

(E) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t \sim N(Const, 1)$ and $y_t = min(yc_t, \epsilon_t)$.

(F) $x_{t2} = -10.0 + 0.2t$, $\epsilon_t \sim N(0,1)$, $yc_t \sim N(Const, 1)$ and $y_t = min(yc_t, \epsilon_t)$.

(G) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t \sim N(Const, 1)$ and $y_t = min(yc_t, 0.5 + 0.5x_{t2} + \epsilon_t)$.

(H) $x_{t2} = -10.0+0.2t$, $\epsilon_t \sim N(0,1)$, $yc_t \sim N(Const, 1)$ and $y_t = min(yc_t, 0.5+0.5x_{t2} + \epsilon_t)$.

(I) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t = Const - 0.5I(1 \leq t \leq 20) - 0.25I(21 \leq t \leq 40) + 0.25I(61 \leq t \leq 80) + 0.5I(81 \leq t \leq 100)$ and $y_t = min(yc_t, \epsilon_t)$.

(J) $x_{t2} = -10.0+0.2t$, $\epsilon_t \sim N(0,1)$, $yc_t = Const-0.5I(1 \leq t \leq 20)-0.25I(21 \leq t \leq 40) + 0.25I(61 \leq t \leq 80) + 0.5I(81 \leq t \leq 100)$ and $y_t = min(yc_t, \epsilon_t)$.

(K) $x_{t2} \sim N(0,1)$, $\epsilon_t \sim N(0,1)$, $yc_t = Const - 0.5I(1 \leq t \leq 20) - 0.25I(21 \leq t \leq 40) + 0.25I(61 \leq t \leq 80) + 0.5I(81 \leq t \leq 100)$ and $y_t = min(yc_t, 0.5 + 0.5x_{t2} + \epsilon_t)$.

Table 1: Results of Simulation Study: Average Share of Censored Observations In Random Samples For various Data Generating Processes (DGP) - In Percent[a]

DGP	Const = 1.0	Const = 0.5	Const = 0.0
(A)	16.0	30.9	48.7
(B)	16.0	31.0	49.1
(C)	32.9	49.1	66.2
(D)	45.6	50.5	55.5
(E)	24.1	36.2	49.6
(F)	24.1	36.2	49.6
(G)	36.8	49.5	62.2
(H)	45.5	50.6	55.7
(I)	17.5	32.1	49.8
(J)	17.5	32.0	49.8
(K)	33.8	49.2	65.6
(L)	44.9	50.5	56.2

(a) See main text for further explanations.

(L) $x_{t2} = -10.0+0.2t\,, \epsilon_t \sim N(0,1)\,, yc_t = Const-0.5I(1 \le t \le 20)-0.25I(21 \le t \le 40) + 0.25I(61 \le t \le 80) + 0.5I(81 \le t \le 100)$ and $y_t = min(yc_t, 0.5 + 0.5x_{t2} + \epsilon_t)$.

The shift variable $Const = 1.0, 0.5, 0.0$ is used to allow for varying degrees of censoring. The lower is $Const$, the higher the average degree of censoring in the random sample. Table 1 contains the average shares of censored observations for the various DGP's. DPG's (A) to (D) assume a fixed censoring point for all observations in the random samples, whereas DPG's (E) to (L) allow for observation specific censoring points. (E) to (H) let yc_t being normally distributed, i.e. in a random sample there are basically no two observations with the same censoring point. (I) to (L) let yc_t take five different values with the frequency being the same. The x_{t2}'s are either normally distributed with expected value zero and variance one, or they take a fixed sequence of values, $(-9.8, -9.6, \ldots, 9.6, 9.8, 10.0)$. The "true" median regression assumes that both intercept and slope are either zero or 0.5.

Table 2 presents the absolute frequencies for various DGP's, that the three algorithms converged, part (i), and that each of them achieves the optimum of the objective function $Q_T(\beta)$, part (ii). From part (i), it is evident that BRCENS always converges, i.e. always yielded an optimal solution with the "marginal cost" reductions being strictly negative for all nonbasic variables in the final simplex tableau. Also MILPA almost always converges, the exceptions are DPG (A) and (B) with one and two samples, respectively, without convergence. For MILPA as well as for ILPA, the algorithm is terminated after 20 iterations, if no convergence is achieved: Experimental evidence showed that this is large enough to determine whether the algorihms converged. The picture is very different for ILPA, which exhibites a lack of convergence the higher the degree of censoring. Also the convergence rate is

Table 2: Results of Simulation Study: For various Data Generating Processes (DGP), Absolute Frequencies that (i) Algorithms Converged and (ii) Achieved Optimum[a]

(i) Convergence Frequency Among 200 Random Samples									
DGP	Const = 1.0			Const = 0.5			Const = 0.0		
	ILPA	MILPA	BRCENS	ILPA	MILPA	BRCENS	ILPA	MILPA	BRCENS
(A)	199	200	200	185	200	200	101	199	200
(B)	200	200	200	190	200	200	100	198	200
(C)	99	200	200	60	200	200	169	200	200
(D)	120	200	200	137	200	200	118	200	200
(E)	148	200	200	112	200	200	94	200	200
(F)	140	200	200	127	200	200	93	200	200
(G)	127	200	200	103	200	200	86	200	200
(H)	142	200	200	148	200	200	126	200	200
(I)	196	200	200	109	200	200	71	200	200
(J)	195	200	200	133	200	200	54	200	200
(K)	106	200	200	66	200	200	42	200	200
(L)	112	200	200	135	200	200	115	200	200

(ii) Frequency Among 200 Random Samples That Optimum Was Achieved									
DGP	Const = 1.0			Const = 0.5			Const = 0.0		
	ILPA	MILPA	BRCENS	ILPA	MILPA	BRCENS	ILPA	MILPA	BRCENS
(A)	198	199	199	174	175	184	31	50	90
(B)	200	200	200	179	179	184	33	53	74
(C)	81	68	128	36	43	82	8	13	84
(D)	95	111	160	109	106	155	93	102	148
(E)	126	138	169	92	115	164	64	81	134
(F)	127	154	183	113	145	174	71	107	149
(G)	108	110	156	77	100	136	63	60	109
(H)	114	124	161	127	117	157	102	109	153
(I)	192	190	196	92	94	157	40	33	104
(J)	191	190	194	110	128	160	43	74	123
(K)	81	78	145	47	57	99	24	28	60
(L)	84	99	155	108	106	155	89	98	147

(a) See main text for further explanations.

Table 3: Results of Simulation Study: For various Data Generating Processes (DGP), Absolute Frequencies that (i) Samples For Which Two Algorithms Converged and that (ii) Among Samples For Which Two Algorithms Converged One Algorithm Achieved A Solution With A Lower Value Of The Objective $Q_T(\beta)$ [a]

Const = 0.5									
	(i) Convergence Frequency Of Two Algorithms			(ii) Frequency That One Better Than Other					
DGP	BRCENS	BRCENS	ILPA	BRCENS	BRCENS	ILPA	MILPA	ILPA	MILPA
	converged with			achieved lower value of objective than					
	ILPA	MILPA	MILPA	ILPA	MILPA	BRCENS	BRCENS	MILPA	ILPA
(A)	185	200	185	4	14	5	2	8	0
(B)	190	200	190	1	7	2	1	3	0
(C)	60	200	60	5	91	30	12	39	0
(D)	137	200	137	2	71	14	1	58	0
(E)	112	200	112	15	86	17	15	42	2
(F)	127	200	127	16	81	18	15	41	3
(G)	103	200	103	9	83	17	20	41	0
(H)	148	200	148	4	58	14	0	48	0
(I)	109	200	109	12	98	14	15	42	2
(J)	133	200	133	15	75	6	11	18	0
(K)	66	200	66	5	96	16	27	38	1
(L)	135	200	135	1	71	14	1	57	0

(a) See main text for further explanations.

smaller for the samples with observation specific censoring points relative to those with a common censoring point for all observations. The result is not surprising, in light of the theoretical analysis in section 3.2.

Part (ii) in table 2 is concerned with the frequencies that the global optimum is achieved by the three algorithms. To determine the global optimum, a grid search is performed among 401×401 equidistant points in the rectangle $[-2,2] \times [-2,2]$.[12] If the minimizer on the grid is on the boundary of the rectangle, the sample is dismissed and a new sample is drawn instead. An algorithm is assumed to have achieved the optimum, if a value of the objective is reached which is less than or equal to the minimized value on the grid. With respect to this criterion, the results for BRCENS are very favorable in comparison to ILPA and MILPA. The relative performance of BRCENS is better, when the degree of censoring is higher and when there are more observation specific censoring points. The ranking between ILPA and MILPA is not clear. All three algorithms perform worse with a higher degree of censoring and with random design in constrast to fixed design, when there is higher

[12]Experiments with a grid of 1001×1001 equidistant points did not change the results reported here.

censoring. The latter result is true in particular for BRCENS. All three perform quite satisfactorily with moderate degrees of censoring, especially in the case with a common censoring point, DGP (A) and (B) and $Const = 1.0$.

Table 3 presents results concerning the relative performance of two algorithms conditional on convergence, for one value of $Const = 0.5$.[13] Since the CQR problem is not a convex optimization problem, it often cannot be expected that a global optimum is reached. Therefore, the relative performance of two algorithms is also of interest. The first three columns of the table indicate how often for a pair of algorithms both converge. For these cases, the next six columns provide the frequencies that one algorithm obtained a better value of the objective than the other. Clearly, ILPA and BRCENS outperform MILPA conditional on convergence, however, the ranking between ILPA and BRCENS is not obvious. It has to be recognized that conditional on convergence, ILPA quite often outperforms BRCENS and that MILPA does not appear to be a completely successful modification of ILPA.

5 Conclusions

This note is concerned with estimating censored quantile regressions (CQR). It establishes a new asymptotical result allowing for observation specific censoring points and fairly arbitrary forms of heteroskdasticity and autocorrelation. The interpolations property (IP) as a characterization of the CQR estimate is derived. As the main contribution, a new algorithm for the CQR estimation problem is developed building on the IP. Also two theretical inconsistencies are found in the iterative linear programming algorithm (ILPA), proposed by Buchinsky (1994). A modification of ILPA, the modified iterative linear programming algorithm (MILPA) is suggested. The subsequent simulation study compares the performance of the three algorithms. BRCENS performs quite well in comparison. It outperforms the two others with respect to the frequencies that the global optimum of the problem is reached. However, conditional on convergence of ILPA and BRCENS, there is no clear ranking between the two algorithms. All algorithms perform the worse, the higher the degree of censoring. Based on this simulation study, the application of BRCENS can be recommended, since it outperforms ILPA significantly for problems with considerable censoring. Despite the better convergence properties, MILPA is not a superior alternative to ILPA. Modifying BRCENS analogous to the most efficient modifications the Barrodale–Roberts-Algorithm for median regressions, as discussed in Dielman (1992), should lead in the future to a numerically more efficient version of BRCENS.

6 Appendix

This appendix contains the proof of theorem 1 in section 2.1 and a short description of the simplex setup and the calculation of the "marginal cost" reduction in

[13]Results are also obtained for $Const = 1.0$ and 0.0, but are not reported here. They are in line from what can be expected from the results in tables 2 and 3.

the algorithm BRCENS.

Proof of Theorem 1: The proof follows Powell (1984), (1986) and Fitzenberger (1993). I only highlight where an adaption to the case considered in this note is necessary. The basic asymptotic results used in this proof, i.e., the law of large numbers, the central limit theorem and the adaption of Huber's (1967) Lemma 3, are presented in Fitzenberger (1993), appendix 1.

a.) Consistency

Define $\epsilon_t^* = y_t - min[x_t'\beta_0, yc_t]$ and the distance function

$$\begin{aligned} S_T(\beta) &= \tfrac{1}{T}\textstyle\sum_{t=1}^{T} sgn_\theta(y_t - min[x_t'\beta, yc_t])\,(y_t - min[x_t'\beta, yc_t]) - \tfrac{1}{T}\sum_{t=1}^{T} sgn_\theta(\epsilon_t^*)\epsilon_t^* \\ &\equiv \tfrac{1}{T}\textstyle\sum_{t=1}^{T} h_t(\epsilon_t^*, x_t'\beta, x_t'\beta_0, yc_t) \end{aligned}$$

Note

$$h_t(\epsilon_t^*, x_t'\beta, x_t'\beta_0, yc_t) = \begin{cases} \theta x_t'(\beta_0 - \beta) & y_t \geq x_t'\beta \text{ and } y_t \geq x_t'\beta_0 \\ (1-\theta)x_t'(\beta - \beta_0) - \epsilon_t^* & y_t < x_t'\beta \leq yc_t \text{ and } y_t \geq x_t'\beta_0 \\ \theta x_t'(\beta_0 - \beta) + \epsilon_t^* & y_t \geq x_t'\beta \text{ and } y_t < x_t'\beta_0 \leq yc_t \\ (1-\theta)x_t'(\beta - \beta_0) & y_t < x_t'\beta \leq yc_t \text{ and } y_t < x_t'\beta_0 \leq yc_t \\ (1-\theta)(yc_t - x_t'\beta_0) \quad \text{for} & yc_t < x_t'\beta \text{ and } y_t < x_t'\beta_0 \leq yc_t \\ (1-\theta)(yc_t - x_t'\beta_0) - \epsilon_t^* & yc_t < x_t'\beta \text{ and } x_t'\beta_0 \leq y_t \\ \theta(yc_t - x_t'\beta) + \epsilon_t^* & y_t \geq x_t'\beta \text{ and } x_t'\beta_0 > yc_t \\ (1-\theta)(x_t'\beta - yc_t) & y_t < x_t'\beta \leq yc_t \text{ and } x_t'\beta_0 > yc_t \\ 0 & yc_t < x_t'\beta \text{ and } x_t'\beta_0 > yc_t \end{cases}$$

Thus, $|h_t|$ is O($\| x_t \|$) and analogous to Fitzenberger (1993), it follows that $S_T(\beta) - ES_T(\beta) \to 0$ a.s. uniformly in $\beta \in B$.

To establish consistency, it suffices to show that β_0 is the identifiably unique minimizer of $ES_T(\beta)$. This can be done analogous to Powell (1984), p. 318–318, and Fitzenberger (1993), appendix 3, by first providing a uniform lower bound of the conditional expectation $E\{h_t|F_t\}$ based on condition (ii)(c), and second using condition (iv)(b) when manipulating the unconditional expectation $ES_T(\beta)$.

b.) Asymptotic Normality

This part makes use of Lemma A.5 in Fitzenberger (1993), appendix 1, an extension of Huber's (1967) Lemma 3, to basically interchange asymptotically the subgradient of the non-smooth distance function, $S_T(\beta)$, by the gradient of the smooth limit function, $ES_T(\beta)$.

Define $\psi_i(\epsilon_t, x_t, \beta, yc_t) = I(x_t'\beta < yc_t)sgn_\theta(\epsilon_t - x_t'(\beta - \beta_0))x_{ti}$, $\psi(\epsilon_t, x_t, \beta, yc_t) \equiv (\psi_1, ..., \psi_k)$ and $\lambda_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} E\psi(\epsilon_t, x_t, \beta, yc_t)$. Analogous to Powell (1994) and Fitzenberger (1993), the argument follows by the following steps:

1) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi(\epsilon_t, x_t, \widehat{\beta_T}, yc_t) \;=\; o_p(1)$

2) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi(\epsilon_t, x_t, \beta_0, yc_t) + \sqrt{T}\lambda_T(\widehat{\beta_T}) \;\xrightarrow{P}\; 0$

3) $\lambda_T(\beta) \;=\; \frac{\partial \lambda_T(\beta)}{\partial \beta} \mid_{\bar{\beta}} (\beta - \beta_0)$ where
$\frac{\partial \lambda_T(\beta)}{\partial \beta} \;=\; -\frac{1}{T} \sum_{t=1}^{T} E\{f_t[x_t'(\beta - \beta_0)]\, I(x_t'\beta < yc_t)\, x_t x_t'\}$ and $\frac{\partial \lambda_T(\beta)}{\partial \beta} \mid_{\beta_0} = -L_T$

4) $J_T^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi(\epsilon_t, x_t, \beta_0, yc_t) \;\xrightarrow{D}\; N(0, I_k)$

The following considerations describe how to establish steps 1 to 4:

1) This follows by considering the left partials of $S_T(\beta)$

$$H_{Ti}(\beta) = \lim_{a \uparrow 0} \frac{S_T(\beta + ae_i) - S_T(\beta)}{a}$$

where $e_i = (0,.,0,1,0,..,0)$ has a one as the i^{th} component. It can be shown that

$$\begin{aligned} H_{Ti}(\beta) = \; & -\tfrac{1}{T} \textstyle\sum_{t=1}^{T} \psi_i(\epsilon_t, x_t, \beta, yc_t) \\ & -\tfrac{1}{T} \textstyle\sum_{t=1}^{T} I(x_t'\beta < yc_t, x_t'\beta = y_t)\, x_{ti} sgn_\theta(x_{ti}) \\ & +\tfrac{1}{T} \textstyle\sum_{t=1}^{T} I(x_t'\beta = yc_t)\{(1-\theta) I(y_t < yc_t, x_{ti} > 0) x_{ti} \\ & \qquad\qquad -\theta I(y_t = yc_t, x_{ti} > 0) x_{ti}\} \end{aligned}$$

The result follows analogous to Powell (1984), p. 320, and Fitzenberger (1993), appendix 3, recognizing the optimality properties of $\widehat{\beta_T}$ and condition (iv)(d), which guarantees that $\sum_{t=1}^{T} I(x_t'\widehat{\beta_T} = yc_t)$ is a.s. finite.

2) and 3) It suffices to show that $\lambda_T(\beta)$ satisfies the conditions of the extension of Huber's (1967) Lemma 3 in Fitzenberger (1993), Lemma A.5. Then the steps are established analogous to Powell (1984), p. 320–322, whereby in particular "E_x" has to be replaced by "$E\{...|F_t\}$" and "$f(.)$" by "$f_t(.)$". Conditions (ii)(b),(c), and (iv)(d) guarantee that Powell's argument can be adapted to the setup in this note.

4) This follows from the strong mixing assumption, condition (ii)(a), and condition (iii) in light of the central limit theorem for strong mixing processes in Fitzenberger (1993), Lemma A.4 .

Q.E.D.

Implementation of the Algorithm BRCENS:

In the following, the setup for the condensed simplex tableau of the algorithm BRCENS is described. Most importantly, I develop how to calculate the "marginal cost" reductions. The presentation corresponds to Barrodale and Roberts (1973). The CQR problem can be formulated "close" to a linear programming problem. Define $u_t = max(y_t - min[x_t'\beta, yc_t], 0)$, $v_t = -min(y_t - min[x_t'\beta, yc_t], 0)$, $b_i = max(\beta_i, 0)$, and $c_i = -min(\beta_i, 0)$; then the estimation problem in equation 1 can be reformulated as

Table 4: Condensed Simplex Tableau for the Algorithm BRCENS

Basis	nb_1 nb_2 ... nb_k	Residuals
bv_1	$\phi_{1,1}$ $\phi_{2,1}$ ... $\phi_{k,1}$	f_1
bv_1	$\phi_{1,2}$ $\phi_{2,2}$ ... $\phi_{k,2}$	f_2
$\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$
bv_T	$\phi_{1,T}$ $\phi_{2,T}$ ... $\phi_{k,T}$	f_T
Marginal Cost reductions	mcr_1 mcr_2 ... mcr_k	

$$\min \textstyle\sum_{t=1}^{T}(\theta u_t + (1-\theta)v_t)$$

$$\text{s.t. } y_t = min[\textstyle\sum_{i=1}^{k} x_{ti}(b_i - c_i), yc_t] + u_t - v_t$$

$$\text{and } b_i,\ c_i,\ u_t,\ v_t \geq 0 \text{ for } t = 1,...,T,\ i = 1,...k$$

The simplex iterations can be performed within a array of dimension $(T+2)\times(k+2)$ which is displayed in table 2. bv_t, $t = 1,...,T$, represent the basis variables and nb_i, $i = 1,...k$, the nonbasic variables. Each one is one of the variables u_t, v_t, b_i, or c_i for some i or t. $\Phi = (\phi_{i,t})_{i=1,...,k,t=1,...,T}$ stores the representation of the nonbasic variables in the current basis. The residuals columns contains θu_t or $(1-\theta)v_t$, if bv_t is a data point, and zero otherwise. Initially the array contains the data, i.e. $bv_t = u_t$ or v_t, $nb_i = b_i$ or c_i, $\Phi = X$, and $sgn_\theta(y_t)\,(y_t)$ as the tth component of the residual column.

It remains to describe the calculation of the "marginal cost" reductions within this setup. For the nonbasic variable nb_i, i=1,...,k, the "marginal cost" reductions prove to be

$$\begin{aligned} mcr_i = \ & \textstyle\sum_{t=1}^{T} \phi_{i,t} I(bv_t \text{ is a data point })[I(x'_{(t)}\beta < yc_{(t)}) \\ & \{\theta I(y_{(t)} > x'_{(t)}\beta) + (1-\theta)I(y_{(t)} < x'_{(t)}\beta) \\ & \theta I(y_{(t)} = x'_{(t)}\beta, bv_t \text{ is some } u_{t'}, \phi_{i,t} < 0) + (1-\theta)I(y_{(t)} = x'_{(t)}\beta, bv_t \text{ is some } v_{t'}, \phi_{i,t} < \\ & +I(x'_{(t)}\beta = yc_{(t)})\{(1-\theta)I(y_{(t)} < yc_{(t)}, \phi_{i,t} > 0) \\ & +\theta I(y_{(t)} = yc_{(t)}, bv_t \text{ is some } u_{t'}, \phi_{i,t} < 0) - \theta I(y_{(t)} = yc_{(t)}, bv_t \text{ is some } v_{t'}, \phi_{i,t} > 0)\}] \\ & -\{\theta I(nb_i \text{ is some } u_{t'}) + (1-\theta)I(nb_i \text{ is some } v_{t'}, y_{(t)} < yc_{(t)})\} \end{aligned}$$

where for some basis variable bv_t, if it represents a data point, this is $(y_{(t)}, yc_{(t)}, x_{(t)})$.

References

[1] **Barrodale, I. and F.D.K. Roberts** (1973). An Improved Algorithm for Discrete l_1 Linear Approximation. *SIAM Journal of Numerical Analysis*, 10:839-848.

[2] **Barrodale, I. and F.D.K. Roberts** (1974). Algorithm 478: Solution of an Overdetermined System of Equations in the l_1 Norm. *Communications of the Association for Computing Machinery*, 17:319–320.

[3] **Buchinsky, M.** (1994). Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression. *Econometrica*, 62:405–458.

[4] **Dielman, T.E.** (1992). Computational Algorithms for Least Absolute Value Regression. In Y. Dodge, editor, *L_1-Statistical Analysis and Related Methods*, pages 311–326. Elsevier Science Publishers B.V., North Holland, Amsterdam.

[5] **Fitzenberger, B.** (1993). The Moving Blocks Bootstrap and Robust Inference in Linear Regressions – Theory. *Center for Economic Policy Research (CEPR) Technical Paper No. 340, Stanford University.*

[6] **Fitzenberger, B., Hujer, R., MaCurdy, T.E., and R. Schnabel**, (1994). The Dynamic Structure of Wages in Germany 1976–1984, A Cohort Analysis. *Mimeo.*

[7] **Huber, P.J.** (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In LeCam, L.M. and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1* , pages 221–233. University of California, Berkeley, University of California.

[8] **Koenker, R. and G. Bassett** (1978). Regression Quantiles. *Econometrica*, 46:33–50.

[9] **Koenker, R. and V. D'Orey** (1987). Algorithm AS 229: Computing Regression Quantiles. *Statistical Algorithms*, pages 383–393.

[10] **Portnoy, S.L.** (1991). Asymptotic Behavior of Regression Quantiles in Non-stationary, Dependent Cases. *Journal of Multivariate Analysis*, 38:100–113.

[11] **Powell, J.L.** (1984). Least Absolute Deviations for the Censored Regression Model. *Journal of Econometrics*, 25:303–325.

[12] **Powell, J.L.** (1986). Censored Regression Quantiles . *Journal of Econometrics*, 32:143–155.

[13] **Weiss, A.A.** (1991). Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation . *Econometric Theory*, 7:46–68.

[14] **Womersley, R.S.** (1986). Censored Discrete l_1 Approximation. *SIAM Journal of Scientific Statistical Computations*, 7:105–122.