Three-Step Censored Quantile Regression and Extramarital Affairs

Victor CHERNOZHUKOV and Han HONG

This article suggests very simple three-step estimators for censored quantile regression models with a separation restriction on the censoring probability. The estimators are theoretically attractive (i.e., asymptotically as efficient as the celebrated Powell's censored least absolute deviation estimator). At the same time, they are conceptually simple and have trivial computational expenses. They are especially useful in samples of small size or models with many regressors, with desirable finite-sample properties and small bias. The separation restriction costs a small reduction of generality relative to the canonical censored regression quantile model, yet its main plausible features remain intact. The estimator can also be used to estimate a large class of traditional models, including the normal Amemiya–Tobin model and many accelerated failure and proportional hazard models. We illustrate the approach with an extramarital affairs example and contrast our findings with those of Fair.

KEY WORDS: Accelerated failure time model; Classification; Discriminant analysis; Fixed censoring; Median regression; Proportional hazard model; Quantile regression; Robustness.

1. INTRODUCTION

Econometrics and statistics have devoted substantial attention to censored data. This article analyzes censored quantile regression (CQR) models with known censoring points, suggesting a simple, easily implementable, and well-behaved three-step estimation procedure. This is achieved by exploring the structured envelope and separation restrictions on the censoring probability. These restrictions preserve the plausible semiparametric, distribution-free, and heteroscedastic features of the model. We illustrate the procedure with an extramarital affairs example.

1.1 Censored Quantile Regression Model

CQR models motivate the main effort of this article. The CQR models allow covariates to shift location, scale, and the entire shape of the distribution and permit distribution-free specifications. As such, CQR models compare favorably to the normal Amemiya–Tobin, Cox, Buckley–James, and other approaches. (See Horowitz and Neumann 1987 and Koenker and Gelling 2001 for excellent expositions.)

Quantile regression research began in the 1970s. In their pathbreaking work, Koenker and Bassett (1978) introduced the general quantile regression (QR) estimation that became the most popular approach. Lehmann (1974) and Doksum (1974) formulated the quantile inference paradigm for the p-sample setting, arguing that location-shift models are insufficient to summarize ubiquitous quantile shift effects. Hogg (1975) suggested instrumental variable–type estimators. Many other works also laid the foundation, including those of Amemiya (1981), Powell (1986), Koenker and Portnoy (1987), Chaudhuri (1991), Portnoy (1991), Jureckova and Prochazka (1994), Chaudhuri, Doksum, and Samarov (1997), Buchinsky and Hahn (1998), Knight (1998), Koenker and Machado (1999), Portnoy and Jureckova (2000), Fitzenberger and Winker (2001), Khan and Powell (2001), Portnoy (2001), and Abadie, Angrist, and Imbens (2002).

The conditional quantile function of the dependent real variable Y given covariates X in $\mathbb{R}^d$, $Q_{Y|X}$, is the inverse of the conditional distribution function $F_{Y|X}$: $Q_{Y|X}(\tau) = \inf_{v\in\mathbb{R}}\{v : F_{Y|X}(v) > \tau\}$. The classical linear model of $Q_{Y|X}$,

$$Q_{Y|X}(\tau) = X'\beta(\tau), \tag{1}$$

is conceptually appealing, incorporating classical linear location-scale models as important special cases. We assume that X includes a constant and note that it may incorporate a wide array of polynomial and alternative transformations of the observable covariates.

Equivariance to monotone transformations is an important property of quantile regression models (see Powell 1986). For a given monotone transformation $\mathcal{T}_c(Y)$ of variable Y, which may depend on other variables C, $Q_{\mathcal{T}_c(Y)|X,C}(\tau) = \mathcal{T}_c(Q_{Y|X,C}(\tau))$. Transformation equivariance naturally leads to Powell's CQR model. In this model, the latent variable Y_i^* is left censored by the *observable*, possibly random, censoring points C_i, and we observe

$$Y_i = Y_i^* \vee C_i, \qquad X_i, \qquad C_i, \qquad \delta_i = 1(Y_i = C_i). \tag{2}$$

Y_i^* is assumed to be *conditionally independent of the censoring point* C_i; that is, for all $y \in \mathbb{R}$,

$$P(Y^* < y|X_i, C_i) = P(Y^* < y|X_i), \text{ so that } Q_{Y^*|X,C_i}(\tau) = X'\beta(\tau). \tag{3}$$

Conditioning on C_i, assumption (3) and the transformation equivariance yield the following CQR model (see Powell 1986):

$$Q_{Y_i|X_i,C_i}(\tau) = X_i'\beta(\tau) \vee C_i. \tag{4}$$

Victor Chernozhukov is Assistant Professor, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA-02142 (E-mail: *vchern@mit.edu*). Han Hong is Assistant Professor, Department of Economics, Princeton University, Princeton, NJ 08544 (E-mail: *doubleh@princeton.edu*). Initial results of this article were included in a 1998 Stanford University working paper titled "Censored Quantile Regression With Estimated Censoring Points." The authors thank Takeshi Amemiya, Roger Koenker, Stephen Portnoy, the editor, the associate editor, and the anonymous referees for very constructive, critical comments. They also thank Joshua Angrist, Jerry Hausman, Jinyong Hahn, Sendhil Mullainathan, and Whitney Newey for very helpful conversations. Of course, the usual disclaimers apply.

Journal of the American Statistical Association
September 2002, Vol. 97, No. 459, Theory and Methods
DOI 10.1198/016214502388618663

The assumption that censoring points are known for all i is realistic in many (but clearly not all) situations. For example, in the famous Stanford survival dataset, we can compute all censoring points, because we know the transplant and the last follow-up dates for each i. In the extramarital affairs example, the censoring point is 0, $C_i = 0$. In this article we do not consider unobserved censoring points; Zhou (1992), Yang (1997), Portnoy (2001), and Honoré, Khan, and Powell (2002) have done important work in this direction.

In fact, any model like (4) can be reduced to a model with a fixed censoring at 0. Subtracting C_i from Y_i, and letting $\widetilde{Y}_i = Y_i - C_i$, $\widetilde{X}_i = (X_i, C_i)$, and $\tilde{\beta}(\tau) = (\beta(\tau)', -1)'$, by equivariance,

$$Q_{\widetilde{Y}|X, C_i}(\tau) = X'\beta(\tau) \vee C_i - C_i = \widetilde{X}'\tilde{\beta}(\tau) \vee 0. \qquad (5)$$

In this model, one coefficient in front of the regressor C_i is known to be -1 and need not be estimated. Thus we may set $C_i = 0$ in the sequel.

1.2 The Estimation Problem

Suppose that we have n observations $\{Y_i, X_i\}$. Sample regression quantiles are defined as the solution to the problem

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta), \qquad (6)$$

where $\rho_\tau(x) \equiv (\tau - 1(x \leq 0))x$ (see Koenker and Bassett 1978). The median or least absolute deviation estimator is a very important special case with $\tau = 1/2$. In the censored model (4), replacement of the linear form with the partially linear form,

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta \vee C_i),$$

leads to the celebrated Powell estimator. Powell (1986) established the asymptotic normality of this estimator and developed an inference theory.

Despite its intuitive appeal, this estimation method has not become popular in empirical research, because of its well-known computational difficulty. We know of only a few applications of censored median regression in econometrics. In contrast, the Amemiya–Tobin and Cox approaches have found hundreds, if not thousands of applications, (see, e.g., a series of remarkable works on computations in Fitzenberger 1997a, b, Fitzenberger and Winker 2001). Buchinsky (1994) and Fitzenberger (1997a) designed ingenious computational algorithms, which Fitzenberger (1997a) recommended for low degrees of censoring while admitting that "all practical algorithms perform quite poorly when a lot of censoring is present." (see, e.g., Fitzenberger 1997a, p. 15). In the case of 50% censoring, one regressor, and small sample $n = 100$, in several important designs (e.g., A, B), the frequency of convergence to the Powell estimator ranged from 5% to 37% for various algorithms. For some of the other designs, results were better, with convergence frequencies ranging from 30% to 70%. These results were obtained for the case of one regressor. In case of many regressors and larger n, the results can be expected to be worse. Fitzenberger's conclusion is well substantiated by a very extensive Monte Carlo experiment with numerous different practical designs. All experiments involved only one regressor. Additional regressors only worsen the performance. In many empirical applications, the censoring is quite heavy and dimensionality is also high. For example, in the affairs example of Section 3, the degree of censoring is 68%, in the well-known Stanford transplant dataset it is 37%. The number of regressors in these two datasets are 9 and 3. Arguably, an important goal is the design of both theoretically elegant and implementable, practically attractive estimators. It is this requirement that makes the problem at hand particularly challenging.

Motivated in part by such limitations, recent remarkable work by Buchinsky and Hahn (1998) and Khan and Powell (2001) suggested a number of alternative estimators. Buchinsky and Hahn (1998) proposed first estimating the propensity score $h(X_i) = P(\delta_i = 1|X_i)$ by a nonparametric kernel regression, then selecting a subset of the whole sample, where $\{i : h(X_i) > 1 - \tau\}$, [i.e., those observations i where the conditional quantile line is above the censoring point 0, $X_i'\beta(\tau) > 0$], and then using a QR on the selected sample. Analogously, Khan and Powell (2001) proposed using any of the following three methods to perform the first-stage selection: maximum score estimators of the regression quantile, nonparametric kernel propensity score estimator for $h(X_i)$, and the nonparametric locally linear conditional quantile estimator of Chaudhuri (1991). The two-stage estimators are somewhat less efficient than the Powell estimator because of smoothing and trimming. Ideologically, these estimators share the ideas behind the construction of the Powell estimator, except that Powell imposed simultaneity to obtain his single-step estimator.

The suggested first stages are attractive but are practical only in low dimensions and have slow convergence rates. Local kernel smoothers apply to (sufficiently) continuous variables only, whereas many applications, including ours, have many (sufficiently) discrete covariates. This is very confounding. Of course, asymptotic theory suggests that the $\sqrt{n}$-consistent estimates could be obtained by averaging within the cells. In the affairs example, there is on average $6800/(8^5) \approx .2$ observations per cell; in the heart example, $69/(15 \times 2^2) \approx 1$. Thus such an asymptotic approach is precluded here. The computational burden is very substantial in high dimensions and large datasets for all of the first-stage estimates. As a result, we simply cannot use any of the available estimators in our example and many other real-life applications because of heavy censoring, high dimensionality, and the polychotomous nature of numerous regressors. From a constructive angle, however, we stress that the aforementioned estimators can be potentially fruitful in many cases.

2. SIMPLE THREE-STEP CENSORED QUANTILE REGRESSION ESTIMATORS

The current approach is based on the structured modeling restrictions that we put on the censoring probability. These restrictions do cause a reduction in generality, but only a small one, because they still incorporate the Amemiya–Tobin model, many Cox models, and accelerated failure time models as very important special cases while preserving the heteroscedasticity and distribution-free character. The end result is that an easily computable (comparable to linear least squares), well-behaved, robust estimator is available. It offers not only an efficient, practical way to estimate the general CQR models, but also a good way to estimate important traditional models.

2.1 The Procedure

This section describes the steps of the estimator.

Step 1. Estimate a parametric classification (probability) model,

$$\delta_i = p(\dot{X}_i'\gamma) + \epsilon_i,$$

where δ_i is the indicator of *not-censoring.* $\dot{X}_i$ indicates desired transforms of (X_i, C_i). Next, *select the sample* $J_0 = \{i : p(\dot{X}_i'\hat{\gamma}) > 1 - \tau + c\}$, where c is strictly between 0 and τ and not too small. The practical choice of p is discussed later. A sensible rule for choosing c is to compare the size of the selected sample $J(c) = \{i : p(\dot{X}'\hat{\gamma}) > 1 - \tau + c\}$ for $c = 0$ and other values. Choosing $c = q$th quantile of all $p(\dot{X}_i'\hat{\gamma})$ such that $p(\dot{X}_i'\hat{\gamma}) > 1 - \tau$ appears to be sound, because it gives a control of percentage of observations from $J(0)$ to be thrown out: $\#J(c)/\#J(0) = (1 - q) \times 100\%$. This rule, with $q = 10\%$, worked well in simulations. Another way to pick the trimming constant as well as a more suitable estimator among various k-step estimators is to evaluate them using the Powell criterion function. The one yielding the minimal value could be used. This suggestion was made by an anonymous referee.

Step 2. Obtain the initial (inefficient) estimator $\hat{\beta}_0(\tau)$ by the standard QR,

$$\min_{\beta} \sum_{i \in J_0} \rho_\tau(Y_i - X_i'\beta). \tag{7}$$

Next, select $J_1 = \{i : X_i'\hat{\beta}_0(\tau) > C_i + \delta_n\}$, where δ_n is a small positive number such that $\sqrt{n} \times \delta_n \to \infty$ and $\delta_n \searrow 0$. In practice, δ_n could be chosen like c, but the percentage of discarded observations should be smaller. Intuitively, this step asymptotically selects those observations that have covariate values (X_i, C_i) such that $X_i'\beta(\tau) > C_i$, *building up the efficiency of the next step.*

Step 3. Run QR (7) with J_1 in place of J_0. In empirical work, a sensible robustness diagnostic is to check whether $J_0 \subset J_I$. If a large proportion of observations J_0 are not in J_I, then one should revise the trimming constants and possibly also the separation models or the conditional quantile models in question.

Denote this three-step estimator by $\hat{\beta}_1(\tau)$.

Step 4. (Optional). Repeat step 3 one or more times, using sample $J_I = \{i : X_i'\hat{\beta}_{I-1}(\tau) > C_i + \delta_n\}$ in place of J_1. $[I = 2, 3, \ldots]$.

In step 4, each repetition involves selecting $J_I = \{i : X_i'\hat{\beta}_{I-1}(\tau) > C_i + \delta_n\}$, and then obtaining $\hat{\beta}_I(\tau)$ from (7) using the sample J_I. Denote the k-step estimators as $\hat{\beta}_I(\tau)$.

Further details are as follows. In step 1 we may use, for example, logit, probit, extreme value, linear (polynomial), or any other model that fits the data $\{\delta_i, X_i, C_i\}$ well. $\dot{X}_i$ denotes a suitable transform of X_i. For example, X_i may consist of X_i, C_i, and its squares (power series and regression spline approximations). In general, this gives an inconsistent estimator of the true propensity score

$$h(X_i, C_i) \equiv P(\delta_i = 1 | X_i, C_i),$$

but the inconsistency is not important as long as the misspecification is not too severe.

Indeed, the goal of step 1 is to select *some*, and not necessarily the largest, subset of observations where $h(X_i, C_i) > 1 - \tau$, that is, where the quantile line $X_i'\beta(\tau)$ is above C_i, so as to obtain a consistent but inefficient estimator $\hat{\beta}_0(\tau)$. For this task to be carried out, it suffices, but is not necessary, that, say, $p(\dot{X}_i'\gamma_0) - c$ is a lower bound on $h(X_i, C_i)$,

$$\text{a.s.} \quad p(\dot{X}_i'\gamma_0) - c < h(X_i, C_i), \quad \text{where } \gamma_0 \equiv \text{plim}\, \hat{\gamma}, \tag{8}$$

and it is *nontrivial*, meaning that the selected set J_0 is sufficiently rich; matrix $EX_iX_i'1\{i \in J_0\}$ is asymptotically invertible. Larger c and better model $p(\cdot)$ simplify the selection task. The envelope restriction may further be replaced by a much *weaker* condition—the separating hyperplane restriction in Theorem 1 (see Fig. 1 for motivation).

The foregoing construction assumes that the estimator $\hat{\gamma}$ is reasonable and converges to a value γ_0 that minimizes a sensible distance between $h(X_i)$ and the model $p(\dot{X}_i'\gamma)$. For example, $\hat{\gamma}$ may be defined by minimizing $\sum_{i=1}^n [\delta_i - p(\dot{X}_i'\gamma)]^2$, in which case, under standard conditions, $\hat{\gamma}_0$ converges to γ_0 that solves $\min_\gamma E[h(X_i) - p(\dot{X}_i'\gamma)]^2$. Alternatively, quasi–maximum likelihood methods may be used. In the empirical section we used a polynomial logistic model and estimated it by the conditional maximum likelihood estimator (MLE).

Another attractive choice is the Fisher–Rao discriminant analysis. The discriminant prospective is justifiable, as follows (treating C_i as a component of X). If $X_i | \{\delta_i = 1\}$ has density $g_1(x)$ and $X_i | \{\delta_i = 0\}$ has density $g_0(x)$, then, by Bayes's rule,

$$P(\delta_i = 1 | X_i = x) = \frac{q_1 g_1(x)}{q_1 g_1(x) + q_0 g_0(x)},$$

where $q_1 = P(\delta_i = 1) = 1 - q_0$. Approximating $g_1(x)$ and $g_0(x)$ by normality with different means and variances leads to the classical logistic linear-quadratic discriminant analysis (LQDA) (see, e.g., Amemiya 1985, p. 282). Other forms of g_1 and g_0 can also be used, but even the normality assumption has been known to produce good results (see Efron 1975; Press and Wilson 1978; Amemiya and Powell 1983). LQDA was also among the top 3 classifiers for 11 of 22 commercially important datasets in the Statlog project (Michie, Spiegelhalter, and Taylor 1994), outcompeting many sophisticated classifiers.

Formally, we do not confine ourselves to a particular estimator. Instead, an assumption such as (8) or its generalization will be required to hold.

Under conditions to be stated, $\hat{\beta}_1(\tau)$ is asymptotically normal with variance equal to that of the Powell estimator. Thus, starting with a good subset of observations, only two recomputations of QR suffice to obtain the Powell-efficient estimator. Estimators $\hat{\beta}_I(\tau)$ are also asymptotically normal with variance equal to that of the Powell estimator.

The QR iterations in step 4 are somewhat analogous to those in the pioneering and remarkable ILPA algorithm that

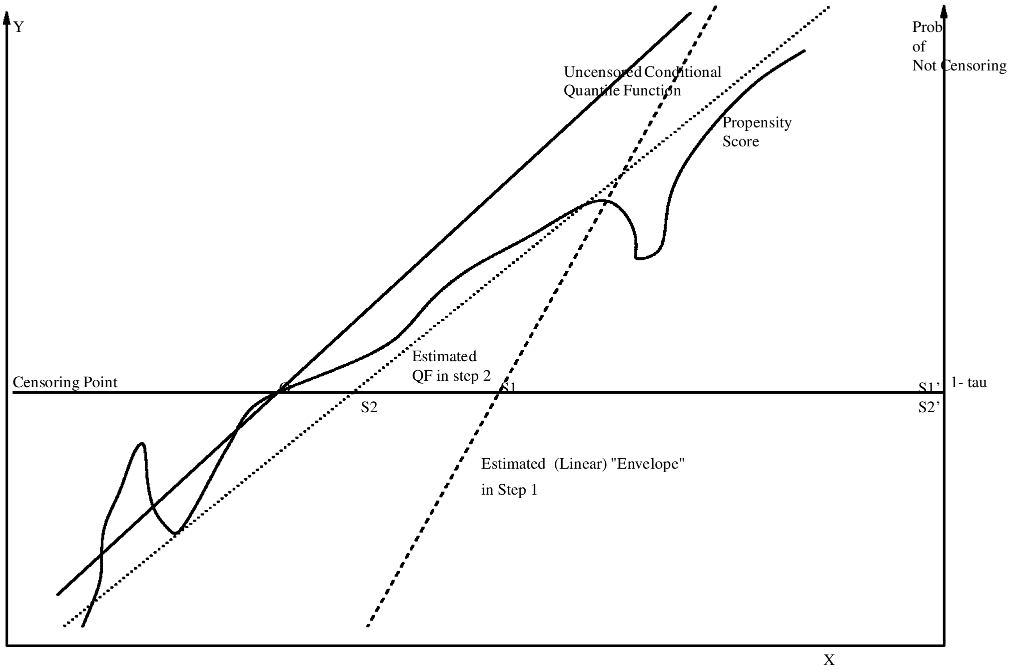

Figure 1. How It Works. The solid line depicts the conditional quantile function and the propensity score. The figure corresponds to (5) and uses notation defined there. The propensity score equals 1 − τ for the value of $\tilde{X}$ such that conditional quantile line $\tilde{X}'\tilde{\beta}(\tau) = 0$. The propensity score is above 1 − τ for $\tilde{X}$ such that $\tilde{X}'\tilde{\beta}(\tau) > 0$, and below 1 − τ if $\tilde{X}'\tilde{\beta}(\tau) < 0$. Once a sample is such that $\tilde{X}_i'\tilde{\beta}(\tau) > 0$, the conditional quantile function of the uncensored model can be estimated by the linear QR. The Initial step involves fitting an envelope of the propensity score, and selecting all i, such that $\tilde{X}_i \in [S1, S1']$. Note that the "envelope" is not correct, but this is irrelevant, because it acts as a good separating hyperplane selecting a subset of i such that $\tilde{X}_i'\tilde{\beta}(\tau) > 0$. The second step fits the QR line, which is used to select all $i : \tilde{X}_i \in [S2, S2']$. The third step uses the selected sample, which asymptotically gets close to the ideal, $i : \tilde{X}_i \in (O, S2']$.

Buchinsky (1994) designed for the Powell problem. (See Buchinsky 1994 and Fitzenberger 1997b for details.) The basic idea is to start at a value $\beta(\tau)$, say 0, and then proceed with iterative linear programming computations until convergence is reached. The convergence to the Powell estimator is not guaranteed and can be quite infrequent (see Fitzenberger 1997b and earlier discussions). The convergence to a local optimum does not lead to a consistent estimator (see Powell 1986). However, going beyond the third or fourth step is not desirable on both computational and statistical grounds, based on our Monte Carlo experience. Our results show that given the first classification step, only two recomputations of quantile regression lead to an efficient estimator (relative to the Powell estimator). In terms of computational aspects, the faster interior point algorithms of Portnoy and Koenker (1997) may be preferred to linear programming.

Finally, because of the distributional equivalence with the Powell CQR estimator, all of the inference procedures developed by Powell apply without modifications. The bootstrap inference of Bilias, Chen, and Ying (2000) can also be very useful in practice. All inference procedures are as for the standard quantile regression procedure, with the only difference being that the selected (rather than complete) sample is used. Therefore, a user of the standard QR software need not make any modification—the standard errors and confidence intervals produced by the last step QR routine are all valid. QR software for R and S-plus environments is available from Statlib or *http://www.econ.uiuc.edu/*. Other available software includes QR modules in STATA, Xplore, and TSP.

In summary, the estimation procedure has two very distinct features: a very simple, parametric classification first step and the additional third step. The estimator is as efficient as the Powell estimator.

2.2 The Model Beneath: How Restrictive is It?

The canonical CQR model in (4), together with the envelope restriction (8), can be thought of as a model. The envelope or the separation assumption in Theorem 1 is a critical ingredient to yield the simplicity. How restrictive is it?

Treat C_i as a component of X_i to simplify notation. The popular Amemiya–Tobin model assumes that conditional on X_i, Y_i is conditionally homoscedastic normal. Then propensity score $h(X)$ is $\Phi(X'\alpha)$ for the normal cdf Φ. A *significantly more general* CQR model can be immediately obtained by simply assuming that $\Phi(\dot{X}'\gamma_0) - c$ is a *nontrivial envelope* of an unknown propensity score $h(X)$, where, say, $\dot{X} = (X, X^2, \dots)$. Such an assumption imposes neither normality nor conditional homoscedasticity nor a location-scale submodel. Similarly, if the benchmark is the Weibull proportional hazard model from duration analysis, then $h(X) = \Gamma(X'\alpha)$ for the Gumbel cdf Γ. A much more flexible CQR model is obtained by assuming that $\Gamma(\dot{X}'\gamma_0) - c$ is a nontrivial envelope of the propensity score $h(X)$.

More generally, Theorem 1 replaces the intuitive envelope restriction by a weaker separation restriction, which requires that once $p(\dot{X}'\gamma_0)$ is above a threshold c, $h(X) > 1-\tau$. This assumption allows the envelope to be an *incorrect* lower bound of the propensity score, but requires only that it do a good job selecting a correct subset of observations. Figure 1 illustrates the situation. For well-behaved models, the further away from 0 the conditional quantile function, the further away from $1-\tau$ the propensity score function, and the easier it is to carry out the classification. Classification problems of this kind are a subject of the modern classification analysis, (e.g., Breiman, Friedman, Olshen, and Stone 1984; LeBlanc and Tibshirani 1996; Ripley 1996; Vapnik 2000). Therefore, in principle, many elaborate, structured strategies for the first classification step are available.

In summary, the model studied here is more restrictive than the canonical Powell's CQR model, yet it leaves the general, plausible features of the CQR intact. This model is also congenial from many other prospectives. The estimator is easily computable and applicable to such examples as extramarital data in section 3 (high censoring, very large sample, many categorical regressors) or Stanford heart data (small sample, high censoring, categorical regressors). It does well in Monte Carlo experiments and sensibly in real life examples. We believe that this estimator will help proliferate the presently scarce applications of the CQR models.

2.3 Large Sample Properties

The following assumptions are made in addition to (1)–(5).

Theorem 1. Suppose that for $u_i(\tau) \equiv Y_i^* - X_i'\beta(\tau)$ and all τ of interest, the following assumptions hold:

(a) $\{(X_i, Y_i^*, C_i)\}$ are iid; $u_i(\tau)$ has density $f_{u_i(\tau)}(u|X_i)$, which is bounded from above, away from 0, and continuous, uniformly in u near 0 and in X_i; $u_i(\tau)$ has τth conditional quantile at 0; the support of the distribution of (X_i, C_i), $\mathbf{X}$, is compact; and X_i includes a constant.

(b) $H_\eta(\tau) \equiv Ef_{u_i(\tau)}(0|X_i)X_iX_i'1[h(X_i, C_i) > (1-\tau)+\eta]$ is positive definite for a fixed constant $\eta \in (0, \tau)$.

(c) We know a pair of the model p and trimming constant c that form a nontrivial envelope or a separating hyperplane of the propensity score, $\exists\, c > 0, v > 0$,

$$p\big(\dot{X}_i'\gamma\big) > (1-\tau)+c \text{ implies } h(X_i, C_i) > (1-\tau)+v \text{ a.s.,}$$

for any γ in a neigborhood of $\gamma_0 \equiv \operatorname{plim}\hat{\gamma}$. $EX_iX_i'1\{p(\dot{X}_i'\gamma_0) > (1-\tau)+c\}$ is invertible, $p(\cdot)$ is strictly increasing and continuous, and $\dot{X}_i$ is a known function of X_i and C_i.

(d) $P(\widetilde{X}_i'\alpha > v)$ is Lipschitz in α uniformly in v, for α in an open neighborhood of γ_0 or $\beta(\tau)$ and $\widetilde{X}_i$ denoting $\dot{X}_i$ and X_i.

Under the stated assumptions, as $\delta_n \times \sqrt{n} \to \infty$ and $\delta_n \downarrow 0$,

$$\sqrt{n}(\hat{\beta}_I(\tau) - \beta(\tau)) \xrightarrow{d} N\big(0, H_0^{-1}(\tau)\Lambda_0(\tau)H_0^{-1}(\tau)\big)$$

for finite $I \geq 1$, where $\Lambda_0(\tau) \equiv \tau(1-\tau)E(X_iX_i'1\{h(X_i) > 1-\tau\})$. Furthermore, the same holds if any other consistent initial estimator $\hat{\beta}_0(\tau)$ is used in step 2, provided that sequence $\delta_n \downarrow 0$ and $|\hat{\beta}_0(\tau) - \beta(\tau)|/\delta_n \xrightarrow{p} 0$. Furthermore, the joint asymptotic distribution of several estimators for $\beta(\tau_j)$, $j \leq J$ is asymptotically normal, with covariance given by $H_0^{-1}(\tau_l)\Lambda_0(\tau_l, \tau_j)H_0^{-1}(\tau_j)$, where $\Lambda_0(\tau, \tau') \equiv [(\tau \wedge \tau') - \tau\tau']E(X_iX_i'1\{h(X_i, C_i) > (1-\tau) \vee (1-\tau')\}$.

Remark 1. Assumptions a and b are standard. Assumption c allows the parametric "probability" model $p(x'\gamma_0)$ to be misspecified (see Fig. 1). Assumption c rationalizes the parametric first step, as we have discussed. This assumption is a main restriction and should not be downplayed, as discussed in the preceding section.

Remark 2. The iid assumption on (X_i, δ_i, C_i) can be relaxed at a notational cost. All that is needed is that data are independent and the limit empirical distribution function of these variables exists. Under such conditions, laws of large numbers apply, and $\hat{\gamma}$ would still converge to a fixed limit γ_0.

Remark 3. No assumptions are made about the rate of convergence of $\hat{\gamma}$ to its probability limit γ_0 or of $\hat{\beta}_0(\tau)$ to $\beta(\tau)$. The trimming device is designed to eliminate the bias, and stochastic equicontinuity eliminates the impact of the variance of the preliminary steps. Assumption (d) requires, for example, the distribution of $\dot{X}_i'\alpha$ to respond smoothly to changes in α in the vicinity of γ_0. The Lipschitz condition can be replaced by the weaker Holder continuity.

Remark 4. Treating C_i as a subcomponent of X_i, it is useful to comment on what happens when the number of parameters in the separation model $\varphi(x) = p(\dot{x}(x)'\gamma) - c$ is increasing with n, for instance, if we model $\dot{x}_n(x)'\gamma_n$ as a power series or spline series. The result is preserved as long as $x \mapsto p(\dot{x}_n(x)'\hat{\gamma}_n)$ converges in Θ uniformly to a fixed function $x \mapsto p(x)$, where Θ is a subset of functions in $C_M^r(\mathbf{X})$, (see van der Vaart and Wellner 1996, p. 154) with $r > \dim(x)/2$, that also have the property that $P(\vartheta(X) > c)$ is uniformly Lipschitz in ϑ over Θ with respect to the $L_2(P)$ metric. Note that the resulting class of envelope models $\mathcal{P} \equiv \{x \mapsto 1(\varphi(x) > c),\ \varphi \in \Theta, c \in [0,1]\}$ is Donsker. This is true because the log of $L_2(P)$ bracketing number of $\mathcal{P}$ is of the same order as that for $\mathcal{F} = \{x \mapsto \varphi(x) - c, \varphi \in \Theta, c \in [0,1]\}$ by monotonicity of the indicator function and the Lipschitz property. The bracketing numbers for $\mathcal{F}$ are given in example 19.9 of van der Vaart (1998) or corollary 2.7.4. of van der Vaart and Wellner (1996). Hence if $r > \dim(x)/2$, then the bracketing entropy integral for $\mathcal{P}$ is finite, and the Donsker property holds in view of a constant envelope. Furthermore, the Donsker property is preserved when $\mathcal{P}$ is multiplied by a bounded random variable, as required in the proof.

2.4 Finite-Sample Properties

Table 1 reports the result of a small Monte Carlo experiment. The setup of our simulation is similar to that of Buchinsky and Hahn (1998). The model is a standard location-scale model with an error term hit by a linear-quadratic heteroscedastic scale, for $X_i = (1, \widetilde{X}_i)'$: $Y_i^* = X_i'\beta + \epsilon_i$. We draw $\widetilde{X}_i \in \mathbb{R}^5$ from independent standard normal distributions, truncated as $\{\widetilde{X}_i : \|X_i\|_\infty < 2\}$. The error term has the multiplicative heteroscedasticity structure, $\epsilon_i = u_i \times (1 + .5\sum_{j=1}^5(\widetilde{X}_{ij} + \widetilde{X}_{ij}^2))$, where $u_i \sim N(0, 25)$. The true parameter vector is chosen at $(1, 1, .5, -1, -.5, .25)$, and the censoring point is $-.75$,

Table 1. Monte Carlo Simulation Results With Five Regressors for the .50 Quantile (1,000 Repetitions) Normal Heteroscedastic Results

	Intercept					*Slope*				
	2-step	*3-step*	*5-step*	*BH**	*ILPA**	*2-step*	*3-step*	*5-step*	*BH**	*ILPA**
$n = 100$										
RMSE	5.08	2.15	2.79	3.25	3.93	2.69	1.89	2.35	2.25	3.05
Mean bias	4.16	1.46	1.59	1.38	.7	.2	−.08	.07	.21	.43
MAE	4.23	1.7	1.86	1.89	1.55	2.02	1.39	1.64	1.09	1.11
Median bias	3.71	1.43	1.33	1.54	1.22	.03	−.43	−.27	0	−.12
$n = 400$										
RMSE	2.28	1.05	1.21	1.28	1.31	1.13	.88	.93	.77	.89
Mean bias	1.94	.64	.75	.62	.74	−.18	−.28	−.18	−.24	−.38
MAE	1.99	.82	.88	.8	.81	.89	.69	.72	.51	.69
Median bias	1.83	.66	.71	.64	.78	−.19	−.43	−.3	−.3	−.54

NOTE: Sample sizes are denoted by $n(100, 400)$. The last two columns are from Buchinsky and Hahn (1998, table 2). BH denotes their estimator Cva, which uses a cross-validated bandwidth adjusted to the undersmoothing assumption. Powell* (ILPA*) is the Powell estimator or, more precisely, an estimator obtained by iterated linear programming.

which produces roughly 45% censoring. We use X and X^2 in the parametric propensity score regressions. We experimented with different probability models p, including logit, probit, and linear models. However, the type of probability model used has very little effect on the performance of the estimators. Therefore, in Table 1 we report the results only for the logit selection model. Note that due to the heteroscedasticity error structure, even the probit model is not consistent with the true propensity score. We report the initial step, the first step, and the third step estimators and compare them to the Buchinsky and Hahn (1998) estimator and to the Powell estimator.

Notably, the results from the Monte Carlo simulation show that for sample size 100, the 3-step estimator outperforms all other estimators in terms of root mean square errors (RMSEs). Its mean absolute deviation (MAE), its mean and median biases, are comparable to that of other estimators. Iterating to step 5 increases both the RMSE and the MAE, with the benefit of reducing both mean and median biases. Overall, the 5-step estimator still compares favorably to both the Buchinsky and Hahn (1998) estimator and the Powell estimator. The initial inefficient estimator does fairly poorly, as expected. In a large sample, $n = 400$, the 3-step estimator and the Buchinsky and Hahn estimator perform equally well and are both more favorable than other estimators in all dimensions. To conclude, the 3-step estimator does better in small samples and quite well in large samples.

3. DETERMINANTS OF EXTRAMARITAL AFFAIRS: A CENSORED QUANTILE REGRESSION ANALYSIS

Extramarital affairs, an important social phenomenon, have received much attention by anthropologists, psychologists, evolutionary biologists, sociologists, and economists (see e.g., Fair 1978; Reiss, Anderson, and Sponaugle 1980; Miller and Klein 1981; Cronk 1991; South and Lloyd 1995.) We present a retrospective analysis of the *Redbook* dataset on extramarital affairs, first analyzed by Fair (1978). We contrast our analysis mainly with his data and model-analytic findings. Fair (1978) presented a utility-based optimization model of the time spent in the affair (affair intensity) as determined by preference for diversity, value of goods consumed in and outside marriage, labor and nonlabor income, and time already spent with the spouse and paramour.

3.1 Data

The dataset was collected by *Redbook* magazine. Fair (1978) described the collection procedures as well as the place of this dataset among only few similar datasets. The dataset covers 6388 first-time married women, of which 68.5% reported to have had no extramarital affairs. This presents a very high degree of "censoring." We define the following variables similar to Fair (1978) to facilitate comparisons with Fair's statistical and model-analytic findings:

- Level of Affair, dependent variable, defined as the number of different partners outside marriage times the approximate number of relationships with each partner, divided by the number of years in the marriage. For 68.5% of the respondents, it is equal to 0. For the rest of the respondents, the density function is sketched by the histogram and a kernel estimator in Figure 2.

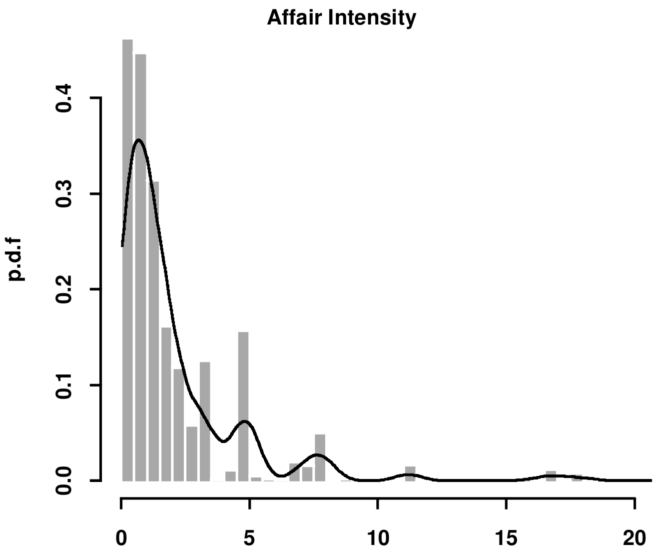

Figure 2.

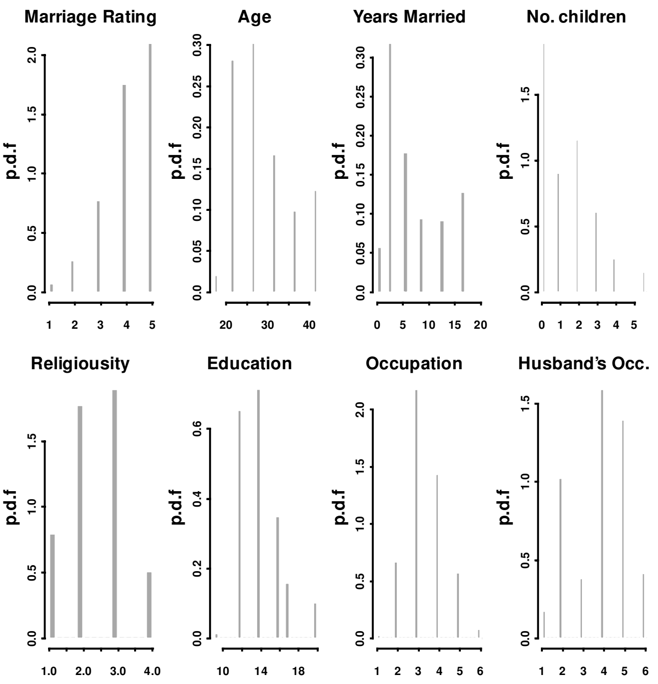

Figure 3.

Simple histograms of the following regressors are given in Figure 3:

- Marriage Rating: respondents' rating of their marriage, on a scale of 1 to 5.
- Age, Years Married, Number of Children.
- Religiousity: respondents' rating of their religiosity, on the scale from 1 to 4.
- Education: 9.0, 12.0, 14.0: grade school, high school, and some college; 16.0, 17.0, 20.0: college graduate, some graduate school, and advanced degree.
- Occupation, Husband's occupation: Hollingshead's socioeconomic status of occupation: 6, professional with advanced degree; 5, managerial, administrative, business; 4, teacher, artist, etc., 3, white collar (administrative, clerical); 2, blue colar (farming, factory); 1, student.

3.2 Models

The CQR model assumes the following form (with $C_i = 0$):

$$Q_{Y|X}(\tau) = (\alpha(\tau) + X'\theta(\tau)) \vee 0. \tag{9}$$

That is, the conditional quantile function of the affair level is either 0 or linear. This functional form is appealing, as we have discussed. We also consider a standard normal model,

$$Q_{Y|X}(\tau) = (\alpha + \sigma\Phi^{-1}(\tau) + X'\theta) \vee 0, \qquad \forall\, \tau, \tag{10}$$

where $\Phi^{-1}(\tau)$ is the inverse of the standard normal distribution. Another benchmark model is the accelerated failure time model [for $\exp(Y)$] from survival analysis,

$$Q_{Y|X}(\tau) = (\alpha + \sigma F^{-1}(\tau) + X'\theta) \vee 0, \qquad \forall\, \tau, \tag{11}$$

where F is an unspecified distribution function. It is easy to estimate the quantile shift effects θ in this model by taking or averaging any of the estimates of $\theta(\tau)$ in the CQR model. This will not be necessary, because neither this nor the normal model is supported by the data.

3.3 Estimation and Model Comparisons

To construct the initial envelope/classifier, we examined the pairwise plots of Y versus X. Many of covariates appeared to be associated with a higher dispersion of Y, which led us to consider a number of polynomial powers in the logistic model $p(\dot{x}'\hat{\gamma})$; $\dot{x}$ consisted of $x_{(i)}$, $x_{(i)}^2$, $x_{(i)}^3$, and certain interactions $x_{(i)}x_{(j)}$ that appeared to significantly improve the fit. The dimension of $\dot{x}$ was 18, which is plausible in view of the large sample size. Sensitivity of the final estimates to further increases in the complexity of the envelope was negligible. The trimming constant c was set to about .1 according to the rule described in Section 2, and $\hat{\gamma}$ was estimated by the conditional MLE.

Due to heavy censoring, it was not possible to estimate all quantile coefficients $\hat{\theta}(\tau)$. Identification depended on the nondegeneracy of the selected design matrix. This condition prevented considering quantiles lower than .4.

The results are summarized graphically in Figure 4. The solid line denotes the 3-step estimates of $\hat{\beta}(\tau) = (\hat{\alpha}(\tau), \hat{\theta}(\tau)')'$, $\tau \in \{.4, \ldots, .9\}$, and the shaded region depicts the pointwise 95% confidence intervals. The dashed line presents the MLEs $\hat{\theta}$ of the quantile shift effects in the normal model (10) obtained by Fair (1978).

$\hat{\theta}(\tau)$ varies significantly across quantiles, especially at higher ones. This presents an evident violation of homoscedasticity assumptions. Therefore, both model (10) and model (11) are strongly unsupportive of the data. It is still interesting to briefly comment on the behavior of MLEs of the normal model. It is well known that the estimates are not robust to violations of both normality (e.g., heavy tails) and homoscedasticity (see Hurd 1979; Goldberger 1983 for proofs and simulation studies). In our example, we have both. It is interesting that for five out of eight variables, the estimates $\hat{\theta}_j$ seem to correspond to the extreme quantile estimates $\hat{\theta}_j(\tau)$, $\tau \approx .9$. For the Marriage Longevity variable, in contrast, the estimate $\hat{\theta}_j$ is far away from any of $\hat{\theta}_j(\tau)$. Furthermore, in several cases, the sign of $\hat{\theta}_j(\tau)$ changes across τ, so $\hat{\theta}_j$ understandably cannot even match the direction of the quantile shift effect, because if the normal model (10) or location-shift model (11) were adequate, then it would have been the case that $\hat{\theta}(\tau) \approx \hat{\theta}$ for all τ. Thus $\hat{\theta}_j$ can hardly be given any meaning in the present setting. This finding is an empirical illustration of the earlier discussion on the breadth and flexibility of the CQR model.

3.4 Analysis

The quantile shift effects estimates $\hat{\theta}(\tau)$ are given in Figure 4. The Religiosity effect is expectedly negative at all quantiles of affair intensity and is especially strong at very high quantiles. The Education quantile shift effects are negative and strongly negative at high quantiles. Note that Education and Religiosity are weakly correlated (.14), but because we condition on Religiosity, the Education effects are net of this and other factors. Education effects are inexplicable within the Fair's model, yet they appear to have a clear meaning in

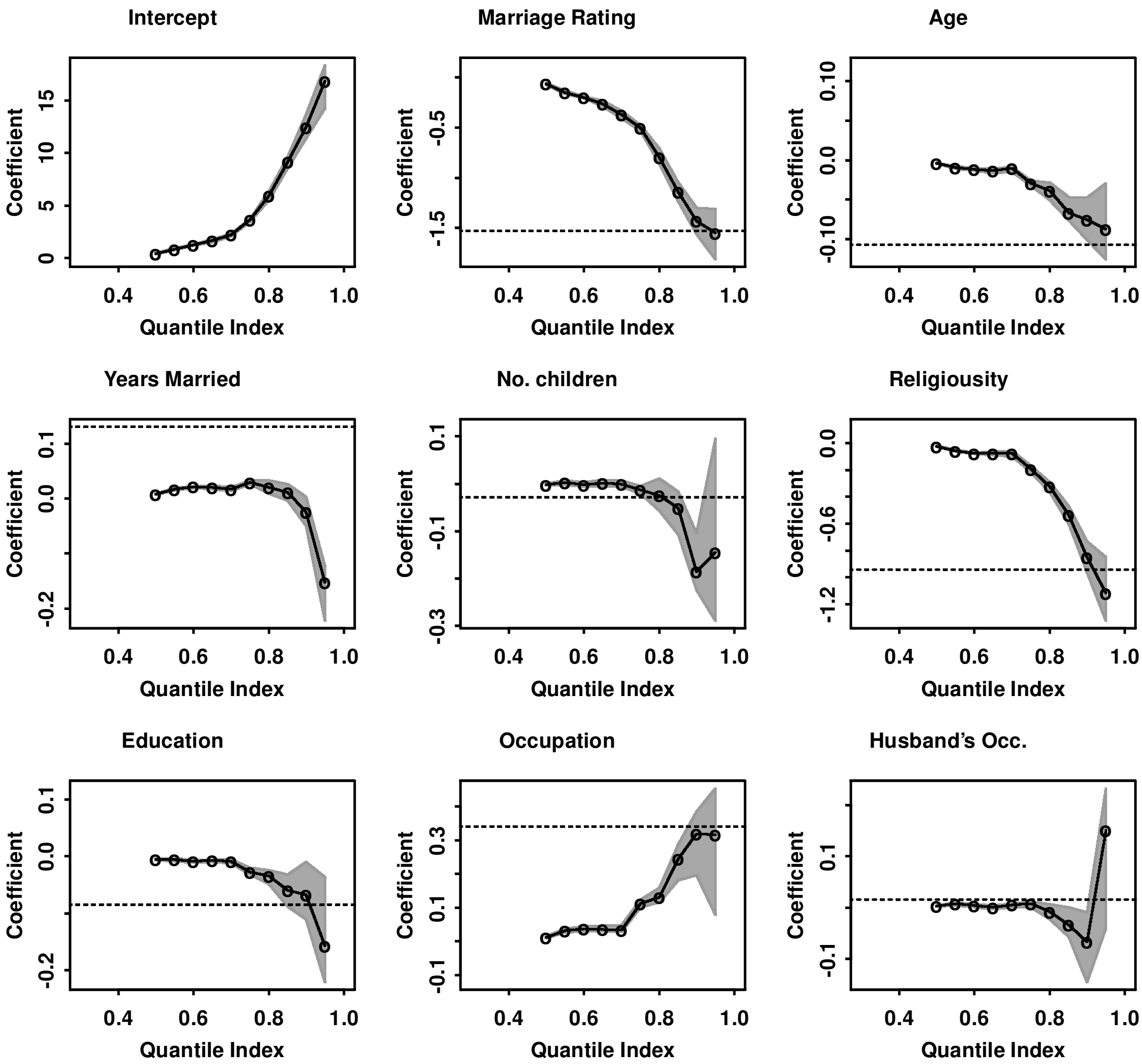

Figure 4.

view of the relational perspectives toward a paramour among the more intelligent and educated individuals (Reiss 1980).

The quantile shift effects for Age are nonpositive across all presented quantiles. Age effects are negative at the middle quantile and strongly negative at high quantiles. This means that the younger respondents are more likely to engage in an affair holding everything else fixed.

Women with an Occupation of higher socioeconomic status are relatively more likely to engage in affairs, especially more intense ones. Explanations for this are to be looked at. One view is that such status creates an interactional advantage, increasing the hazard of an affair and subsequent marital dissolution (South and Lloyd 1995). Fair's analytic model does not necessarily yield predictions about the direction of the status effect (because he treats the status as proxy for labor income). To the extent that higher status is associated with nonlabor income, or to the degree that income effects dominate substitution effects, Fair's model may predict a positive effect.

Husband's Occupational Status has a very small positive or a negligible effect across almost all quantiles, except at very extreme ones, where it becomes very negative. (It is positive but insignificant at .95.) Fair's analytic model predicts the positive effects of a husband's status (income) on the affair level, because a higher value of goods consumed in marriage causes wives to substitute labor activities for time spent with family and paramour. The Fair model, however, ignores the negative value of the dissolution option, which is real, as other studies have pointed out (South and Lloyd 1995). Our results suggest that Fair's model explains only the middle quantiles and does not apply to the high quantiles. It seems plausible that incorporation of the dissolution risk into the Fair model, more along the lines of Becker's (1968) crime and punishment model, can make it conform to the present findings.

The effect of Marriage Longevity is slightly positive at .6–.8 quantiles and strongly negative at high quantiles. Fair postulated that marriage longevity may positively relate to diversity considerations, leading to an increased affair level. However, it is not entirely clear why the effect is very negative at high quantiles. This may relate to the fact that only married and undivorced respondents were selected to the sample, so that marriage longevity correlates with the marriage match quality and thus has a deleterious effect on affairs. Such an outcome would be a clear prediction of the search (for spousal alternatives) theory.

Our finding thus partially challenges both the analytic and statistical predictions of Fair (1978) derived from the normal model.

4. DISCUSSION

This article had two goals. One goal was to offer an empirical CQR analysis of the determinants of a very serious social phenomenon—extramarital affairs. This is an important topic within sociology, psychology, and economics of marriage and family. To our regret, we found no previous implementable estimators that can be used in the settings of heavy censoring, many polychotomous or continuous regressors, and large or small samples. Such datasets seem to prevail in many areas of applied statistics. This justified the other goal, the pursuit of a practical, implementable, well-behaved estimator. The suggested estimator can be used to robustly estimate the CQR models, as well as many traditional models.

APPENDIX: PROOF OF THEOREM 1

Here, C, *const*, and K are generic positive constants, and C_i denotes the censoring point. The first part of the proof, up to (A.2), uses convexity arguments. The remainder of the proof shows that the estimated selector does not affect the asymptotic distribution, invoking the empirical process arguments.

Part 1. First, consider $\hat\beta_0(\tau)$. The rescaled statistic $Z_n^0 = \sqrt{n}(\hat\beta_0(\tau)-\beta(\tau))$ minimizes

$$Q_n(z,\hat\gamma) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_{in}(z)1[p(\dot X_i'\hat\gamma) > 1-\tau+c],$$

where $V_{in}(z) \equiv \sqrt{n}[\rho_\tau(\epsilon_i - X_i'z/\sqrt{n}) - \rho_\tau(\epsilon_i)]$ and $\epsilon_i \equiv Y_i - X_i'\beta(\tau)$. The claim is that for any finite collection of points z_j, $j \le l$,

$$(Q_n(z_j,\hat\gamma), j\le l) \xrightarrow{d} (Q_\infty(z_j), j \le l), \tag{A.1}$$

where $Q_\infty(z) \equiv \dot W'z + \frac{1}{2}z'\dot Jz$, $\dot W \stackrel{d}{=} N(0,\dot\Lambda)$, $\dot J \equiv Ef_u(0|X_i)\times X_iX_i'1[p(\dot X_i'\gamma_0) > 1-\tau+c]$, and $\dot\Lambda \equiv \tau(1-\tau)EX_iX_i'1[p(\dot X_i'\gamma_0) > 1-\tau+c]$. $\dot J$ is invertible by conditions b and d. Because Q_n and Q_∞ are convex, finite, and continuous in z, and because Q_∞ is uniquely minimized at $-\dot J^{-1}\dot W = O_p(1)$, (A.1) implies that

$$Z_n^0 \xrightarrow{d} -\dot J^{-1}\dot W$$

by the convexity theorem (e.g., Davis, Knight, and Liu 1992, Pollard 1991). But if $\hat\gamma = \gamma_0$, then (A.1) follows by the law of large numbers, the central limit theorem, and some standard calculations, and so it remains only to verify that

$$Q_n(z,\hat\gamma) - Q_n(z,\gamma_0) \xrightarrow{p} 0 \quad \text{for any fixed } z. \tag{A.2}$$

(a) For any fixed z, $\{Q_n(z,\gamma) - EQ_n(z,\gamma), \gamma\in\mathcal{G}\}$ is stochastically equicontinuous in γ, where $\mathcal{G} \equiv \{\gamma : |\gamma-\gamma_0| \le \delta\}$ and $\delta > 0$ is small. Indeed, $\mathcal{F} = \{\dot x \mapsto 1[p(\dot x'\gamma) > 1-\tau+c], \quad \gamma\in\mathcal{G}\}$ is a Vapnik-Červonenkis subgraph class, and hence by theorem 2 of Andrews (1994), it satisfies Pollard's entropy condition with a constant envelope. This property is *retained by the product* of $\mathcal{F}$ with random variable $V_{in}(z)$, $V_{in}(z)\otimes\mathcal{F}$, by theorem 3 of Andrews (1994), because by assumption a $|V_{in}(z)|$ has a constant envelope,

$$|V_{in}(z)| \le 2|X_iz| < \text{const.} \tag{A.3}$$

Hence, Part (a) is verified by theorem 1 of Andrews (1994).

(b) Space $\mathcal{G}$ with pseudometric $\rho(\gamma_1,\gamma_2) \equiv \sup_{n\ge1} E|V_{in}(z)\times[1\{p(\dot X_i'\gamma_1) > 1-\tau+c\} - 1\{p(\dot X_i'\gamma_2) > 1-\tau+c\}]|^2 \le$ const $\times\|\gamma_2 - \gamma_1\|_2$ is totally bounded, where the inequality easily follows from (A.3) and assumption d using compactness of $\mathbf{X}$. Parts (a) and (b) together imply that

$$\sup_{|\gamma-\gamma_0|\to0} |Q_n(z,\gamma) - Q_n(z,\gamma_0) - EQ_n(z,\gamma) + EQ_n(z,\gamma_0)| = o_p(1).$$

Thus to complete the proof of (A.2), it only remains to show that

$$|EQ_n(z,\gamma) - EQ_n(z,\gamma_0)|_{\gamma=\hat\gamma} = o_p(1). \tag{A.4}$$

We show that for $s_i(\gamma,\gamma_0) \equiv 1[p(\dot X_i'\gamma) > 1-\tau+c] - 1[p(\dot X_i'\gamma_0) > 1-\tau+c]$,

$$\begin{aligned} EQ_n(z,\gamma)|_{\gamma=\hat\gamma} - EQ_n(z,\gamma_0) &\equiv \sqrt{n}EV_{in}(z)s_i(\gamma,\gamma_0)|_{\gamma=\hat\gamma} \\ &= O_p(\hat\gamma-\gamma_0), \end{aligned} \tag{A.5}$$

Write $\sqrt{n}V_{in}(z) \equiv -\sqrt{n}[\{\tau - 1[\epsilon_i \le 0]\}X_i'z] + \sqrt{n}[-\eta_i(z)\{X_i'z - \epsilon_i\sqrt{n}\}] \equiv \sqrt{n}V_{in}'(z) + \sqrt{n}V_{in}''(z)$, where $\eta_i(z) \equiv [1(\epsilon_i \le 0) - 1(\epsilon_i \le X_i'z/\sqrt{n})]$. For γ close enough to γ_0, $p(\dot X_i'\gamma) > (1-\tau)+c$ implies $X_i'\beta(\tau) > C_i + v$ a.s. for $v > 0$ small for all i, so that

$$E[\sqrt{n}V_{in}'(z)s_i(\gamma,\gamma_0)|X_i, C_i] = 0 \text{ uniformly in } i, \tag{A.6}$$

because $P[\epsilon_i \le 0|X_i, C_i, X_i\beta(\tau) > C_i+v] = \tau$ [if $X_i\beta(\tau) > C_i$, ϵ_i has τth conditional quantile at 0]. Also, $E[\sqrt{n}V_{in}''(z)s_i(\gamma,\gamma_0)|X_i,C_i] = O[f_u(0|X_i)z'X_iX_i'z1(X_i'\beta(\tau) > C_i+v)]\times s_i(\gamma,\gamma_0)$, uniformly in i. Therefore, by assumptions c and d:

$$EE[\sqrt{n}V_{in}''(z)s_i(\gamma,\gamma_0)|X_i,C_i] = O(E[s_i(\gamma,\gamma_0)]) = O(\gamma-\gamma_0). \tag{A.7}$$

(A.6) and (A.7) imply (A.4).

Part 2. It suffices to show the result for $\hat\beta_1(\tau)$ with $\hat\beta_0(\tau)$ as the selector. The proof for $\hat\beta_I(\tau)$, $I > 1$ is identical. The proof of Part 2 is similar to that of Part 1, so only important differences are given. Rescaled statistic $Z_n^1 = \sqrt{n}(\hat\beta_1(\tau)-\beta(\tau))$ minimizes

$$Q_n(z,\hat\beta_0(\tau),\delta_n) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_{in}(z)1[X_i'\hat\beta_0(\tau) > C_i+\delta_n].$$

Consider δ_n as a parameter sequence. Proceed identically as in Part 1 up to (A.2), only replacing $1[p(\dot X_i'\hat\gamma) > 1-\tau+c]$ by $1[X_i'\hat\beta_0(\tau) > C_i+\delta_n]$, $1[p(\dot X_i'\gamma) > 1-\tau+c]$ by $1[X_i'\beta(\tau) > C_i]$ and $\dot W$, $\dot J$, and $\dot\Lambda$ by $W \stackrel{d}{=} N(0, H_0^{-1}\Lambda_0H_0^{-1})$, H_0, and Λ_0. It remains to show that

$$Q_n(z,\hat\beta_0(\tau),\delta_n) - Q_n(z,\beta(\tau),0) \xrightarrow{p} 0 \quad \text{for any fixed } z. \tag{A.8}$$

(a) For any fixed z, $\{Q_n(z,\beta,\delta) - EQ_n(z,\beta,\delta), (\beta,\delta)\in\mathcal{B}\times\mathcal{D}\}$ is stochastically equicontinuous in β,δ, where $\mathcal{B} \equiv \{\beta : |\beta-\beta(\tau)| \le C'\}$, $\mathcal{D} \equiv \{\delta : 0 \le \delta \le C''\}$, and $C', C'' > 0$ are small. Indeed, $\mathcal{F} = \{(X,C)\mapsto 1[X'\beta > C+\delta], \quad (\beta,\delta)\in\mathcal{B}\times\mathcal{D}\}$ is a VC subgraph class. Hence $\mathcal{F}$ has a finite uniform covering entropy integral and a constant envelope; that is, it satisfies Pollard's entropy condition. This property is *retained by the product of space* $\mathcal{F}$ with random variable $V_{in}(z)$: $V_{in}(z)\otimes\mathcal{F}$, by theorem 3 of Andrews (1994), because $|V_{in}(z)|$ has a constant envelope. Thus part (a) is verified by theorem 1 of Andrews (1994).

(b) Space $\mathcal{B} \times \mathcal{D}$ is totally bounded under the L_2 pseudometric,

$$\begin{aligned}\rho((\beta_1, \delta_1), (\beta_2, \delta_2)) &\equiv \sup_n (E|V_{in}(z) \times [1\{X_i'\beta_1 > \delta_1\} - 1\{X_i'\beta_2 > \delta_2\}]|^2) \\ &\leq \text{const} \times \|\beta_2 - \beta_1\|_2 + \text{const} \times \|\delta_2 - \delta_1\|_2, \qquad \text{(A.9)}\end{aligned}$$

where (A.9) follows from (A.3) and from assumption d, treating δ as shifting the intercept parameter. Part (b), along with Part (a), imply that

$$\sup_{|\beta-\beta(\tau)|\to 0} |Q_n(z, \beta, \delta_n) - Q_n(z, \beta(\tau), 0) - EQ_n(z, \beta, \delta_n) + EQ_n(z, \beta(\tau), 0)| = o_p(1).$$

Thus, to complete the proof of (A.8), it remains to show that

$$|EQ_n(z, \beta, \delta_n) - EQ_n(z, \beta(\tau), 0)|_{\beta=\hat{\beta}(\tau)} = o_p(1). \qquad \text{(A.10)}$$

Let $s_i(\beta, \beta(\tau)) \equiv 1(X_i'\beta > C_i + \delta_n) - 1(X_i'\beta(\tau) > C_i + 0)$. By assumption on the sequence δ_n, with probability $\to 1$, $\hat{\beta}_0(\tau)$ is inside the ball with radius $\kappa'\delta_n$, centered at $\beta(\tau)$, where $\kappa' > 0$ is small. By the compactness assumption on X_i, κ' can be chosen so that with probability $\to 1$, $\sup_{x\in\mathbf{X}} |x'(\hat{\beta}_0(\tau) - \beta(\tau))| < \frac{1}{2}\delta_n$. So set β inside this ball. Then for small enough κ' chosen as such, $x'\beta + c > \delta_n$ implies $x'\beta(\tau) > c$, and $x'\beta(\tau) \leq c$ implies $x'\beta \leq c + \delta_n$. Thus $s_i(\beta, \beta(\tau)) \neq 0$ necessarily implies $X_i'\beta(\tau) > C_i$ a.s. Hence, uniformly in i,

$$\begin{aligned}&E[\sqrt{n}V_{in}'(z)s_i(\beta, \beta(\tau))|X_i, C_i] \\ &\quad = E[\sqrt{n}V_{in}'(z)1(X_i'\beta(\tau) > C_i)|X_i, C_i] \times s_i(\beta, \beta(\tau)) = 0, \qquad \text{(A.11)}\end{aligned}$$

because $P[\epsilon_i \leq 0|X_i, C_i, X_i\beta(\tau) > C_i] = \tau$. Also,

$$\begin{aligned}&E[\sqrt{n}V_{in}''(z)s_i(\beta, \beta(\tau))|X_i, C_i] \\ &\quad = O[f_u(0|X_i)z'X_iX_i'z1(X_i'\beta(\tau) > C_i)] \times s_i(\beta, \beta(\tau)), \qquad \text{(A.12)}\end{aligned}$$

uniformly in i. Therefore, by assumption c and Lipschitz condition d,

$$\begin{aligned}EE[\sqrt{n}V_{in}''(z)s_i(\beta, \beta(\tau))|X_i, C_i] &= O[Es_i(\beta, \beta(\tau))] \\ &= O(\beta - \beta(\tau)) + O(\delta_n). \qquad \text{(A.13)}\end{aligned}$$

(A.11) and (A.12) give (A.10).

Finally, joint convergence results can be obtained analogously.

[Received June 2001. Revised December 2001.]

REFERENCES

Abadie, A., Angrist, J., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Amemiya, T. (1981), "Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

——— (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

Amemiya, T., and Powell, J. L. (1983), "A Comparison of the Logit Model and Normal Discriminant Analysis When the Independent Variables are Binary," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, New York: Academic Press, pp. 3–30.

Andrews, D. (1994), "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. 4*, eds. R. Engle and D. McFadden, Amesterdam: North-Holland.

Becker, G. S. (1968), "Crime and Punishment: An Economic Approach," *The Journal of Political Economy*, 76, 169–217.

Bilias, Y., Chen, S., and Ying, Z. (2000), "Simple Resampling Methods for Censored Regression Quantiles," *Journal of Econometrics*, 99, 373–386.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), *Classification and Regression Trees*, Belmont, CA: Wadsworth.

Buchinsky, M. (1994), "Changes in U.S. Wage Structure 1963–87: An Application of Quantile Regression," *Econometrica*, 62, 405–458.

Buchinsky, M., and Hahn, J. (1998), "An Alternative Estimator for the Censored Regression Model," *Econometrica*, 66, 653–671.

Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *The Annals of Statistics*, 19, 760–777.

Chaudhuri, P., Doksum, K., and Samarov, A. (1997), "On Average Derivative Quantile Regression," *The Annals of Statistics*, 25, 715–744.

Cronk, L. (1991), "Human Behavioral Ecology," *Annual Review of Anthropology*, 20, 25–53.

Davis, R. A., Knight, K., and Liu, J. (1992), "*M*-Estimation for Autoregressions With Infinite Variance," *Stochastic Processes and Their Applications*, 40, 145–180.

Doksum, K. (1974), "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-sample case," *Annals of Statistics*, 2, 267–277.

Efron, B. (1975), "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis," *Journal of the American Statistical Association*, 70, 892–898.

Fair, R. C. (1978), "Theory of Extramarital Affairs," *Journal of Political Economy*, 86, 45–61.

Fitzenberger, B. (1997a), "Computational Aspects of Censored Quantile Regression," in *Proceedings of the 3rd International Conference on Statistical Data Analysis Based on the L1-Norm and Related Methods*, ed. Y. Dodge, Hayward, CA: IMS, pp. 171–186.

——— (1997b), "A Guide to Censored Quantile Regressions," in *Handbook of Statistics, Vol. 14, Robust Inference*, Amsterdam: North-Holland, pp. 405–437.

Fitzenberger, B., and Winker, P. (2001), "Improving the Computation of Censored Quantile Regression," preprint.

Goldberger, A. (1983), "Abnormal Selection Bias," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. T. A. S. Karlin and L. Goodman, New York: Academic Press.

Hogg, R. V. (1975), "Estimates of Percentile Regression Lines Using Salary Data," *Journal of the American Statistical Association*, 70, 56–59.

Honoré, B., Khan, S., and Powell, J. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67–105.

Horowitz, J., and Neumann, G. (1987), "Semiparametric Estimation of Employment Duration Models," *Econometric Reviews*, 6, 5–40.

Hurd, M. (1979), "Estimation in Truncated Samples When There is Heteroscedasticity," *Journal of Econometrics*, 11, 247–258.

Jureckova, J., and Prochazka, B. (1994), "Regression Quantiles and Trimmed Least Squares in Nonlinear Regression Model," *Journal of Nonparametric Statistics*, 3, 201–222.

Khan, S., and Powell, J. L. (2001), "Two-Step Estimation of Semiparametric Censored Regression Models," *Journal of Econometrics*, 100, 319–355.

Knight, K. (1998), "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *The Annals of Statistics*, 26, 755–770.

Koenker, R., and Bassett, G. S. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and Geling, O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458–468.

Koenker, R., and Machado, J. A. F. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1310.

Koenker, R., and Portnoy, S. (1987), "*L*-Estimation for Linear Models," *Journal of the American Statistical Association*, 82, 851–857.

LeBlanc, M., and Tibshirani, R. (1996), "Combining Estimates in Regression and Classification," *Journal of the American Statistical Association*, 91, 1641–1650.

Lehmann, E. L. (1974), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

Michie, Spiegelhalter, and Taylor (1994), *Machine Learning, Neural, and Statistical Classifcation*, Ellis Horwood, Gregory Piatetsky-Shapiro.

Miller, B. C., and Klein, D. M. (1981), "A Survey of Recent Marriage and Family Texts (in Survey Essays)," *Contemporary Sociology*, 10, 8–21.

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

Portnoy, S. (1991), "Asymptotic Behavior of Regression Quantiles in Nonstationary, Dependent Cases," *Journal of Multivariate Analysis*, 38, 100–113.

——— (2001), "Censored Regression Quantiles," preprint.

Portnoy, S., and Jureckova, J. (2000), "On Extreme Regression Quantiles," *Extremes*, 227–243.

Portnoy, S., and Koenker, R. (1997), "The Gaussian Hare and the Laplacian Tortoise," *Statistical Science*, 12, 279–300.

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

——— (1991), "Estimation of Monotonic Regression Models Under Quantile Restrictions," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics (Durham, NC, 1988)*, Cambridge, U.K.: Cambridge University Press, pp. 357–384.

Press, S. J., and Wilson, S. (1978), "Choosing Between Logistic Regression and Discriminant Analysis," *Journal of the American Statistical Association*, 73, 699–705.

Reiss, I. (1980), *Family Systems in America*, New York: Holt, Reinhart and Winston.

Reiss, I., Anderson, R., and Sponaugle, G. (1980), "A Multivariate Model of the Determinants of Extramarital Permissiveness," *Journal of Marriage and Family*, 42, 395–410.

Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge, U.K.: Cambridge University Press.

South, S. J., and Lloyd, K. M. (1995), "Spousal Alternatives and Marital Dissolution," *American Sociological Review*, 60, 21–35.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge, U.K.: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Vapnik, V. N. (2000), *The Nature of Statistical Learning Theory*, (2nd ed.), New York: Springer-Verlag.

Yang, S. (1997), "Extended Weighted Log-Rank Estimating Functions in Censored Regression," *Journal of the American Statistical Association*, 92, 977–984.

Zhou, M. (1992), "M-Estimation in Censored Linear Models," *Biometrika*, 79, 837–841.