First Draft
Comments Are Welcome

Quantile Regression Model with Unknown Censoring Point[1]

Moshe Buchinsky[2]
Yale University and NBER

and

Jinyong Hahn[3]
University of Pennsylvania

April, 1995

[1] We thank Don Andrews for sharing his knowledge with us and helping with several delicate technical problems. Comments by Gary Chamberlain and Mark Rosenzweig are greatly appreciated. Financial support to the first author from the NSF under grant SBR9320386 is acknowledged.

[2] Department of Economics, Yale University, P.O. Box 208264, Yale Station, New Haven, CT 06520-8264.

[3] Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297.

Abstract

The paper introduces an estimator for the linear censored quantile regression model when the censoring point is an unknown function of a set of regressors. The objective function minimized is convex and the minimization problem is a linear programming problem, for which there is a global minimum. The suggested procedure applies also to the special case of a fixed known censoring point. Under fairly weak conditions the estimator is shown to have $\sqrt{n}$-convergence rate and is asymptotically normal. In the special case of a fixed censoring point it is aymptotically equivalent to the estimator suggested by Powell (1984, 1986a). A Monte Carlo study performed shows that the suggested estimator has very desirable small sample properties. It precisely corrects for the bias induced by censoring, even when there is a large amount of censoring, and for relatively small sample sizes. The estimator outperforms that suggested by Powell in cases where both apply.

1 Introduction

The "Tobit" model has received much attention in the literature since its introduction (Tobin (1958)), and a variety of parametric and semiparametric estimation methods have been proposed. While the older parametric estimators entail rather stringent distributional assumptions, the semiparametric estimators considerably relax these assumptions. The more recent estimators require weaker conditions on the underlying stochastic elements but, in general, impose more direct restrictions on the implied moments or shapes of the distributions. Although the specific assumptions vary greatly from one paper to another, a common feature is that the censoring points are known (even for non-censored observations) and are not independent of the regressors. We follow Honoré and Powell (1993) in referring to these models as fixed censoring models.

The most notable papers in the semiparametric literature on censored models—Powell (1984, 1986a), Nawata (1990), and Newey and Powell (1990)—put forward only conditional quantile restrictions. Papers that apply conditional symmetry restrictions on the distribution of the error term are Powell (1986b) and Newey (1989). Horowitz (1986, 1988), Honoré and Powell (1993), and Moon (1989) exploited an assumption of independence between the error term and the regressors to provide estimators for the underlying linear models. A somewhat different literature on censored models has been developed in biometrics and statistics, wherein the censoring values are observed only for the observations that are censored. Honoré and Powell (1993) have recently developed an estimator for such models that assume a random censoring point.[4]

The estimation procedure of the Tobit model in the context of quantile regression is motivated by two main factors: (a) relaxing the assumptions under which the Tobit model can be consistently estimated, and (b) the growing interest in quantile regression, in which censoring problems seem frequently to surface.[5]

The present paper considers the distribution of the dependent variable, conditional on not being censored, to arrive at a weighting scheme of the residuals that corrects for censoring. We address the censoring problem in a context similar to that introduced by Powell (1984 and 1986a); however, we relax the assumptions in the existing literature by allowing the censoring point to be an unknown, fixed function of a known set of regressors.

Using the seminal framework of Pollard (1991), we provide proofs, under fairly weak conditions, for consistency and asymptotic normality for the estimate of the coefficient vector in the linear quantile regression model. Furthermore, our estimator is shown to be a solution of a linear programming problem, which can be solved using a wide variety of linear programming algorithms. A special advantage of such an estimator is that it is a global minimizer, since the objective function is convex. Also, convergence to a solution occurs in a finite number of simplex iterations. We show that for a model with a fixed known censoring value, our estimator has the same asymptotic distribution as Powell's (1984, 1986a) estimator.

In a sequence of Monte Carlo simulations we find that our estimator performs very well. It corrects for the bias associated with a large amount of censoring, even in small samples. The bias of the suggested estimator is small and the variance is also relatively small. In cases where both our estimator and Powell's apply, our estimator performs better in terms of both mean squared

[4] See Honoré and Powell (1993) for an extensive reference list of work in this area.

[5] Recent works that involve different aspects of the estimation of quantile regressions include: Buchinsky (1994, 1995a, and 1995b), Chamberlain (1991), Hahn (1995) and Powell (1984 and 1986a, 1991), for linear quantile regression models, and Oberhofer (1982), and Koenker and Park (1994) for nonlinear quantile regression.

error and variance. Furthermore, the performance of our estimator increases faster than that of Powell's estimator for larger sample sizes.

The paper is organized as follows. Section 2 describes the model and gives some motivation and intuition for the proposed estimator. It also provides the notation and definitions that are used. Section 3 discusses the asymptotic properties. Section 4 compares our estimator with Powell's (1986a). In Section 5 we present Monte Carlo experiments designed to investigate the small sample properties. Section 6 consists of a summary and conclusions. An appendix presents proofs for several lemmas and a corollary.

2 Description, Definitions and Motivation

The censored regression known as the "Tobit" model can be written (e.g., Powell (1984)) as:

$$y_i = \max\{y^0, x_i'\beta + \epsilon_i\}.$$

That is, y_i is observed only if it is greater (for left censoring) than some threshold y^0; or

$$y_i = \min\{y^0, x_i'\beta + \epsilon_i\},$$

for right censored data.[6] Powell (1984 and 1986a) suggested an intuitive estimator for the censored quantile version of the Tobit model: With the quantile restriction $\text{Quant}_\theta(\epsilon_\theta \mid x) = 0$ for the latent variable $y^* = x'\beta_\theta + \epsilon_\theta$, Powell's estimator is obtained by solving

$$\min_\beta S_n(\beta, \theta) = \min_\beta \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_\theta(y_i - \max\{y^0, x_i'\beta\}) \right\}, \tag{1}$$

where $\rho_\theta(\lambda)$ is the check function

$$\rho_\theta(\lambda) = [\theta - I(\lambda < 0)]\lambda,$$

and $I(\cdot)$ is the usual indicator function. This estimator is the sample analog of the population minimization problem given by

$$\min_\beta \Phi(\beta, \theta) = \min_\beta \mathrm{E}\left[\rho_\theta \left(y_i - \max\{y^0, x_i'\beta\}\right) \mid x_i\right],$$

which defines the θth conditional quantile of y_i, conditional on $x_i, x_i'\beta_\theta$, where

$$\beta_\theta = \arg\min_\beta \Phi(\beta, \theta).$$

Powell derived the asymptotic properties of the estimator $\hat{\beta}_\theta$ obtained from solving the problem in (1) and showed that it is consistent and asymptotically normal.

While Powell's work has established an ingenious way for dealing with censoring in the context of quantile regression, his estimator has a few disadvantages due to the form of the objective function. In the first place the censoring point y_i^0 must be known. There are many cases where

[6] The censoring point y^0 need not be the same for all i ($i = 1, \ldots, n$), but it needs to be known.

the only information available to the econometrician is the set of variables that determine the censoring (or threshold) point. Such a situation arises in eligibility requirements for some social programs that are set exogenously by the government. Another situation is in using data, say on wages, that are self selected. For example, in a typical dynamic programming model a decision about participation in the labor force is made by comparing the value function from not working against that of a particular wage offered to the individual.

Another problem is that the objective function is not convex so that one is not guaranteed a global minimizer, $\hat{\beta}_\theta$, for $S_n(\beta, \theta)$ in (1). Lastly, it is not easy computationally to obtain a solution to the problem in (1). A relatively simple algorithm suggested in Buchinsky (1994) uses a linear programming (LP) algorithm in an iterative fashion (ILPA algorithm), but does not guarantee convergence to a solution. A more sophisticated algorithm suggested by Womerersley (1986) also provides only a local minimum.

Our new estimator improves on Powell's (1984, 1986a) in all of the above areas. First, we do not assume that the censoring point is known, but only that it is a fixed function of a known set of variables. Second, we design an estimator based on an objective function that is globally convex, so that a global minimum can be found. Finally, it can be shown that the problem is an LP problem and hence simplifies the solution computationally and guarantees convergence in a finite number of simplex iterations.

The Model:

The model is most easily described in the latent variable framework

$$Y^* = X'\beta_\theta + \epsilon_\theta, \tag{2}$$

where Y^* is a latent variable which is observed only if above a certain threshold, that is,

$$Y = \begin{cases} Y^* & \text{if } Y^* > \varphi_0(X) \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

with Y the observed dependent variable, and $\varphi_0(X)$ a fixed unknown function of the regressors X. We assume, as in Koenker and Bassett (1978), that the conditional quantile of the error term ϵ_θ, conditional on X, is

$$\text{Quant}_\theta(\epsilon_\theta \mid X) = 0, \tag{4}$$

for some $0 < \theta < 1$. Rewriting condition (3) in terms of ϵ_θ gives

$$Y = Y^* I(\epsilon_\theta > \phi_0(X)),$$

where

$$\phi_0(x) = \varphi_0(x) - x'\beta_\theta$$

and $I(\cdot)$ denotes, as before, the usual indicator function.

For simplicity of presentation we define

$$D \equiv I(Y^* > \varphi_0(X))$$

and let the vector Z be defined by

$$Z \equiv (Y, D, X)'.$$

This is the Tobit model except that here $\varphi_0(X)$ is unknown. The estimator suggested below will apply also to the case that the censoring point is known.

Motivation:

Denote the probability that $D = 1$ conditional on $X = x$ by

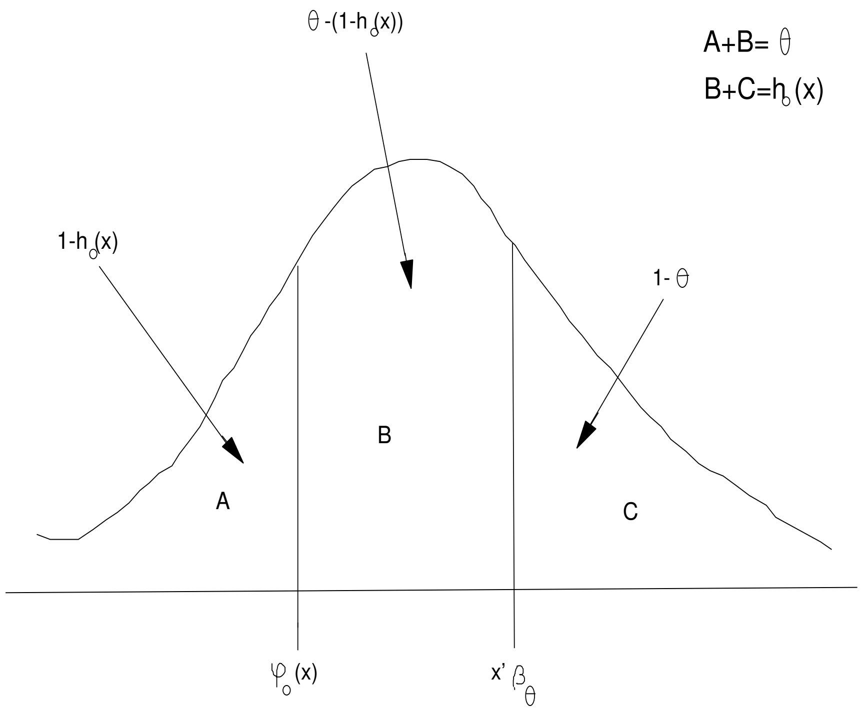
$$h_0(x) = \Pr(D = 1 \mid X = x).$$

Consider the conditional distribution of Y (i.e., Y^*D), conditional on $X = x$, $D = 1$ and $h_0(X) > 1-\theta$ for $0 < \theta < 1$, as defined in equations (2)–(4). For this distribution, $x'\beta_\theta$ is the $\pi_\theta(x)$ quantile where

$$\pi_\theta(x) = \frac{h_0(x) - (1-\theta)}{h_0(x)}, \tag{5}$$

because $X'\beta_\theta$ is the θth conditional quantile of Y^*, conditional on X. This point is illustrated in the diagram of the conditional distribution of Y^*, conditional on X, depicted in Figure 1. Note from this figure that the conditional probability that Y^* will be left of $x'\beta_\theta$, given that $Y^* > \varphi_0(x)$, is $\pi_\theta(x)$. Hence $x'\beta_\theta$ is the $\pi_\theta(x)$ quantile of the conditional distribution of Y^*, conditional on $X = x$, $D = 1$ and $\theta-(1-h_0(x)) > 0$. We use the last condition in the estimation as it insures that $\varphi_0(x) < x'\beta_\theta$, in which case $x'\beta_\theta$ is in the part of the distribution that is observed and therefore can be identified. This leads to the practical conclusion that the observations for which $\pi_\theta(x) < 0$ cannot be used in the estimation.

Figure 1: Density of Y^*, Conditional on $X = x$



Thus, the population parameter β_θ must solve the problem given by

$$\min_\beta \mathrm{E}\left[\pi_\theta(X)(Y - X'\beta)^+ + (1-\pi_\theta(X))(Y - X'\beta)^- \mid X, D=1, \pi_\theta(X) > 0\right],$$

where $a^+ \equiv \max\{0, a\}$ and $a^- \equiv \max\{0, -a\}$, and β_θ can then be written as

$$\beta_\theta = \arg\min_\beta \mathrm{E}\left[DI(\pi_\theta(X) > 0)\left(\pi_\theta(X)(Y - X'\beta)^+ + (1-\pi_\theta(X))(Y - X'\beta)^-\right)\right]. \quad (6)$$

If the conditional probabilities $h_0(X)$ were known, then β_θ could have been estimated using the sample analog of (6) (see, for example, Manski (1988)), that is,

$$\hat{\beta}_\theta = \arg\min_\beta \frac{1}{n}\sum_{i=1}^n D_i I\left(\pi_\theta(X_i) > 0\right)\left[\pi_\theta(X_i)(Y_i - X_i'\beta)^+ + (1-\pi_\theta(X_i))(Y_i - X_i'\beta)^-\right]. \quad (7)$$

The problem we have is that $h_0(X)$, and therefore $\pi_\theta(X)$, is not known so that the estimation in (7) cannot be carried out. One can, however, estimate $h_0(X)$ using a nonparametric method. Denote this estimator by $\hat{h}(X)$ and denote the corresponding estimator for $\pi_\theta(X)$ by $\hat{\pi}_\theta(X)$ so that the sample estimator is provided now by

$$\hat{\beta}_\theta = \arg\min_\beta \frac{1}{n}\sum_{i=1}^n D_i I\left(\hat{\pi}_\theta(X) > 0\right)\left[\hat{\pi}_\theta(X)(Y_i - X_i'\beta)^+ + (1-\hat{\pi}_\theta(X))(Y_i - X_i'\beta)^-\right].$$

For various technical reasons the actual estimator we propose—and which we refer to as the CQR-UCP (censored quantile regression with unknown censoring point) estimator—is given by

$$\hat{\beta}_\theta = \arg\min_\beta \sum_{i=1}^n D_i I(X \in \mathcal{X})\Lambda(\hat{\pi}_\theta(X_i))\left[\hat{\pi}_\theta(X)(Y_i - X_i'\beta)^+ + (1-\hat{\pi}_\theta(X))(Y_i - X_i'\beta)^-\right], \quad (8)$$

where $\Lambda(\cdot)$ is some smooth non-negative valued "trimming" function $\Lambda : \Re \to [0,1]$. This function takes the value 0 for all values of $\hat{\pi}_\theta < c$, for some small positive number c, and otherwise $0 < \Lambda(\pi_\theta) < 1$. The trimming of $\hat{\pi}_\theta$ insures that one does not use observations for which $\pi_\theta(X_i) < 0$.[7] The additional fixed trimming of X $(I(X_i \in \mathcal{X}))$ helps in simplifying the proofs.

3 Large Sample Property of the Estimator

In this section we establish the asymptotic properties of the estimator defined in (8). In particular we show that the CQR-UCP estimator is $\sqrt{n}$-consistent and asymptotically normal. We will extensively use the "convexity" framework established by Pollard (1991). Two main reasons motivated us to adopt this approach: (a) Unlike the conventional techniques we are able to establish in a single step consistency, $\sqrt{n}$-consistency and asymptotic normality; and (b) Pollard's approach seems especially appropriate since the objective function minimized in our formulation is globally convex (and piecewise linear). First, however, we impose regularity conditions which are rather standard in this literature. Proofs of the Lemmas and Corollary appearing below are given in the Appendix.

[7] Note that we do not require that $c \to 0$ as $n \to \infty$; the consistency and asymptotic normality is established for a fixed c.

3.1 Assumptions on the Error Term, Regressors and Trimming Function

Consider the latent variable model defined earlier in (2):

$$Y_i^* = X_i'\beta_\theta + \epsilon_{\theta i} \qquad (i = 1, \ldots, n).$$

Assumption ER: Error term and regressors

(i) The $(k+1)$-dimensional random vectors $\{(X_i', \epsilon_{\theta i})\}$ are independent and identically distributed.

(ii) The error term $\epsilon_{\theta i}$ has a continuous conditional density $f_{\epsilon_\theta}(\cdot \mid X_i)$ with a unique conditional θ-quantile equal to 0.

(iii) There exists a function $H_f(x)$, such that $f_{\epsilon_\theta}(0 \mid x) \leq H_f(x)$ and $\mathrm{E}[H_f(X_i)|X_i|^2] < \infty$.

(iv) The regressors X_i have a continuous density $f_X(x)$.

(v) The trimming set $\mathcal{X}$ of X, is a compact set.

(vi) $\mathrm{E}[|X_i|^4] < \infty$.

Assumption TF: Trimming function

The trimming function $\Lambda(\cdot)$ is bounded, with bounded and continuous first and second derivatives.

Assumption PD: Positive definiteness

The matrix

$$J \equiv \mathrm{E}\left[I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))f_{\epsilon_\theta}(0 \mid X_i)X_iX_i'\right]$$

is positive definite.

3.2 Kernel Estimation of $\pi_\theta(\cdot)$

As explained in the previous section, obtaining the estimator in (8) requires the estimation of $\pi_\theta(X_i)$. Denote the density function of X_i by $f_X(\cdot)$ and denote the kernel estimator for $f_X(\cdot)$ by $\hat{f}(\cdot)$. At a point x the estimator takes the form

$$\hat{f}(x) = \frac{1}{n\delta^k}\sum_{j=1}^{n} K\left(\frac{x - X_j}{\delta}\right), \tag{9}$$

where $K(\cdot)$ is an appropriately defined kernel function of dimension k (the dimension of X_i) and δ is a suitable bandwidth.

Let $\hat{h}(\cdot)$ denote the Nadaraya-Watson estimator for the regression $h_0(\cdot)$ of D_i on X_i, that is

$$\hat{h}(x) = \frac{\sum_{j=1}^{n} K\left((x - X_j)/\delta\right) D_j / n\delta^k}{\sum_{j=1}^{n} K\left((x - X_j)/\delta\right) / n\delta^k} = \frac{\hat{A}(x)}{\hat{f}(x)},$$

where

$$\hat{A}(x) = \frac{1}{n\delta^k} \sum_{j=1}^{n} K\left(\frac{x - X_j}{\delta}\right) D_j$$

is an estimator for

$$A_0(x) \equiv h_0(x) f_X(x)$$

and $\hat{f}(x)$ is defined in (9). Note that we can rewrite $\pi_\theta(\cdot)$ from (5), using the definition of $F_X(\cdot)$ and $A_0(\cdot)$, as

$$\pi_\theta(\cdot) = \frac{h_0(\cdot) - 1 + \theta}{h_0(\cdot)} = 1 - (1 - \theta)\frac{f_X(\cdot)}{A_0(\cdot)}.$$

We have already imposed an assumption on the density $f_X(\cdot)$ of X_i (Assumption ER). In the following we impose further restrictions on the function $A_0(\cdot)$ and the kernel function $K(\cdot)$.

Assumption K: Kernel function

The kernel function $K(\cdot)$ satisfies:

(i) $K(\cdot)$ is continuous and is zero outside a bounded set.

(ii) $\int K(u)du = 1$.

(iii) There is a positive integer m such that

$$\int u_1^{\ell_1} \ldots u_k^{\ell_k} K(u)du = 0, \quad \text{for} \quad \ell_1 + \ldots + \ell_k < m.$$

(iv) $f_X(x)$ and $A_0(x)$ are bounded away from 0 on $\mathcal{X}$.

(v) There are versions of the functions f_X and A_0 which are continuous on an open set containing $\mathcal{X}$.

(vi) $n\delta^{2k}/(\log n)^2 \to \infty$ and $n\delta^{2m} \to 0$.

3.3 Asymptotic Distribution of the Estimator

In this section we show, using Pollard's (1991) framework, that the estimator $\hat{\beta}_\theta$ obtained as a solution to (8), is asymptotically normal. Theorem 1 below establishes this result, while Corollary 1 computes the specific form of the asymptotic covariance matrix for $\hat{\beta}_\theta$.

Theorem 1

Suppose that Assumptions ER, TF, PD and K are satisfied. Then

$$n^{1/2}(\hat{\beta}_\theta - \beta_\theta) =$$
$$n^{-1/2} \sum_{i=1}^{n} J^{-1} X_i I(X_i \in \mathcal{X}) \Lambda(\pi_\theta(X_i)) \left(D_i(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)) + (1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)}\right)$$
$$+ o_p(1),$$

where $\hat{\beta}_\theta$ is defined by (8) and J is defined in Assumption PD.

Proof:

Define

$$\begin{aligned} G_n(\tau,\pi) &\equiv \sum_{i=1}^{n} D_i I(X_i \in \mathcal{X})\Lambda(\pi(X_i)) \\ &\quad \cdot \left[\pi(X_i)[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^+ - \epsilon_{\theta i}^+] + (1-\pi(X_i))[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^- - \epsilon_{\theta i}^-]\right] \\ &= \sum_{i=1}^{n} g_i(\tau,\pi), \end{aligned} \tag{10}$$

where

$$\begin{aligned} g_i(\tau,\pi) &= D_i I(X_i \in \mathcal{X})\Lambda(\pi(X_i)) \\ &\quad \times \left[\pi(X_i)[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^+ - \epsilon_{\theta i}^+] + (1-\pi(X_i))[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^- - \epsilon_{\theta i}^-]\right]. \end{aligned}$$

Note that $G_n(\tau,\hat{\pi}_\theta)$ is: (a) a convex function in τ; and (b) is minimized at

$$\tau_n = n^{1/2}(\hat{\beta}_\theta - \beta_\theta). \tag{11}$$

In the following we approximate the function $G_n(\tau,\pi)$ by a quadratic function whose minimizing value, η_n, is arbitrarily close to the value τ_n in (11). Furthermore, we show that η_n has an asymptotic normal distribution. Combining these two facts yields the conclusion that τ_n has an asymptotic normal distribution.

Let

$$\Gamma_n(\tau,\pi) = \mathrm{E}\left[G_n(\tau,\pi)\right].$$

Note that the function

$$\gamma(t,\pi_\theta) = \mathrm{E}[g_i(t,\pi_\theta)]$$

is minimized, as a function of t, at $t=0$ and with a second order derivative given by

$$\left.\frac{\partial^2\gamma(t,\pi_\theta)}{\partial t \partial t'}\right|_{t=0} = J \equiv \mathrm{E}\left[I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i)) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\right].$$

It follows that for fixed τ we have

$$\Gamma_n(\tau,\pi_\theta) = \frac{1}{2}\tau' J \tau + o(1). \tag{12}$$

In order to simplify the proof we define the following three quantities:

$$\zeta_n(Z_i,\pi) \equiv n^{-1/2}(\pi(X_i) - I(\epsilon_{\theta i} < 0))X_i, \tag{13}$$

$$W_n(\pi) \equiv \sum_{i=1}^{n} D_i I(X_i \in \mathcal{X})\Lambda(\pi(X_i))\zeta_n(Z_i,\pi), \tag{14}$$

$$\begin{aligned} R_n(Z_i,\pi,\tau) &\equiv D_i I(X_i \in \mathcal{X})\Lambda(\pi(X_i))\Big(\pi(X_i)\left[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^+ - \epsilon_{\theta i}^+\right] \\ &\quad + (1-\pi(X_i))\left[(\epsilon_{\theta i} - n^{-1/2}X_i'\tau)^- - \epsilon_{\theta i}^-\right] + \tau'\zeta_n(Z_i,\pi)\Big). \end{aligned} \tag{15}$$

Since

$$\mathrm{E}[\zeta_n(Z_i,\pi_\theta)|X_i,\pi_\theta(X_i)>0,D_i=1]=0,$$

we can write $G_n(\tau,\pi)$ from (10), using the notation in (13)–(15), as

$$G_n(\tau,\pi)=\Gamma_n(\tau,\pi_\theta)-\tau' W_n(\pi)+\sum_{i=1}^{n}\Big(R_n(Z_i,\pi,\tau)-\mathrm{E}[R_n(Z_i,\pi_\theta,\tau)]\Big). \tag{16}$$

We now state two lemmas (proofs given in the Appendix) which will allow us to write $G_n(\tau,\hat{\pi}_\theta)$ as

$$G_n(\tau,\hat{\pi}_\theta)=\frac{1}{2}\tau' J\tau-\tau' W_n(\hat{\pi}_\theta)+o_p(1),$$

and which will then enable us to directly apply Pollard's Convexity Lemma.

Lemma 1

Suppose that Assumptions ER, TF, and K are satisfied. Then,

$$\begin{aligned} W_n(\hat{\pi}_\theta) &= n^{-1/2}\sum_{i=1}^{n}X_i I(X_i\in\mathcal{X})\Lambda(\pi_\theta(X_i))\Big(D_i\cdot(\pi_\theta(X_i)-I(\epsilon_{\theta i}<0))+(1-\theta)\frac{D_i-h_0(X_i)}{h_0(X_i)}\Big) \\ &\quad + o_p(1). \end{aligned}$$

From Lemma 1 and the Central Limit Theorem it follows that $W_n(\hat{\pi}_\theta)$ has an asymptotic normal distribution. The arguments below also show that τ_n lies close to $J^{-1}W_n(\hat{\pi}_\theta)$, that is,

$$\tau_n=J^{-1}W_n(\hat{\pi}_\theta)+o_p(1),$$

and therefore that $\tau_n=n^{1/2}(\hat{\beta}_\theta-\beta_\theta)$ has an asymptotic normal distribution.

Lemma 2

Suppose that Assumptions ER(i)–(v), TF, and K are satisfied. Then for a fixed τ

$$\sum_{i=1}^{n}\left(R_n(Z_i,\pi,\tau)-\mathrm{E}\left[R_n(Z_i,\pi_\theta,\tau)\right]\right)=o_p(1).$$

Lemma 2 and equation (12) imply that G_n from (16) can be rewritten as

$$G_n(\tau,\hat{\pi}_\theta)=\frac{1}{2}\tau' J\tau-\tau' W_n(\hat{\pi}_\theta)+o_p(1),$$

for each fixed τ.

Pollard's Convexity Lemma (Pollard (1991, p. 187)) strengthens the pointwise convergence into uniform convergence on compact subsets $T \subset R^k$. That is, letting $\eta_n = J^{-1}W_n(\hat{\pi}_\theta)$, we may write the resulting convergence as

$$G_n(\tau, \hat{\pi}_\theta) = \frac{1}{2}(\tau - \eta_n)'J(\tau - \eta_n) - \frac{1}{2}\eta_n' J \eta_n + r_n(\tau), \tag{17}$$

where for each compact set $T \subset R^r$

$$\sup_{\tau \in T} |r_n(\tau)| = o_p(1). \tag{18}$$

The argument to establish Theorem 1 will be complete if we show, as in Pollard's case, that

$$\tau_n = J^{-1}W_n(\hat{\pi}_\theta) + o_p(1).$$

This is achieved in the following lemma (proof in the Appendix):

Lemma 3

Suppose that (17) and (18) are satisfied. Then for each $\varepsilon > 0$, $|\tau_n - \eta_n| < \varepsilon$ with probability tending to 1.

Q.E.D.

It remains to derive the asymptotic covariance matrix for the estimator $\hat{\beta}_\theta$. This is provided in the following corollary (proof in the Appendix):

Corollary 1

Given the assumptions stated in Theorem 1,

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \Omega_\theta), \quad \textit{as } n \to \infty,$$

where

$$\Omega_\theta = \theta(1-\theta)J^{-1}\Delta J^{-1},$$
$$\Delta = \mathrm{E}\left[I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))X_i X_i'\right],$$

and J is given in Assumption PD.

4 Comparison of the Estimator with Powell's Estimator

The estimator suggested by Powell (1984 and 1986a) applies only to the case where the censoring points are known for all observations, both censored and non-censored, that is, $\varphi_0(X_i) = y_i^0$, $i = 1, \ldots, n$. Our CQR-UCP estimator works for this case, as well as for the more general case where $\varphi_0(X_i)$ may not be known.

There are two main theoretical differences between the two estimators. Firstly, Powell's estimator does not require an estimate of the probability that an observation be in the sample, whereas our CQR-UCP estimator does require such an estimate. Secondly, the objective function used in obtaining the CQR-UCP estimator is globally convex while that for Powell's estimator is not. Consequently, Powell's estimator is guaranteed to be only a local minimizer. In contrast, the CQR-UCP estimator is a global minimizer. Also, since our problem is formulated as a linear programming problem, the solution is reached in a finite number of simplex iterations.

We compare the asymptotic properties of the two estimators when both apply, i.e., for known censoring points implying $\varphi_0(X_i) = y_i^0$. It turns out that the two estimators have (almost) the same asymptotic distribution, for suitable choices of $\mathcal{X}$ and $\Lambda(\cdot)$. In the next section we examine the small sample properties of the two estimators by way of Monte Carlo experiments.

Powell's (1986a) quantile regression estimator, say $\hat{\beta}_\theta^{\mathrm{P}}$, is known to have the asymptotic representation

$$\sqrt{n}(\hat{\beta}_\theta^{\mathrm{P}} - \beta_\theta) = \sqrt{n}\sum_{i=1}^{n} H^{-1} I\left(X_i'\beta_\theta > y_i^0\right)\left(\theta - I\left(\epsilon_{\theta i} < 0\right)\right) X_i,$$

where

$$H \equiv \mathrm{E}\Big[I\left(X_i'\beta_\theta > y_i^0\right) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\Big].$$

Thus the asymptotic covariance of $\hat{\beta}_\theta^{\mathrm{P}}$ is given by

$$\Omega_\theta^{\mathrm{P}} = \theta(1-\theta) H^{-1} \Delta_{\mathrm{P}} H^{-1},$$

where

$$\Delta_{\mathrm{P}} = \mathrm{E}\left[I\left(X_i'\beta_\theta > y_i^0\right) X_i X_i'\right].$$

The asymptotic distribution of our CQR-UCP estimator, say $\hat{\beta}_\theta^*$ depends on the choice of $\Lambda(\pi_\theta)$ and the trimming of X, $I(X \in \mathcal{X})$. Thus the two estimators can be compared only when: (a) $\Lambda(\pi_\theta)$ is very close to $I\left(\pi_\theta > 0\right)$; and (b) $\mathcal{X}$ contains the support of X_i. Putting $\Lambda(\pi_\theta) = I\left(\pi_\theta > 0\right)$ and $\mathcal{X} = R^k$, we can compare the asymptotic variance of $\hat{\beta}_\theta^*$ with that of $\hat{\beta}_\theta^{\mathrm{P}}$. For this particular model we have

$$\pi_\theta(X_i) > 0 \quad \text{if and only if} \quad X_i'\beta_\theta > y_i^0.$$

Thus,

$$\begin{aligned} J &= \mathrm{E}\left[I\left(\pi_\theta(X_i) > 0\right) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\right] \\ &= \mathrm{E}\left[I\left(X_i'\beta_\theta > y_i^0\right) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\right] \\ &= H. \end{aligned} \tag{19}$$

Similarly it follows that

$$\begin{aligned} \Delta &= \mathrm{E}\left[I(\pi_\theta(X_i) > 0) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\right] \\ &= \mathrm{E}\left[I(X_i'\beta_\theta > y_i^0) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i'\right] \\ &= \Delta_{\mathrm{P}}. \end{aligned} \tag{20}$$

We conclude therefore, from (19), (20), and Corollary 1, that the asymptotic covariances of Powell's estimator and CQR-UCP estimator are close if $\Lambda(\pi) \approx I\left(\pi > 0\right)$, and $\Pr\left(X_i \in \mathcal{X}\right) \approx 1$.

5 Finite Sample Properties of the Estimator

Section 3 provided the large sample properties of the proposed estimator. The CQR-UCP estimator has an asymptotic normal distribution, even though the reweighting scheme incorporates a nonparametric estimate for the probability of not being censored, conditional on the observed regressors. In practice, since including a nonparametric estimate when the sample is small can lead to severe bias in the estimate, it is important for empirical applications to evaluate the performance of the estimator for small samples.

This evaluation has been done in a Monte Carlo study which examines the small sample performance in two ways:

a. Comparing the Monte Carlo results for the CQR-UCP estimator with two alternative unadjusted quantile regression estimates, when the censoring point is a function of the regressors; and

b. Comparing Powell's (1986a) and the CQR-UCP estimators when the censoring values are known.

Two concerns that have affected the design of the Monte Carlo simulations carried out here are: the similarity with real data problems, and the feasibility of the Monte Carlo experiment, i.e., the examples need to be simple enough to perform a satisfactory Monte Carlo study. Some elements were kept unchanged throughout the experiment, while others were varied and the results are reported here in detail.

The model outlined in Section 2 was used:

$$Y_i^* = \alpha_0 + \beta_0 X_i + \epsilon_{\theta i},$$

$$Y_i = Y_i^* I(Y_i^* > \varphi_0(X_i)), \tag{21}$$

where the X_i are drawn from a standard normal distribution. For the error term, $\epsilon_{\theta i}$, a multiplicative heteroskedasticity formulation was adopted:

$$\epsilon_{\theta i} = u_i v(X_i)$$

where

$$v(X_i) = a_0 + a_1 x_i + a_2 x_i^2,$$

with $a_0 = 1$, $a_1 = .5$, and $a_2 = .5$. The u_i term was drawn from a normal distribution with zero mean and with a variance σ^2 that was varied, with $\sigma = 1, 2, 3, 4, 5$.

Two sets of experiments were carried out for the model in (21). In the first set $\varphi(x)$ was calculated for each observation according to the formula

$$\varphi_0(x) = b_0 + b_1 x + v_2 x^2$$

with $b_0 = .5$, $b_1 = .5$, and $b_2 = -2$. The aim was to compare the performance of the CQR-UCP estimator relative to quantile regression estimates that did not take the censoring into account. Two unadjusted quantile regressions were considered for this part: one that did not use the censored observations (reported as 0), and one that used them with the value 0 assigned to the dependent variable Y.

The second set of experiments was devoted to comparing Powell estimator and our CQR-UCP estimator for the special case that they both apply, namely, $\varphi(x)$ set equal to -1, and assumed to be known. For Powell's estimator we used the Iterated Linear Programming Algorithm (ILPA).[8]

For the solution of the CQR-UCP estimator, a few more choices were needed. In estimating $\mathrm{E}[D_i \mid X_i]$ nonparametrically we used the Epanechnikov kernel given by

$$K_e\left(t/h_n\right) = \frac{3}{4}\left(1-(t/h_n)^2\right) I\left((t/h_n)^2 < 1\right).$$

The bandwidth of the kernel h_n was selected using the likelihood cross validation procedure (see, for example, Silverman (1986)).[9] For the $\Lambda_0(\pi)$ trimming function we used

$$\begin{aligned}\Lambda_0(\pi) \quad = \quad & \left(\frac{\exp(\pi-2c)}{1+\exp(\pi-2c)} - \frac{\exp(-c)}{1+\exp(-c)}\right)\left(\frac{2+\exp(c)+\exp(-c)}{\exp(c)-\exp(-c)}\right) I(c<\pi<3c) \\ & +I(\pi>3c), \end{aligned} \tag{22}$$

with $c = .005$. This function has more than two bounded derivatives (as required by Assumption TF). Varying the trimming value c has no effect on the results. In fact, eliminating the function $\Lambda_0(\pi)$ entirely from the estimation also has no effect. This suggests that in practice it can be omitted. For purpose of internal consistency with our proofs we kept $\Lambda_0(\pi)$ in the form of (22).

a. Monte Carlo Experiments with Unknown $\varphi_0(x)$

The results for these Monte Carlo experiments are reported in Tables I through V, for $\sigma = 1,2,3,4$ and 5, respectively. For the larger values of σ more observations are affected by censoring since $\Pr(D_i = 1 \mid X_i)$ declines. Each table reports four statistics for the constant and slope: root mean squared error (RMSE), mean bias, median absolute error (MAE), and median bias. Three alternative solutions are considered in each of the five experiments: (a) Using all observations, ignoring the fact that some observations are censored; (b) using only the observations which are not censored, i.e. $Y_i > 0$; and (c) using the CQR-UCP estimator. Each table reports the number of observations used, in effect, by these three methods (i.e., the number of observations estimated to be above the censoring point).

Each Monte Carlo experiment was repeated for 3000 times and for five sample sizes: 50, 100, 200, 300, and 400. The results for $\sigma = 1$ (Table I) indicate that when only a few observations are censored, it does not matter significantly whether one corrects for the censoring or merely uses all observations at their reported values. Using only the uncensored observations (as is currently the most common practice) does yield slightly larger values of RMSE, MAE, and mean and median bias than for the CQR-UCP estimator, even for the very small sample sizes, for both the constant (columns 1-3) and the slope (columns 4-6).

---

[8] See Buchinsky (1994) for a detailed description. We also tried using the Nelder-Meade algorithm in solving for Powell's estimator. The results were virtually the same as for the ILPA algorithm and are therefore omitted from the following discussion.

[9] Other experiments using a normal kernel and/or the least–squares cross validation procedure yielded almost the same results and are also omitted.

For larger σ, where there is more censoring, there is a greater need for the correction. For example, for $\sigma = 2$, with approximately 20% of the observations censored, a very large bias occurs in the estimated slope, if one does not correct for censoring. This bias vanishes almost completely when the CQR-UCP estimation procedure is applied, even for very small samples. Increasing the sample size from 50 to 400 observations, for example, improves the performance of the CQR-UCP estimator, while for the unadjusted estimates the RMSE, MAE, and the mean and absolute biases do not change.

When 40% of the data is censored (see Table V for $\sigma = 5$)—not uncommon in empirical studies—the need for correction using the CQR-UCP estimator becomes crucial. For a sample size of 50 observations there is an extremely large bias induced by censoring, which is not fully corrected by the CQR-UCP estimator. A significant improvement occurs with larger sample sizes. For example, for sample size of 200, when all the data is used, the mean bias for the slope is about 40%, and is 120% when only the censored observations are used. It is only 12% for the CQR-UCP estimator, which decreases to 5% for a sample size of 400; no significant changes are observed for the other two alternatives.

It is important to note in Tables I–V that the number of observations used by the CQR-UCP method is on average very close to the number of uncensored points. That is to say, the reweighting of the positive and negative residuals is very effective using CQR-UCP. Furthermore, estimating the probability that an individual observation will be uncensored introduces no practical difficulties, even for relatively small sample sizes. This implies that the CQR-UCP estimator does not need large data sets in order to obtain good approximations to the asymptotic distribution.

Note that the traditional measures of precision—RMSE and mean bias—as well as the robust measures—MAE and median bias—provide essentially identical evaluations. Although the cases considered here are relatively simple, they do appear to be fairly typical in applications.

b. Monte Carlo Experiments with Known $\varphi_0(X)$

In this part of our Monte Carlo study we consider the case when $\varphi_0(X_i) = y_i^0$ are known, so that both Powell's estimator (CQR) and our CQR-UCP estimator apply and are asymptotically equivalent. The results are reported in Tables VI–X for $\sigma = 1$, 2, 3, 4, and 5, respectively, for five different sample sizes: 50, 100, 200, 300, and 400. The tables are organized similarly to Tables I–V.

The performance of the two estimators is comparable for small σ. For large σ, however (see Tables IX and X), the RMSE and the MAE are much larger for the CQR estimator than for the CQR-UCP estimator. These differences become smaller for larger sample sizes (see the bottom part of Tables IX and X). While the larger RMSE can be attributed mostly to a larger variance, the bias of the CQR estimator is also larger than the bias of the CQR-UCP estimator.

The number of observations used is noticeably different for the two estimators. The CQR estimator clearly tends to use far more observations than the CQR-UCP estimator. For example, for $\sigma = 5$ and a sample size of 400, the CQR estimator uses an average of 382 observations while the CQR-UCP estimator uses only 250 observations. This is to say that the CQR-UCP method selects, on average, the correct number of censored observations, while the CQR estimator uses too many. We do not attempt to investigate here why this phenomena occurs, but merely point out that the CQR-UCP estimator does not exhibit this distortion and hence yields, in general, more accurate results.

Finally, we note that the performance of the CQR-UCP estimator does not change much in these experiments relative to the previous experiments with an unknown $\varphi_0(x)$. All the measures of accuracy are very similar for the two simulations.

Overall, we have shown the importance of correcting the estimation for censoring especially when the censored observations are a large fraction of the available data. Moreover, we have demonstrated the suitability and accuracy of the CQR-UCP estimator.

6 Summary and Conclusions

This paper introduces a new estimator for the linear quantile regression model when censoring of an unknown form exists. As other studies in this area have also shown, there is a close relationship between the censoring point and the regressors, and this is true when the censoring point may be an unknown function of a known set of regressors.

We have developed an estimator that takes into account the probability of observing a non-censored observation, conditional on the regressors. We do this by adjusting the weights on the positive and negative residuals, based on the conditional probability of the observed dependent Y. The new estimator solves the problem (equation (8)):

$$\min_{\beta} \sum_{i=1}^{n} D_i I(X \in \mathcal{X}) \Lambda(\hat{\pi}_\theta(X_i)) \left[\hat{\pi}_\theta(X)(Y_i - X_i'\beta)^+ + (1 - \hat{\pi}_\theta(X))(Y_i - X_i'\beta)^- \right].$$

Minimization of this function requires a consistent estimate for the probability of being observed, $\mathrm{E}(D_i = 1 \mid X_i)$, in order to form a consistent estimate for $\pi_\theta(x) = (h_0(x) - 1 + \theta))/h_0(x)$. We obtain $\hat{h}(x)$ as a kernel estimate for $h_0(x)$.

The estimator $\hat{\beta}_\theta$ is unaffected by the preliminary estimation of h_0 even though $\hat{h}$ does not have $\sqrt{n}$-convergence rate. Our estimator is shown to have the desired $\sqrt{n}$-convergence rate and is asymptotically normal. Specifically,

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{\mathcal{L}} \mathrm{N}(0, \theta(1-\theta) J^{-1} \Delta J^{-1}),$$

where

$$J \equiv \mathrm{E}\left[I(X_i \in \mathcal{X}) \Lambda(\pi_\theta(X_i)) f_{\epsilon_\theta}(0 \mid X_i) X_i X_i' \right]$$

and

$$\Delta \equiv \mathrm{E}\left[I(X_i \in \mathcal{X}) \Lambda(\pi_\theta(X_i)) X_i X_i' \right].$$

While the new estimator is suited to deal with unknown censoring point, it also applies to the case considered by Powell (1984, 1986a), namely, fixed and known censoring values. We show that our estimator has (almost) the same asymptotic distribution as the estimator suggested by Powell.

There are several advantages to the new estimator. First, it minimizes a globally (piecewise linear) convex function; consequently, the estimator obtained provides a global minimum. Second, since the minimization problem is, in fact, a linear programming problem, a solution is obtained in a finite number of simplex iterations. Third, the technique is extremely easy to use, as cross-validation techniques are readily available for the estimation of $h_0(x)$.

In a sequence of Monte Carlo simulations it is shown that the suggested estimator has very desirable small sample properties. The estimates are very precise even for relatively small sample sizes and for data contaminated by a large amount of censoring. The root mean squared error and the median absolute error are quite small even for small sample sizes, and they rapidly decrease as the sample size increases. The bias of the estimates is also minimal. Thus, the the estimator should be a useful tool in empirical applications that require the use of quantile regression with censored data, especially when the censoring values are unknown.

TABLE I

Monte Carlo Simulation for .50 Quantile Regression
Unknown Censoring $\varphi_0(X)$, $\sigma = 1$

	All[a]	*Non Cens.*[b]	*UCP*[c]	*All*[a]	*Non Cens.*[b]	*UCP*[c]
	Constant			*Slope*		
50 Observations:						
RMSE	0.241	0.266	0.245	0.353	0.362	0.352
Mean bias	0.001	0.123	-0.008	0.003	0.062	0.006
Median abs. error	0.161	0.179	0.164	0.235	0.235	0.236
Median bias	0.011	0.121	0.002	0.004	0.063	0.004
Observations	50.0	46.5	46.6	50.0	46.5	46.6
100 Observations:						
RMSE	0.173	0.207	0.175	0.257	0.255	0.257
Mean bias	0.001	0.124	-0.006	0.001	0.051	0.003
Median abs. error	0.119	0.147	0.119	0.174	0.169	0.176
Median bias	-0.000	0.124	-0.007	0.003	0.050	0.005
Observations	100.0	93.1	93.1	100.0	93.1	93.1
200 Observations:						
RMSE	0.121	0.166	0.122	0.173	0.184	0.173
Mean bias	0.002	0.118	-0.004	-0.002	0.055	-0.000
Median abs. error	0.081	0.122	0.082	0.116	0.125	0.116
Median bias	0.002	0.118	-0.004	-0.000	0.058	0.002
Observations	200.0	186.0	186.1	200.0	186.0	186.1
300 Observations:						
RMSE	0.097	0.151	0.097	0.148	0.149	0.149
Mean bias	0.001	0.119	-0.004	-0.005	0.051	-0.005
Median abs. error	0.066	0.119	0.066	0.103	0.099	0.103
Median bias	0.002	0.118	-0.003	-0.005	0.052	-0.003
Observations	300.0	278.9	279.1	300.0	278.9	279.1
400 Observations:						
RMSE	0.085	0.146	0.086	0.123	0.135	0.123
Mean bias	0.001	0.121	-0.003	-0.004	0.054	-0.004
Median abs. error	0.059	0.121	0.059	0.081	0.091	0.081
Median bias	0.001	0.121	-0.002	-0.004	0.055	-0.003
Observations	400.0	372.0	372.2	400.0	372.0	372.2

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Using all observations including the censored observations.
[b] Using only non-censored observations.
[c] Using CQR-UCP estimator.

TABLE II

Monte Carlo Simulation for .50 Quantile Regression
Unknown Censoring $\varphi_0(X)$, $\sigma = 2$

	All[a]	*Non Cens.*[b]	*UCP*[c]	*All*[a]	*Non Cens.*[b]	*UCP*[c]
	Constant			*Slope*		
50 Observations:						
RMSE	0.462	0.850	0.498	0.647	0.765	0.697
Mean bias	0.046	0.715	-0.018	-0.088	0.331	0.023
Median abs. error	0.317	0.698	0.332	0.427	0.503	0.453
Median bias	0.022	0.698	-0.033	-0.093	0.325	0.008
Observations	50.0	39.9	39.8	50.0	39.9	39.8
100 Observations:						
RMSE	0.328	0.767	0.339	0.455	0.588	0.488
Mean bias	0.034	0.698	-0.019	-0.082	0.333	0.015
Median abs. error	0.218	0.689	0.223	0.315	0.398	0.333
Median bias	0.031	0.689	-0.023	-0.087	0.313	0.012
Observations	100.0	79.7	79.9	100.0	79.7	79.9
200 Observations:						
RMSE	0.237	0.744	0.242	0.332	0.476	0.354
Mean bias	0.030	0.708	-0.015	-0.076	0.332	0.015
Median abs. error	0.160	0.706	0.163	0.220	0.353	0.235
Median bias	0.026	0.706	-0.015	-0.082	0.335	0.013
Observations	200.0	159.7	159.9	200.0	159.7	159.9
300 Observations:						
RMSE	0.195	0.731	0.198	0.266	0.436	0.282
Mean bias	0.026	0.707	-0.016	-0.073	0.338	0.013
Median abs. error	0.134	0.702	0.134	0.181	0.342	0.188
Median bias	0.027	0.702	-0.013	-0.081	0.338	0.014
Observations	300.0	239.7	239.4	300.0	239.7	239.4
400 Observations:						
RMSE	0.171	0.721	0.173	0.237	0.412	0.249
Mean bias	0.029	0.704	-0.010	-0.074	0.332	0.009
Median abs. error	0.111	0.704	0.113	0.167	0.330	0.175
Median bias	0.024	0.704	-0.012	-0.075	0.328	0.008
Observations	400.0	319.3	319.3	400.0	319.3	319.3

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Using all observations including the censored observations.
[b] Using only non-censored observations.
[c] Using CQR-UCP estimator.

TABLE III

Monte Carlo Simulation for .50 Quantile Regression
Unknown Censoring $\varphi_0(X)$, $\sigma = 3$

	All[a]	*Non Cens.*[b]	*UCP*[c]	*All*[a]	*Non Cens.*[b]	*UCP*[c]
	Constant			*Slope*		
50 Observations:						
RMSE	0.674	1.662	0.752	0.846	1.185	1.015
Mean bias	0.161	1.522	0.046	-0.204	0.648	0.075
Median abs. error	0.465	1.513	0.479	0.616	0.798	0.679
Median bias	0.113	1.513	0.054	-0.235	0.642	0.069
Observations	50.0	36.0	35.6	50.0	36.0	35.6
100 Observations:						
RMSE	0.497	1.570	0.527	0.625	0.946	0.725
Mean bias	0.085	1.500	-0.015	-0.213	0.647	0.067
Median abs. error	0.337	1.483	0.335	0.434	0.681	0.489
Median bias	0.069	1.483	-0.015	-0.221	0.631	0.051
Observations	100.0	71.9	71.4	100.0	71.9	71.4
200 Observations:						
RMSE	0.360	1.521	0.360	0.469	0.803	0.513
Mean bias	0.063	1.484	-0.019	-0.228	0.632	0.033
Median abs. error	0.244	1.474	0.242	0.327	0.633	0.350
Median bias	0.057	1.474	-0.028	-0.239	0.627	0.033
Observations	200.0	143.9	143.7	200.0	143.9	143.7
300 Observations:						
RMSE	0.303	1.493	0.299	0.402	0.748	0.429
Mean bias	0.058	1.468	-0.018	-0.218	0.636	0.026
Median abs. error	0.205	1.464	0.205	0.287	0.627	0.294
Median bias	0.062	1.464	-0.022	-0.228	0.627	0.027
Observations	300.0	215.5	215.5	300.0	215.5	215.5
400 Observations:						
RMSE	0.258	1.495	0.257	0.357	0.722	0.366
Mean bias	0.049	1.476	-0.024	-0.219	0.633	0.026
Median abs. error	0.181	1.471	0.176	0.257	0.632	0.249
Median bias	0.052	1.471	-0.025	-0.220	0.632	0.025
Observations	400.0	287.5	287.3	400.0	287.5	287.3

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Using all observations including the censored observations.
[b] Using only non-censored observations.
[c] Using CQR-UCP estimator.

TABLE IV

Monte Carlo Simulation for .50 Quantile Regression
Unknown Censoring $\varphi_0(X)$, $\sigma = 4$

	All[a]	*Non Cens.*[b]	*UCP*[c]	*All*[a]	*Non Cens.*[b]	*UCP*[c]
	Constant			*Slope*		
50 Observations:						
RMSE	0.878	2.553	1.252	0.991	1.591	1.426
Mean bias	0.263	2.386	0.124	-0.258	0.906	0.214
Median abs. error	0.613	2.341	0.650	0.761	1.110	0.895
Median bias	0.208	2.341	0.178	-0.369	0.885	0.198
Observations	50.0	33.5	32.3	50.0	33.5	32.3
100 Observations:						
RMSE	0.633	2.439	0.689	0.728	1.306	0.937
Mean bias	0.119	2.356	0.008	-0.333	0.926	0.102
Median abs. error	0.440	2.350	0.436	0.547	0.984	0.621
Median bias	0.099	2.350	0.020	-0.378	0.940	0.105
Observations	100.0	67.1	65.7	100.0	67.1	65.7
200 Observations:						
RMSE	0.487	2.349	0.483	0.580	1.106	0.686
Mean bias	0.090	2.307	-0.001	-0.333	0.905	0.075
Median abs. error	0.332	2.293	0.333	0.424	0.906	0.464
Median bias	0.091	2.293	0.010	-0.353	0.902	0.100
Observations	200.0	134.1	133.4	200.0	134.1	133.4
300 Observations:						
RMSE	0.406	2.349	0.397	0.505	1.051	0.554
Mean bias	0.078	2.321	-0.007	-0.332	0.918	0.044
Median abs. error	0.274	2.311	0.275	0.373	0.922	0.367
Median bias	0.079	2.311	-0.013	-0.343	0.922	0.039
Observations	300.0	201.4	200.9	300.0	201.4	200.9
400 Observations:						
RMSE	0.342	2.326	0.336	0.470	1.015	0.490
Mean bias	0.066	2.305	-0.018	-0.335	0.907	0.040
Median abs. error	0.237	2.302	0.228	0.355	0.907	0.330
Median bias	0.071	2.302	-0.012	-0.340	0.907	0.045
Observations	400.0	268.1	267.8	400.0	268.1	267.8

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Using all observations including the censored observations.
[b] Using only non-censored observations.
[c] Using CQR-UCP estimator.

TABLE V

Monte Carlo Simulation for .50 Quantile Regression Unknown Censoring $\varphi_0(X)$, $\sigma = 5$

	All[a]	*Non Cens.*[b]	*UCP*[c]	*All*[a]	*Non Cens.*[b]	*UCP*[c]
	Constant			*Slope*		
50 Observations:						
RMSE	1.041	3.414	2.159	1.089	1.990	2.234
Mean bias	0.304	3.236	0.173	-0.366	1.184	0.287
Median abs. error	0.721	3.171	0.763	0.925	1.395	1.094
Median bias	0.187	3.171	0.249	-0.582	1.156	0.259
Observations	50.0	31.8	29.4	50.0	31.8	29.4
100 Observations:						
RMSE	0.771	3.285	0.921	0.829	1.653	1.130
Mean bias	0.188	3.190	0.089	-0.403	1.218	0.158
Median abs. error	0.532	3.144	0.525	0.628	1.251	0.756
Median bias	0.150	3.144	0.098	-0.459	1.210	0.164
Observations	100.0	64.1	61.6	100.0	64.1	61.6
200 Observations:						
RMSE	0.588	3.242	0.604	0.655	1.453	0.820
Mean bias	0.127	3.195	0.039	-0.395	1.217	0.118
Median abs. error	0.407	3.194	0.403	0.470	1.216	0.572
Median bias	0.114	3.194	0.044	-0.399	1.214	0.133
Observations	200.0	128.0	125.8	200.0	128.0	125.8
300 Observations:						
RMSE	0.497	3.204	0.492	0.581	1.320	0.675
Mean bias	0.094	3.172	0.008	-0.395	1.157	0.091
Median abs. error	0.335	3.163	0.321	0.440	1.143	0.438
Median bias	0.079	3.163	0.011	-0.410	1.143	0.081
Observations	300.0	191.9	190.3	300.0	191.9	190.3
400 Observations:						
RMSE	0.422	3.189	0.416	0.549	1.307	0.599
Mean bias	0.069	3.165	-0.015	-0.411	1.179	0.053
Median abs. error	0.286	3.161	0.284	0.417	1.180	0.398
Median bias	0.071	3.161	-0.015	-0.405	1.180	0.071
Observations	400.0	255.9	254.1	400.0	255.9	254.1

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Using all observations including the censored observations.
[b] Using only non-censored observations.
[c] Using CQR-UCP estimator.

TABLE VI

Monte Carlo Simulation for .50 Quantile Regression
Fixed Censoring ($\varphi_0(X) = y^0$), $\sigma = 1$

	CQR[a]	*UCP*[b]	*CQR*[a]	*UCP*[b]
	Constant		*Slope*	
50 Observations:				
RMSE	0.244	0.244	0.415	0.369
Mean bias	-0.015	0.017	0.019	-0.062
Median abs. error	0.167	0.164	0.256	0.244
Median bias	-0.016	0.012	-0.023	-0.067
Observations	48.2	44.7	48.2	44.7
100 Observations:				
RMSE	0.172	0.173	0.267	0.249
Mean bias	0.004	0.024	0.013	-0.041
Median abs. error	0.116	0.115	0.176	0.172
Median bias	0.003	0.023	0.001	-0.043
Observations	97.1	89.4	97.1	89.4
200 Observations:				
RMSE	0.116	0.117	0.184	0.178
Mean bias	-0.004	0.011	-0.004	-0.044
Median abs. error	0.076	0.077	0.117	0.117
Median bias	-0.002	0.013	-0.008	-0.044
Observations	195.0	178.5	195.0	178.5
300 Observations:				
RMSE	0.099	0.099	0.153	0.150
Mean bias	-0.000	0.013	0.002	-0.034
Median abs. error	0.066	0.066	0.101	0.102
Median bias	-0.000	0.013	-0.002	-0.037
Observations	292.5	267.8	292.5	267.8
400 Observations:				
RMSE	0.083	0.084	0.127	0.126
Mean bias	0.001	0.012	0.000	-0.031
Median abs. error	0.056	0.057	0.085	0.086
Median bias	0.000	0.012	-0.004	-0.034
Observations	390.5	356.9	390.5	356.9

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Powell estimator.
[b] CQR-UCP estimator.

TABLE VII

Monte Carlo Simulation for .50 Quantile Regression: Fixed Censoring ($\varphi_0(X) = y^0$), $\sigma = 2$

	CQR[a]	*UCP*[b]	*CQR*[a]	*UCP*[b]
	Constant		*Slope*	
50 Observations:				
RMSE	0.768	0.461	1.007	0.717
Mean bias	-0.071	0.068	0.128	-0.040
Median abs. error	0.324	0.304	0.506	0.465
Median bias	-0.039	0.075	-0.002	-0.084
Observations	46.7	37.5	46.7	37.5
100 Observations:				
RMSE	0.350	0.337	0.611	0.486
Mean bias	-0.022	0.053	0.066	-0.047
Median abs. error	0.236	0.224	0.365	0.326
Median bias	-0.012	0.054	-0.003	-0.066
Observations	95.2	75.0	95.2	75.0
200 Observations:				
RMSE	0.247	0.244	0.401	0.350
Mean bias	0.001	0.045	0.025	-0.047
Median abs. error	0.165	0.164	0.243	0.232
Median bias	0.003	0.047	-0.003	-0.056
Observations	193.0	150.4	193.0	150.4
300 Observations:				
RMSE	0.193	0.196	0.314	0.281
Mean bias	0.006	0.042	0.013	-0.047
Median abs. error	0.133	0.133	0.202	0.191
Median bias	0.004	0.038	-0.007	-0.054
Observations	290.9	226.0	290.9	226.0
400 Observations:				
RMSE	0.167	0.168	0.269	0.247
Mean bias	-0.002	0.029	0.003	-0.044
Median abs. error	0.111	0.113	0.170	0.167
Median bias	-0.004	0.031	-0.016	-0.049
Observations	388.8	300.7	388.8	300.7

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Powell estimator.
[b] CQR-UCP estimator.

TABLE VIII

Monte Carlo Simulation for .50 Quantile Regression: Fixed Censoring ($\varphi_0(X) = y^0$), $\sigma = 3$

	CQR[a]	*UCP*[b]	*CQR*[a]	*UCP*[b]
	Constant		*Slope*	
50 Observations:				
RMSE	2.416	0.733	2.049	1.123
Mean bias	-0.266	0.118	0.307	0.010
Median abs. error	0.498	0.461	0.764	0.621
Median bias	-0.076	0.115	-0.020	-0.093
Observations	44.9	33.1	44.9	33.1
100 Observations:				
RMSE	1.487	0.502	1.164	0.715
Mean bias	-0.108	0.085	0.136	-0.040
Median abs. error	0.348	0.339	0.529	0.450
Median bias	-0.049	0.081	-0.015	-0.083
Observations	93.3	66.8	93.3	66.8
200 Observations:				
RMSE	0.379	0.357	0.665	0.510
Mean bias	-0.035	0.059	0.058	-0.051
Median abs. error	0.242	0.242	0.375	0.332
Median bias	-0.032	0.056	-0.016	-0.066
Observations	190.2	133.9	190.2	133.9
300 Observations:				
RMSE	0.297	0.288	0.514	0.416
Mean bias	-0.009	0.058	0.036	-0.045
Median abs. error	0.201	0.197	0.306	0.287
Median bias	-0.004	0.059	-0.007	-0.067
Observations	287.7	201.6	287.7	201.6
400 Observations:				
RMSE	0.251	0.253	0.419	0.365
Mean bias	-0.004	0.051	0.015	-0.051
Median abs. error	0.165	0.169	0.258	0.249
Median bias	0.004	0.052	-0.020	-0.065
Observations	386.0	269.1	386.0	269.1

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.

[a] Powell estimator.

[b] CQR-UCP estimator.

TABLE IX

Monte Carlo Simulation for .50 Quantile Regression: Fixed Censoring ($\varphi_0(X) = y^0$), $\sigma = 4$

	CQR[a]	*UCP*[b]	*CQR*[a]	*UCP*[b]
	Constant		*Slope*	
50 Observations:				
RMSE	2.847	1.506	2.979	1.878
Mean bias	-0.401	0.201	0.466	0.125
Median abs. error	0.667	0.634	1.000	0.816
Median bias	-0.061	0.244	-0.046	-0.046
Observations	43.7	30.3	43.7	30.3
100 Observations:				
RMSE	1.785	0.663	1.858	0.978
Mean bias	-0.220	0.149	0.271	0.020
Median abs. error	0.465	0.446	0.751	0.608
Median bias	-0.055	0.126	-0.047	-0.069
Observations	90.9	61.6	90.9	61.6
200 Observations:				
RMSE	0.568	0.469	0.958	0.669
Mean bias	-0.041	0.109	0.129	-0.024
Median abs. error	0.329	0.317	0.505	0.427
Median bias	-0.012	0.100	0.012	-0.057
Observations	187.5	124.4	187.5	124.4
300 Observations:				
RMSE	0.433	0.391	0.716	0.536
Mean bias	-0.029	0.081	0.065	-0.048
Median abs. error	0.266	0.262	0.404	0.353
Median bias	-0.019	0.076	-0.003	-0.065
Observations	284.7	187.2	284.7	187.2
400 Observations:				
RMSE	0.360	0.332	0.590	0.471
Mean bias	-0.026	0.063	0.049	-0.042
Median abs. error	0.226	0.220	0.347	0.308
Median bias	-0.026	0.066	-0.012	-0.057
Observations	381.9	249.5	381.9	249.5

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.

[a] Powell estimator.

[b] CQR-UCP estimator.

TABLE X

Monte Carlo Simulation for .50 Quantile Regression: Fixed Censoring ($\varphi_0(X) = y^0$), $\sigma = 5$

	CQR[a]	*UCP*[b]	*CQR*[a]	*UCP*[b]
	Constant		*Slope*	
50 Observations:				
RMSE	4.667	3.187	4.140	3.126
Mean bias	-0.685	0.238	0.555	0.294
Median abs. error	0.831	0.800	1.209	1.091
Median bias	-0.110	0.386	-0.069	-0.002
Observations	42.3	28.0	42.3	28.0
100 Observations:				
RMSE	3.455	1.075	2.587	1.351
Mean bias	-0.321	0.217	0.311	0.068
Median abs. error	0.593	0.540	0.914	0.784
Median bias	-0.083	0.204	-0.060	-0.074
Observations	89.2	57.6	89.2	57.6
200 Observations:				
RMSE	2.107	0.591	1.757	0.832
Mean bias	-0.201	0.141	0.175	-0.009
Median abs. error	0.422	0.409	0.647	0.537
Median bias	-0.054	0.142	0.010	-0.062
Observations	183.6	117.2	183.6	117.2
300 Observations:				
RMSE	0.619	0.482	0.983	0.679
Mean bias	-0.069	0.105	0.082	-0.049
Median abs. error	0.344	0.331	0.527	0.464
Median bias	-0.040	0.099	-0.054	-0.102
Observations	281.5	176.9	281.5	176.9
400 Observations:				
RMSE	0.513	0.420	0.858	0.607
Mean bias	-0.036	0.101	0.095	-0.028
Median abs. error	0.295	0.294	0.465	0.407
Median bias	-0.019	0.110	-0.026	-0.073
Observations	377.1	237.7	377.1	237.7

Note: The true value of the constant is 1, and the true value of the slope is 1. The simulation is performed for 3000 repetitions.
[a] Powell estimator.
[b] CQR-UCP estimator.

Appendix—Proofs of Lemmas and Corollary

Proof of Lemma 1

Let

$$\begin{aligned}\xi(Z_i, f, A) &\equiv D_i X_i I(X_i \in \mathcal{X}) \Lambda\left(\pi(X_i)\right)\left(\pi(X_i) - I(\epsilon_{\theta i} < 0)\right) \\ &\equiv \lambda_1(\pi(X_i)) D_i X_i I(X_i \in \mathcal{X}) - \lambda_2(\pi(x)) D_i X_i I(X_i \in \mathcal{X}) I(\epsilon_{\theta i} < 0),\end{aligned}$$

where $\lambda_1(\pi) = \Lambda(\pi)\pi$, $\lambda_2(\pi) = \Lambda(\pi)$, and $\pi = 1 - (1-\theta) f_X / A$. We consider the asymptotic distribution of

$$n^{-1/2} \sum_{i=1}^{n} \xi(Z_i, \hat{f}, \hat{A})$$

using Newey and McFadden (1994, Theorem 8.11). Because Assumptions ER and K are satisfied, it suffices to show the existence of a vector of functionals $\Xi(z, f, A)$ which is linear in f and A such that:

A. For f and A with $\|(f, A) - (f_X, A_0)\|$ small, there exists some $b(z)$ such that

$$|\xi(z, f, A) - \xi(z, f_X, A_0) - \Xi(z, f - f_X, A - A_0)| = b(z) \|(f, A) - (f_X, A_0)\|^2$$

with $\mathrm{E}[b(Z_i)] < \infty$, where the norm $\|\gamma\|$ of γ is defined as

$$\|\gamma\| = \sup_{x \in \mathcal{X}} |\gamma(x)|.$$

B. There exists $c(\cdot)$ such that

$$|\Xi(z, f, A)| \leq c(z) \|(f, A)\|,$$

with $\mathrm{E}[|c(Z_i)|^2] < \infty$.

C. There exist $\nu_1(x)$ and $\nu_2(x)$ with

$$\mathrm{E}[\Xi(Z_i, f, A)] = \int \Big(\nu_1(x) f(x) + \nu_2(x) A(x)\Big)\, dx.$$

D. ν_j is continuous almost everywhere, $\int |\nu_j(x)| dx < \infty$, and there is $\varepsilon > 0$ such that

$$\mathrm{E}[\sup_{|v| \leq \varepsilon} |\nu_j(X_i + v)|^4] < \infty, \quad j = 1, 2.$$

If Conditions A–D are satisfied then we can write

$$n^{-1/2} \sum_{i=1}^{n} \xi(Z_i, \hat{f}, \hat{A}) =$$

$$n^{-1/2} \sum_{i=1}^{n} \Big(\xi(Z_i, f_X, A_0) + \nu_1(X_i) + \nu_2(X_i) D_i - \mathrm{E}\left[\nu_1(X_i) + \nu_2(X_i) D_i\right]\Big) + o_p(1).$$

To verify Condition A, we note that for a fixed Z_i,

$$\begin{aligned}
\lambda_j(\pi)-\lambda_j(\pi_\theta) &= (1-\theta)\lambda_j'(\pi_\theta)\left(-\frac{1}{A_0}(f-f_X)+\frac{1}{\pi_\theta A_0}(A-A_0)\right) \\
&\quad+\frac{1}{2}\lambda_j''(\bar\pi)\left(\frac{1-\theta}{\bar A}\right)^2(f-f_X)^2 \\
&\quad+\left(\lambda_j'(\bar\pi)\frac{1-\theta}{\bar A^2}-\lambda_j''(\bar\pi)\frac{(1-\theta)^2\bar f}{\bar A^3}\right)(f-f_X)(A-A_0) \\
&\quad+\frac{1}{2}\left(-2\lambda_j'(\bar\pi)\frac{(1-\theta)\bar f}{\bar A^3}+\lambda_j''(\bar\pi)\frac{(1-\theta)^2\bar f^2}{\bar A^4}\right)(A-A_0)^2,
\end{aligned}$$

for $(\bar f,\bar A)$ on the line segment adjoining (f_X,A_0) and (f,A), and $\bar\pi=1-(1-\theta)\bar f/\bar A$. Using the fact that the first and second derivatives of $\Lambda(\cdot)$ are bounded and A_0 is bounded away from zero on $\mathcal{X}$, we obtain

$$|\xi(Z_i,f,A)-\xi(Z_i,f_X,A_0)-\Xi(Z_i,f-f_X,A-A_0)|\le C|X_i|\|(f,A)-(f_X,A_0)\|^2$$

for some constant C, and

$$\begin{aligned}
&\Xi(z,f-f_X,A-A_0)= \\
&\quad-(1-\theta)D_iX_iI(X_i\in\mathcal{X})\left(\lambda_1'(\pi_\theta(X_i))-I(\epsilon_{\theta i}<0)\lambda_2'(\pi_\theta(X_i))\right)\frac{1}{A_0(X_i)}\left(f(X_i)-f_X(X_i)\right) \\
&\quad+(1-\theta)D_iX_iI(X_i\in\mathcal{X})\left(\lambda_1'(\pi_\theta(X_i))-I(\epsilon_{\theta i}<0)\lambda_2'(\pi_\theta(X_i))\right)\frac{f_X(X_i)}{A_0^2(X_i)}\left(A(X_i)-A_0(X_i)\right).
\end{aligned}$$

Since the first derivative of $\Lambda(\cdot)$ is bounded, Condition B is also satisfied.

To verify Condition C, notice that

$$\begin{aligned}
&\mathrm{E}\Big[D_iI(X_i\in\mathcal{X})X_i\left(\lambda_1'(\pi_\theta(X_i))-I(\epsilon_{\theta i}<0)\lambda_2'(\pi_\theta(X_i))\right)\Big|\,X_i=x\Big] \\
&= I(x\in\mathcal{X})h_0(x)\lambda_1'(\pi_\theta(x))x-I(x\in\mathcal{X})\lambda_2'(\pi_\theta(x))\left(h_0(x)-1+\theta\right)x \\
&= xI(x\in\mathcal{X})h_0(x)\Lambda(\pi_\theta(x)) \\
&\equiv \mu(x).
\end{aligned}$$

From the fact that

$$\mathrm{E}[\Xi(Z_i,f,A)]=\int\Big(-(1-\theta)\mu(x)\frac{1}{A_0(x)}f(x)f_X(x)+(1-\theta)\mu(x)\frac{f_X(x)}{A_0^2(x)}A(x)f_X(x)\Big)\,dx,$$

we obtain that

$$\nu_1(X_i)=-(1-\theta)\frac{\mu(X_i)}{h_0(X_i)}=-(1-\theta)X_iI(X_i\in\mathcal{X})\Lambda(\pi_\theta(X_i))$$

and

$$\nu_2(X_i)=(1-\theta)\frac{\mu(X_i)}{h_0^2(X_i)}=(1-\theta)\frac{1}{h_0(X_i)}X_iI(X_i\in\mathcal{X})\Lambda(\pi_\theta(X_i)).$$

Consequently Condition C follows.

Notice that Condition D is satisfied because X_i has a finite fourth moment (Assumption ER(vi)), $\Lambda(\cdot)$ is bounded, and $h_0(\cdot)$ is bounded away from zero.

Note that since

$$\mathrm{E}\left[\nu_1(X_i)+\nu_2(X_i)D_i\right]=(1-\theta)\mathrm{E}\left[\frac{D_i-h_0(X_i)}{h_0(X_i)}X_i I(X_i\in\mathcal{X})\Lambda(\pi_\theta(X_i))\right]=0$$

we obtain that

$$n^{-1/2}\sum_{i=1}^{n}\xi(Z_i,\hat{f},\hat{A})=$$
$$n^{-1/2}\sum_{i=1}^{n}\left(\xi(Z_i,f_X,A_0)+(1-\theta)\frac{D_i-h_0(X_i)}{h_0(X_i)}X_i I(X_i\in\mathcal{X})\Lambda(\pi_\theta(X_i))\right)+o_p(1),$$

from which the lemma follows. **Q.E.D.**

Proof of Lemma 2

We first consider $\sum_{i=1}^{n}\left[R_n(Z_i,\hat{\pi}_\theta,\tau)-R_n(Z_i,\pi_\theta,\tau)\right]$.

Observe that

$$\begin{aligned}&D_i I\left(X_i\in\mathcal{X}\right)\left(\pi(X_i)[(\epsilon_{\theta i}-n^{-1/2}X_i'\tau)^{+}-\epsilon_{\theta i}^{+}]+(1-\pi(X_i))[(\epsilon_{\theta i}-n^{-1/2}X_i'\tau)^{-}-\epsilon_{\theta i}^{-}]+\tau'\zeta(Z_i,\pi)\right)\\&\quad=D_i I\left(X_i\in\mathcal{X}\right)(n^{-1/2}X_i'\tau-\epsilon_{\theta i})\left(I\left(0<\epsilon_{\theta i}<n^{-1/2}X_i'\tau\right)-I\left(n^{-1/2}X_i'\tau<\epsilon_{\theta i}<0\right)\right)\end{aligned}$$

does not depend on π, and is bounded by

$$n^{-1/2}|X_i'\tau|I\left(|\epsilon_{\theta i}|<n^{-1/2}|X_i'\tau|\right).$$

Thus, we may write

$$R_n(Z_i,\hat{\pi}_\theta,\tau)-R_n(Z_i,\pi_\theta,\tau)=S_n(Z_i,\tau)\left[\Lambda(\hat{\pi}_\theta(X_i)-\Lambda(\pi_\theta(X_i))\right],$$

where

$$S_n(Z_i,\tau)\equiv D_i I\left(X_i\in\mathcal{X}\right)(n^{-1/2}X_i'\tau-\epsilon_{\theta i})\left(I\left(0<\epsilon_{\theta i}<n^{-1/2}X_i'\tau\right)-I\left(n^{-1/2}X_i'\tau<\epsilon_{\theta i}<0\right)\right).$$

We then have

$$\begin{aligned}\left|\sum_{i=1}^{n}\left[R_n(Z_i,\hat{\pi}_\theta,\tau)-R_n(Z_i\pi_\theta,\tau)\right]\right| &\le \sum_{i=1}^{n}|S_n(Z_i,\tau)|\cdot|\Lambda(\hat{\pi}_\theta(X_i))-\Lambda(\pi_\theta(X_i))|\\&\le \sup_{x\in\mathcal{X}}|\Lambda(\hat{\pi}_\theta)-\Lambda(\pi_\theta)|\cdot\sum_{i=1}^{n}|S_n(Z_i,\tau)|\\&\le C\sup_{x\in\mathcal{X}}|\hat{\pi}_\theta-\pi_\theta|\cdot\sum_{i=1}^{n}|S_n(Z_i,\tau)|,\end{aligned}$$

where C is the bound of the first derivative of $\Lambda(\cdot)$.

Note now that because

$$\sup_{x \in \mathcal{X}} |\hat{\pi}_\theta - \pi_\theta| = o_p(1),$$

by Newey and McFadden (1994, Lemma 8.10), and because

$$|S_n(Z_i, \tau)| \leq n^{-1/2}|X_i'\tau| I\left(|\epsilon_{\theta i}| < n^{-1/2}|X_i'\tau|\right),$$

it suffices to establish that

$$n^{-1/2}\sum_{i=1}^{n} |X_i'\tau| I\left(|\epsilon_{\theta i}| < n^{-1/2}|X_i'\tau|\right) = O_p(1).$$

Let $U_i = |X_i'\tau|$ and let $a_n = n^{-1/2}$. Then it suffices to show that

$$n^{-1/2}\sum_{i=1}^{n} U_i I\left(|\epsilon_{\theta i}| < a_n U_i\right) = O_p(1).$$

Notice that the family of functions $UI\left(|\epsilon| < aU\right)$ indexed by a satisfies Pollard's entropy condition (See Andrews (1989)), and thus satisfies stochastic equicontinuity:

$$n^{-1/2}\sum_{i=1}^{n} |U_i| I\left(|\epsilon_{\theta i}| < a_n|U_i|\right) = n^{1/2}\mathrm{E}[|U_i| I\left(|\epsilon_{\theta i}| < a_n|U_i|\right)] + o_p(1).$$

Letting $F_{\epsilon_\theta}(\cdot|x)$ denote the conditional cumulative distribution function of $\epsilon_{\theta i}$ given $X_i = x$, we have

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n} |U_i| I\left(|\epsilon_{\theta i}| < a_n|U_i|\right) &= n^{1/2}\mathrm{E}\Big[|U_i|[F_{\epsilon_\theta}(a_n|U_i| \;\Big|\; X_i) - F_{\epsilon_\theta}(-a_n|U_i| \;\Big|\; X_i)]\Big] + o_p(1) \\ &= n^{1/2}a_n 2\mathrm{E}\Big[f_{\epsilon_\theta}(0|X_i)|X_i|\;|U_i|\Big] + o_p(1) \\ &= O_p(1). \end{aligned}$$

It follows then that the term

$$\sum_{i=1}^{n} [R_n(Z_i, \hat{\pi}_\theta, \tau) - R_n(Z_i, \pi_\theta, \tau)]$$

contributes $o_p(1)$ to $G_n(\tau, \hat{\pi}_\theta)$.

Consider now the sum $\sum_{i=1}^{n} T_{n,i} = \sum_{i=1}^{n} [R_n(Z_i, \pi_\theta, \tau) - \mathrm{E}[R_n(Z_i, \pi_\theta, \tau)]]$. Because of the cancellation of cross-product terms, we get

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} T_{n,i}\right)^2\right] &\leq \sum_{i=1}^{n} \mathrm{E}[R_n(Z_i, \pi_\theta, \tau)^2] \\ &\leq \mathrm{E}\left[|X_i'\tau|^2 I\left(|\epsilon_{\theta i}| < n^{-1/2}|X_i'\tau|\right)\right] \\ &= o(1). \end{aligned}$$

Thus, the term $\sum_{i=1}^{n} T_{n,i}$ also contributes $o_p(1)$, and the conclusion follows. **Q.E.D.**

Proof of Lemma 3

The arguments made in this Appendix are similar to Pollard's arguments in Pollard (1991, Theorem 1, p.192)). We present this lemma for completeness since minor changes are made.

The proof uses the following definition of norm $\| \cdot \|$:

Definition (norm): Let $t \in R^k$ and let J be a matrix of suitable dimensions. Then, $\| t \| \equiv t'Jt$.

Given this definition we can write

$$G_n(\tau, \hat{\pi}) = \frac{1}{2} \| \tau - \eta_n \|^2 - \frac{1}{2} \| \eta_n \|^2 + r_n(\tau),$$

where for each compact set $T \subset R^k$,

$$\sup_{\tau \in T} \| r_n(\tau) \| = o_p(1).$$

We need to show that for each $\varepsilon > 0$

$$\Pr\{\| \tau_n - \eta_n \| > \varepsilon\} \to 0.$$

This is a consequence of two elements: (a) the convexity of G_n; and (b) the behavior of r_n in a small close neighborhood of η_n.

Let $B(n)$ denote the closed ball (with respect to the norm $\| \cdot \|$) with center η_n and radius ε. Because η_n converges in distribution, it is stochastically bounded. The compact set T can be chosen to contain $B(n)$ with probability arbitrarily close to one, thereby implying that

$$\Delta_n \equiv \sup_{\tau \in B(n)} \| r_n(\tau) \| = o_p(1).$$

Now consider the behavior of G_n outside $B(n)$. Suppose $\tau = \eta_n + \mu v$, with $\mu > \varepsilon$ and v a unit vector (with respect to the $\| \cdot \|$ norm). Define τ^* as the boundary point of $B(n)$ that lies on the line segment from η_n to τ, that is, $\tau^* = \eta_n + \varepsilon v$. The convexity of G_n and the definition of Δ_n imply

$$\begin{aligned}
\frac{\varepsilon}{\mu} G_n(\tau) + \left(1 - \frac{\varepsilon}{\mu}\right) G_n(\eta_n) &\geq G_n(\tau^*) \\
&\geq \frac{1}{2}\varepsilon^2 - \frac{1}{2} \| \eta_n \|^2 - \Delta_n \\
&\geq \frac{1}{2}\varepsilon^2 - G_n(\eta_n) - 2\Delta_n.
\end{aligned}$$

This last expression does not depend on τ. It follows that

$$\inf_{\|\tau - \eta_n\| > \varepsilon} G_n(\tau, \hat{\pi}_\theta) \geq G_n(\eta_n, \hat{\pi}_\theta) + \frac{\mu}{\varepsilon}\left[\frac{1}{2}\varepsilon^2 - 2\Delta_n\right].$$

When $2\Delta_n < \frac{1}{2}\varepsilon^2$, which occurs with probability tending to one, the minimum of G_n cannot occur at any τ with $\| \tau - \eta_n \| > \varepsilon$; thus with probability tending to one, $\| \tau - \eta_n \| \leq \varepsilon$ as required.

Q.E.D.

Proof of Corollary 1

It follows from Theorem 1 that the asymptotic covariance for $\hat{\beta}_\theta$ is given by

$$\Omega_\theta = \text{Va}(\hat{\beta}) = J^{-1}\text{E}(W_i W_i')J^{-1}, \tag{23}$$

where

$$W_i = I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))\left[D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right) + (1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)}\right]X_i$$

and

$$J \equiv \text{E}\left[I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))f_{\epsilon_\theta}(0 \mid X_i)X_i X_i'\right].$$

Note that by the law of iterated expectation

$$\text{E}(W_i W_i') = \text{E}\left[\text{E}\left(W_i W_i' \mid X_i, \pi_\theta(X_i) > 0\right)\right]. \tag{24}$$

We will first compute the inner expectation in (24):

$$\begin{aligned}
&\text{E}\left(W_i W_i' \mid X_i, \pi_\theta(X_i) > 0\right) \\
&= X_i X_i' I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i)) \\
&\quad \text{E}\left[\left(D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right) + (1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)}\right)^2 \mid X_i, \pi_\theta(X_i) > 0\right] \\
&= X_i X_i' I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))\Bigg[\text{Var}\left(D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right) \mid X_i, \pi_\theta(X_i) > 0\right) \\
&\quad +\text{Var}\left((1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)} \mid X_i, \pi_\theta(X_i) > 0\right) \\
&\quad +2\text{Cov}\left(D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right), (1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)} \,\Big|\, X_i, \pi_\theta(X_i) > 0\right)\Bigg] \\
&= X_i X_i' I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))\left(V_1 + V_2 + 2CV\right),
\end{aligned} \tag{25}$$

where

$$V_1 \equiv \text{Var}\left(D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right) \,\Big|\, X_i, \pi_\theta(X_i) > 0\right), \tag{26}$$

$$V_2 \equiv \text{Var}\left((1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)} \,\Big|\, X_i, \pi_\theta(X_i) > 0\right) \tag{27}$$

and

$$CV \equiv \text{Cov}\left(D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right), (1-\theta)\frac{D_i - h_0(X_i)}{h_0(X_i)} \,\Big|\, X_i, \pi_\theta(X_i) > 0\right). \tag{28}$$

Recall that

$$h_0(X_i) = \Pr\left(D_i = 1 \mid X_i = X_i\right)$$

and note that

$$\pi_\theta(X_i) > 0 \quad \text{if and only if} \quad h_0(X_i) > 1 - \theta.$$

Using these facts, V_1 in (26) can be rewritten as

$$\begin{aligned}
V_1 &= \mathrm{E}\left[D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right)^2 \mid X_i, h_0(X_i) > 1-\theta\right] \\
&= \mathrm{E}\left[\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right)^2 \mid X_i, h_0(X_i) > 1-\theta, D_i = 1\right] \Pr\left(D_i = 1 \mid X_i, h_0(X_i) > 1-\theta\right) \\
&= \mathrm{Var}\left(I(\epsilon_{\theta i} < 0) \mid X_i, h_0(X_i) > 1-\theta, D_i = 1\right) h_0(X_i) \\
&= (1-\theta)\frac{h_0(X_i) - (1-\theta)}{h_0(X_i)}. \qquad (29)
\end{aligned}$$

Similarly V_2 in (27) can be rewritten as

$$\begin{aligned}
V_2 &= \frac{(1-\theta)^2}{(h_0(X_i))^2}\mathrm{Var}\left(D_i \mid X_i, h_0(X_i) > 1-\theta\right) \\
&= (1-\theta)^2\frac{1-h_0(X_i)}{h_0(X_i)}. \qquad (30)
\end{aligned}$$

For the term CV in (28) we have

$$\begin{aligned}
CV &= \mathrm{E}\left[(1-\theta)D_i\left(\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0)\right)\frac{D_i - h_0(X_i)}{h_0(X_i)} \;\Big|\; X_i, h_0(X_i) > 1-\theta\right] \\
&\quad \times \Pr\left(D_i = 1 \mid X_i, h_0(X_i) > 1-\theta\right) \\
&= (1-\theta)\frac{1-h_0(X_i)}{h_0(X_i)}\mathrm{E}\left[\pi_\theta(X_i) - I(\epsilon_{\theta i} < 0) \mid X_i, h_0(X_i) > 1-\theta\right] h_0(X_i) \\
&= 0. \qquad (31)
\end{aligned}$$

Substituting equations (29), (30) and (31) into (25) yields

$$\mathrm{E}\left[W_iW_i' \mid X_i, \pi_\theta(X_i) > 0\right] = \theta(1-\theta)X_iX_i'I(X_i \in \mathcal{X})\Lambda\left(\pi_\theta(X_i)\right)$$

and therefore for $\mathrm{E}(W_iW_i')$ in (24) we have

$$\mathrm{E}(X_iX_i') = \theta(1-\theta)\mathrm{E}\left[X_iX_i'I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))\right] = \theta(1-\theta)\Delta,$$

where

$$\Delta = \mathrm{E}\left[I(X_i \in \mathcal{X})\Lambda(\pi_\theta(X_i))X_iX_i'\right].$$

For the variance of $\hat{\beta}_\theta$ in (23) we have finally

$$\Omega_\theta = \theta(1-\theta)J^{-1}\Delta J^{-1}.$$

Q.E.D.

7 References

Andrews, D.W.K. (1989): "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity," Unpublished manuscript, Yale University.

Andrews, D.W.K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 64, 43–72.

Buchinsky, M. (1995): "Quantile Regression, Box-Cox Transformation Model, and the U.S. Wage Structure, 1963-1987," *Journal of Econometrics*, 109–154.

Buchinsky, M. (1995): "A Monte Carlo Study of the Asymptotic Covariance Matrix Estimators for Quantile Regression Estimator," *Journal of Econometrics*, forthcoming.

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

Chamberlain, G. (1991): "Quantile Regression, Censoring, and the Structure of Wages," Harvard Institute of Economic Research, Discussion Paper No. 1558.

Hahn, J. (1994): "Bootstrapping Quantile Regression Models," Unpublished manuscript, University of Pennsylvania.

Honoré, B. and J.L. Powell (1993): "Quantile Regression Under Random Censoring," Unpublished manuscript, Princeton University.

Honoré, B. and J.L. Powell (1993): "Pairwise Difference Estimators of Linear, Censored, and Truncated Regression Models," *Journal of Econometrics*, forthcoming.

Horowitz, J.L. (1988): "Semiparametric M-Estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45–83.

Horowitz, J.L. (1986): "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models," *Journal of Econometrics*, 32, 59–84.

Koenker, R. and B. Park (1993): "An Interior Point Algorithm for Nonlinear Quantile Regression," Unpublished manuscript, University of Illinois, Champaign.

Manski, C. (1988): *Analog Estimation Methods in Econometrics*, New York: Chapman and Hall.

Moon, C-G. (1989): "A Monte Carlo Comparison of Semiparametric Tobit Estimators," *Journal of Applied Econometrics*, 4, 361–382.

Nadaraya, E.A. (1965): "On Nonparametric Estimation of Density Function and Regression," *Theory of Probability and its Application*, 10, 186–190.

Nawata, K. (1990): "Robust Estimation Based on Group-Adjusted Data in Censored Regression Models," *Journal of Econometrics*, 43, 337–362.

Newey, W.K. (1989): "Efficient Estimation of Tobit Models Under Symmetry, in Barnett, W.A., J.L. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press.

Newey, W.K. and D. McFadden (1994): "Estimation and Hypothesis Testing in Large Samples," in Engle, R.F. and D. McFadden (eds.), *Handbook of Econometrics*, Vol. IV.

Newey, W.K. and J.L. Powell (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

Oberhofer, W. (1982): "The Consistency of Nonlinear Regression Minimizing the L_1 Norm," *Annals of Statistics*, 10, 316–319.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimator," *Econometric Theory*, 7, 186–199.

Powell, J.L. (1991): "Estimation of Monotonic Regression Models under Quantile Restrictions, in Barnett, W., J.L. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press.

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

Powell, J.L. (1986b): "Symmetrically Trimmed Least Squares Estimation of Tobit Models," *Econometrica*, 1435–1460.

Powell, J.L. (1984): "Least Absolute Deviation Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

Silverman, B.W. (1986): *Density Estimation*, London: Chapman and Hall.

Tobin, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.

Watson, G.S. (1964): "Smooth Regression Analysis," *Sankhyā*, Series A, 26, 359–372.