Minimum Distance Estimation of Randomly Censored Regression Models

Shakeeb Khan*
University of Rochester
Elie Tamer
Northwestern University

March 2005
Preliminary and Incomplete

Abstract

This paper proposes minimum distance estimation procedures for the slope coefficients and location parameter in a censored regression model. The distribution of the error term is assumed to satisfy a conditional median restriction but generalizes existing work in terms of the restrictions imposed on the distribution of the error term and the censoring variable. The estimator is shown to converge at the root-n rate with an asymptotic normal distribution. A small scale simulation study demonstrates adequate finite sample performance.

JEL Classification: C14, C25, C13.

Key Words: Semiparametric, accelerate failure time model, covariate dependent censoring, minimum distance, heteroskedasticity.

*Corresponding author. S. Khan, Department of Economics, University of Rochester, Rochester, NY, 14627, USA. (585)-275-5781, skhan@troi.cc.rochester.edu. Both authors gratefully acknowledge support from the NSF. Any errors are our own.

1 Introduction

Much of the recent econometrics, statistics, and biostatistics literature has been concerned with distribution-free estimation of the parameter vector β_0 in the linear regression model

$$y = x'\beta_0 + \epsilon$$

where the dependent variable y is subject to censoring that can potentially be random. For example, in the duration literature, this model is known as the *accelerated failure time* (or AFT) model where y, typically the logarithm of survival time, is right censored at varying censoring points.

The semiparametric literature which studies variations of this model is quite extensive and can be classified by the set of assumptions that a given paper imposes on the joint distribution of (x, ϵ, c) where c is the censoring variable. Work in this area includes the papers by Buckley and James (1979), Powell(1981), Koul et al.(1981), Ying et al.(1995), Yang(1999), Honoré et al.(2002) and, more recently Portnoy(2003) among many others. Unfortunately, each of the estimation methods introduced in the literature impose some assumption which may be considered too strong and not reasonably characterized by the data, and furthermore, the proposed methods will yield inconsistent estimators of the parameters of interest if these assumptions do not hold. Examples of assumptions which may be regarded as too strong are homoskedastic errors, censoring variables that are independent of the regressors, strong support conditions on the censoring variable which rule out fixed censoring.

This paper adds to this important literature in different dimensions. We propose an estimation procedure for the censored regression model which does not require any of these strong conditions. Specifically, it permits conditional heteroskedasticity in the data, permits the censoring variable to depend on the covariates in an arbitrary way, can be applied to both fixed and randomly censored data. Our proposed minimum distance estimator does not require nonparametric estimation of the censoring variable or error distribution, and consequently does not require the selection of smoothing parameters nor trimming procedures.

The following section describes the censored model studied in this paper in detail, and introduces the proposed minimum distance estimation procedure. It also compare this procedure to others recently proposed in the literature. Section 2 establishes the asymptotic properties for the proposed procedure, specifying sufficient regularity conditions for root-n consistency and asymptotic normality. Section 3 explores the relative finite sample performance of the estimator using a simulation study, and section 4 concludes by summarizing results and discussing areas for future research. A mathematical appendix provides the details of the proofs of the asymptotic theory results.

2 Model and Minimum Distance Estimation Method

This paper will estimate the parameters in an accelerate failure time model, which is characterized by the observed random variables v_i, d_i, x_i, where x_i is a k-dimensional vector of covariates, v_i is scalar variable, and d_i is a binary variable the indicates whether an observation is censored or not. The censored AFT studied here can be expressed by the following two equations:

$$v_i = min(y_i, c_i) = min(x_i'\beta_0 + \epsilon_i, c_i) \tag{2.1}$$

$$d_i = I[x_i'\beta_0 + \epsilon_i < c_i] \tag{2.2}$$

where β_0 denotes the unknown k-dimensional parameter vector of interest, c_i denotes the censoring variable that is only observed for censored observations, and ϵ_i denotes the unobserved error term. In the absence of censoring, $x_i'\beta_0 + \epsilon_i$ would be equal to the observed dependent variable, which in the AFT context usually is the log of the survival time. In the censored model, the log-survival time is only partially observed.

The parameter of interest is β_0. We provide an identification result that will be based primarily on the following conditions, which characterize the censoring and error term behavior:

A1 $\text{med}(\epsilon_i|x_i) = 0$

A2 $c_i \perp \epsilon_i|x_i$

A3 The matrix $E[I[x_i'\beta_0 \leq c_i]x_i x_i']$ is invertible.

The first of the above conditions is the conditional median assumption imposed previously in the literature- e.g. Honoré et al.(2002) and Ying et al.(1995). It permits very general forms of heteroskedasticity, and is weaker than the assumption $\epsilon_i \perp x_i$ as was imposed in Buckley and James(1979), Yang(1999), Portnoy(2003). The second condition allows the censoring variable to depend on the regressors in an arbitrary way, and is weaker than the condition $c_i \perp x_i$ imposed in Honoré et al.(2002), Ying et al.(1995)[1]. Thus we can see that by permitting both conditional heteroskedasticity and covariate dependent censoring, our assumptions are weaker than existing work on the censored AFT model. Furthermore the third condition imposes weaker support conditions on the censoring variable when compared to existing work in the literature. For example, the estimator in Koul et al.(1981) requires the support of the censoring variable to be sufficiently large, thereby ruling out fixed censoring. The estimator in Ying et al.(1995) requires the censoring variable to exceed the index $x_i'\beta_0$ with probability 1, thus often ruling out the fixed censoring case as well.

We now describe our procedure for identifying β_0. Before doing so, we introduce functions of the data we will be using to establish an identification result that estimation will be based on. For any possible parameter value β, define the functions of x_i

$$\tau_1(x_i, \beta) = E[I[v_i \geq x_i'\beta]|x_i] - \frac{1}{2} \tag{2.3}$$

and let

$$\begin{aligned} \tau_0(x_i, \beta) &= E[(1 - d_i) + d_i I[v_i \geq x_i'\beta]|x_i] - \frac{1}{2} &(2.4)\\ &= \frac{1}{2} - E[d_i I[v_i \leq x_i'\beta]|x_i] &(2.5) \end{aligned}$$

We next provide an objective function under which our estimator will be based on.

Lemma 2.1 *Let the function $g(x_i, \beta)$ be defined as:*

$$g(x_i, \beta) = \tau_1(x_i, \beta)I[\tau_1(x_i, \beta) \geq 0] + \tau_0(x_i, \beta)I[\tau_0(x_i, \beta) \leq 0] \tag{2.6}$$

then under Assumptions A1-A3,

1. $g(x_i, \beta_0) = 0$ *x_i-a.s.*

2. $P_X(g(x_i, \beta) > 0) > 0$ *for all* $\beta \neq \beta_0$.

[1]Both of these papers suggest methods to allow for the censoring variable to depend on the covariates, which involve replacing the Kaplan-Meier procedure they require with a conditional Kaplan Meier. This will require the choice of smoothing parameters to localize the Kaplan-Meier procedure, as well as the additional regularity conditions in Beran(1981).

The lemma is stating that at the true value of the parameter, the function $g(\cdot, \beta_0)$ is 0 for all values in the support of x_i, whereas at $\beta \neq \beta_0$ the function is positive *somewhere* in the support of x_i. As this result is fundamental to what follows in the paper, we prove it here.

Proof: First we note note

$$P(v_i \geq x_i'\beta_0|x_i) = P(\epsilon_i \geq 0|x_i)P(c_i \geq x_i'\beta_0|x_i) = \frac{1}{2}P(c_i \geq x_i'\beta_0|x_i)$$

This results in $\tau_1(x_i, \beta_0)I[\tau_1(x_i, \beta_0) \geq 0] = 0$ Similarly, since

$$d_i v_i = d_i(x_i'\beta_0 + \epsilon_i) \leq x_i'\beta_0 + \epsilon_i$$

one can show that $\tau_0(x_i, \beta_0)$ is always non-negative, and thus $\tau_0(x_i, \beta_0)I[\tau_0(x_i, \beta_0) \leq 0] = 0$.

Next, consider the "imposter" value $\beta \neq \beta_0$. Let $\delta = \beta - \beta_0$. By Assumption A3, we can find a subset of the support of (c_i, x_i'), which we denote by S_{cx}^*, that has positive measure, where $x_i'\beta_0 \leq c_i$ for all $(c_i, x_i) \in S_{cx}$ and x_i does not lie in a linear subspace of $\mathbf{R}^k$ when restricted to this set. Consequently, for any $(c^*, x^*) \in S_{cx}$, we have $x^*\delta \neq 0$.

We will establish that $g(x^*, \beta) \neq 0$. If $x^*\delta < 0$ then since $c^* \geq x^*\beta_0$,

$\tau_1(x^*, \beta) = P(\epsilon_i \geq x^*\delta|x_i = x^*) - \frac{1}{2} > 0$. by A2. If $x^*\delta > 0$, first note since for any x_i,

$d_i I[v_i \leq x_i'\beta] = I[x_i'\beta_0 + \epsilon_i \leq c_i]I[x_i'\beta_0 + \epsilon_i \leq x_i'\beta] = I[\epsilon_i \leq min(c_i - x_i'\beta_0, x_i'\delta)]$

So at c^*, x^*, where $x^*\beta_0 < c^*$, $x^*\delta > 0$, the expected value of the above expression is greater than $\frac{1}{2}$. Consequently $\tau_0(x^*, \beta) < 0$. Therefore, we have established $g(x^*, \beta) \neq 0$ for any $\beta \neq \beta_0$ and any $(c^*, x^*) \in S_{cx}$, which has positive measure. ■

Remark 2.1 *Our identification result in Lemma 2.1 uses available information in the two functions $\tau_1(\cdot, \cdot)$ and $\tau_0(\cdot, \cdot)$. We can contrast this with the procedure in Ying et al.(1995) which is only based on the function $\tau_1(\cdot)$, and consequently requires to reweight the data using the Kaplan Meier(1958) estimator. As alluded to previously, this imposes strong support conditions on the censoring variable does not allow for covariate dependent censoring, unless one uses the conditional Kaplan Meier estimator in Beran(1981) to reweight the data, which introduces the complication of selecting smoothing parameters and trimming procedures.*

Lemma 2.1 establishes a conditional moment condition which we aim to base our estimator for β_0 on. Recent work on conditional moment estimation includes Donald et al.(2003), and recently in Dominguez and Lobato(2004)[2] (DL hereafter). As mentioned in DL, some previous work in the literature (e.g. Chamberlain(1987)), do attain the semiparametric efficiency bound, but at the expense of selecting smoothing parameters.

Our identification result and estimation procedure will be similar to the framework considered in DL, who recognized the difficulty in translating a conditional moment model into an unconditional moment model while ensuring global identification of the parameters of interest. Our model is based on a set of conditional moment inequalities that are satisfied uniquely for all x at the truth. To ensure global identification from a set of conditional moment inequalities, we provide a procedure that modifies the estimator in DL to take account of moment inequalities and preserve global point identification[3].

To explain how we attain global point identification, we define the following functions of the parameter vector, and two vectors of the same dimension as x_i. Specifically let t_1, t_2 denote two vectors the same dimension as x_i and define the following functions:

$$H_1(\beta, t_1, t_2) = E\left\{[I[v_i \geq x_i'\beta] - \frac{1}{2}]I[t_1 \leq x_i \leq t_2]\right\} \tag{2.7}$$

$$H_0(\beta, t_1, t_2) = E\left\{[\frac{1}{2} - d_i I[v_i \leq x_i'\beta]]I[t_1 \leq x_i \leq t_2]\right\} \tag{2.8}$$

where above the inequality $t_1 \leq x_i$, corresponds to each component of the two vectors. From Billingsley(1995, Theorem 16.10) a conditional moment condition can be related to unconditional moment conditions with indicator functions as above comparing regressor values to all vectors in $\mathbf{R}^k$. Our identification result will need to make use of two distinct vectors in $\mathbf{R}^k$. As we will see below, this will translate into estimation procedure involving a third order U-process.

Our global identification result is based on the following objective function of distinct realizations of the observed regressors, denoted here by x_j, x_k:

$$Q(\beta) = E\left[H_1(\beta, x_j, x_k)^2 I[H_1(\beta, x_j, x_k) \geq 0] + H_0(\beta, x_j, x_k)^2 I[H_0(\beta, x_j, x_k) \leq 0]\right] \tag{2.9}$$

[2]Similar results are also given in Koul(2002) and Stute(1996).

[3]In interesting work, Rosen(2005) has extended the DL framework to estimate models that are set identified.

The intuition of our identification result is the following simple case. Suppose that uniquely at $\theta = \theta_0$, the following inequality moment condition is satisfied $E[m(y;\theta)|x] \leq 0$ for all x a.e. , then also uniquely at $\theta = \theta_0$, we have $H(x_i, x_j; \theta) = E[I[x_i \leq x \leq x_j]m(y;\theta)] = 0$ for all x_i and x_j a.e. The main identification result is stated in the following lemma:

Lemma 2.2 *Under Assumptions A1-A3, $Q(\beta)$ is uniquely minimized at $\beta = \beta_0$.*

Proof: We first show that

$$Q(\beta_0) = 0 \tag{2.10}$$

To see why note this follows directly from the previous lemma which established that $\tau_1(x_i, \beta_0) = \tau_0(x_i, \beta_0) = 0$ for all values of x_i on its support. Similarly, as established in that lemma, there exists a regressor value x^*, such that for all x in a sufficiently small neighborhood of x^*, $\max(\tau_1(x,\beta)I[\tau_1(x,\beta) \geq 0], -\tau_0(x,\beta)I[\tau_0(x,\beta) \leq 0]) > 0$ for any $\beta \neq \beta_0$. Let $\mathcal{X}_\delta$ denote this neighborhood of x^*. Since x_i has the same support across observations, if we let $\mathcal{X}_{jk}$ denote the set of values that $x_k - x_j$ takes, it follows that the set $\mathcal{X}_{jk} \cap \mathcal{X}_\delta$ has positive measure, establishing that $Q(\beta) > 0$ for $\beta \neq \beta_0$. ■

Having shown global identification, we propose an estimation procedure, which is based on the analogy principle, and minimizes the sample analog of $Q(\beta)$. Our estimator involves a third order U-statistic which selects the values of t_1, t_2 that ensures conditioning on all possible regressor values, and hence global identification. Specifically, we propose the following estimation procedure:

First, define the functions:

$$\hat{H}_1(\beta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})I[x_j \leq x_i \leq x_k] \tag{2.11}$$

$$\hat{H}_0(\beta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2} - d_i I[v_i \leq x_i'\beta])I[x_j \leq x_i \leq x_k] \tag{2.12}$$

$$
\begin{aligned}
\hat{\beta} &= \arg\min_{\beta\in\mathcal{B}} \hat{Q}_n(\beta) &(2.13)\\
&= \arg\min_{\beta\in\mathcal{B}} \frac{1}{n(n-1)} \sum_{j\neq k} \Big\{ \hat{H}_1(\beta, x_j, x_k)^2 I[\hat{H}_1(\beta, x_j, x_k) \geq 0] &(2.14)\\
&+ \hat{H}_0(\beta, x_j, x_k)^2 I[\hat{H}_0(\beta, x_j, x_k) \leq 0] \Big\}
\end{aligned}
$$

As we can see, the above estimation procedure minimizes a third order U-process, which is computationally expensive. As is the case with DL, (which minimizes a second order U-process) this provides us with estimation procedure without the need to select smoothing parameters.

We next turn attention to the asymptotic properties of the estimator. We begin by establishing consistency under the following assumptions.

C1 The parameter space $\mathcal{B}$ is a compact subset of $\mathbf{R}^k$.

C2 The sample vector $(d_i, v_i, x_i')'$ is i.i.d.

C3 $Q(\beta)$ is continuous.

The following theorem establishes consistency of the estimator; its proof is left to the appendix.

Theorem 2.1 *Under Assumptions A1-A3, and C1-C3,*

$$\hat{\beta} \xrightarrow{p} \beta_0$$

We next turn attention to root-n consistency and asymptotic normality. Our results our based on the following additional regularity conditions:

D1 β_0 is an interior point of the parameter space $\mathcal{B}$.

D2 The error terms $\{\epsilon_i\}$ are absolutely continuously distributed with conditional density function $f(\epsilon \mid x)$ given the regressors $x_i = x$ which has median equal to zero, is bounded above, Lipschitz continuous in ϵ, and is bounded away from zero in a neighborhood of zero, uniformly in x_i.

D3 The censoring values $\{c_i\}$ are distributed independently of ϵ_i conditionally on the regressors. We note this weak restriction permits both conditional heteroskedasticity and covariate dependent censoring.

D4 The regressors $\{x_i\}$ and censoring values $\{c_i\}$ satisfy

(i) $\Pr\{| c_i - x_i'\beta |\leq d\} = O(d) \quad \text{if} \quad \|\beta - \beta_0\| < \eta_0, \quad \text{some} \quad \eta_0 > 0$; and

(ii) $E[I[c_i - x_i'\beta > \eta_0] \cdot x_i x_i'] = E\left[S_{c|x}(x_i'\beta_0 + \eta_0) \cdot x_i x_i'\right]$ is nonsingular for some $\eta_0 > 0$

The following theorem establishes the root-n consistency and asymptotic normality of our proposed minimum distance estimator. Due to its technical nature, we leave the proof to the appendix.

Theorem 2.2 *Under Assumptions A1-A3, C1-C3, and D1-D4*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{2.15}$$

where we define V as follows. Let $\mathcal{C}$ denote the subset of $\mathcal{X}$ where for all $x \in \mathcal{C}$, $P(c_i \geq x_i'\beta_0 | x_i = x) = 1$. Define the function

$$G(x_j, x_k) = \int f_\epsilon(0|x_i) x_i I[x_j \leq x_i \leq x_k] f_X(x_i) dx_i \tag{2.16}$$

where $f_X(\cdot)$ denotes the regressor density function. Then

$$V = 4E[I[[x_j, x_k] \subseteq \mathcal{C}] G(x_j, x_k) G(x_j, x_k)'] \tag{2.17}$$

Next we define the outer score term Ω. Now define the function define

$$G(x_j, x_k, x_i) = G(x_j, x_k) I[x_j \leq x_i \leq x_k]$$

from which we define

$$\bar{G}(x_i) = E[G(x_j, x_k, x_i) | x_i]$$

The outer score term is defined as

$$\Omega = E[\delta_{0i}\delta_{0i}']$$

with

$$\delta_{0i} = 2\bar{G}(x_i)(I[v_i \geq x_i'\beta_0] - d_i I[v_i \leq x_i'\beta_0])$$

To conduct inference, one can either adopt the bootstrap or consistently estimate the variance matrix.

3 Finite Sample Performance

The theoretical results of the previous section give conditions under which the randomly-censored regression quantile estimator will be well-behaved in large samples. In this section, we investigate the small-sample performance of this estimator by reporting results of a small-scale Monte Carlo study.

The model used in this simulation study is

$$y_i = \min\{\alpha_0 + x_i\beta_0 + \varepsilon_i, c_i\} \tag{3.1}$$

where the scalar regressor x_i has a standard normal distribution. The true values α_0 and β_0 of the parameters are 0.5 and 1, respectively. We considered two types of censoring- covariate independent censoring, where c_i was distributed chi-squared, one degree of freedom, and covariate dependent censoring, where we set $c_i = I[x_i \geq 0]*exp(x_i) - I[x_i \leq 0]*exp(x) + x_i*z_i$, where z_i was distributed standard normal.

We assumed the error distribution of ϵ_i was standard normal. In addition, we simulated designs with heteroskedastic errors as well: $\varepsilon_i = \sigma(x_i)\cdot\eta_i$, with η_i having a standard normal distribution and $\sigma(x_i) = \exp(x_i)$. For these designs, the overall censoring probabilities vary between 25% and 35%. For each replication of the model, the following estimators were calculated[4]:

a) The minimum distance least absolute deviations (MD) estimator introduced in this paper.

b) The randomly censored LAD introduced in Honoré et al.(2002), refereed to in the tables as HKP.

c) The estimator proposed by Buckley and James (1979);

d) The estimator proposed by Ying et al. (1995);

Both the Ying et al.(1995) and MD estimators were computed using the Nelder Meade simplex algorithm.[5]

[4]The simulation study was performed in GAUSS and C++. Codes for the estimators introduced in this paper are available from the authors upon request.

[5]OLS, LAD, and true parameter values were used in constructing the initial simplex for the results reported.

The randomly-censored least absolute deviations estimator (HKP) was computed using the iterative Barrodale-Roberts algorithm described by Buchinsky(1995)[6]; in the random censoring setting, the objective function can be transformed into a weighted version of the objective function for the censored quantile estimator with fixed censoring.

The results of 401 replications of these estimators for each design, with sample sizes of 50, 100, 200, and 400, are summarized in Tables I-IV, which report the mean bias, median bias, root-mean-squared error, and mean absolute error. These 4 tables corresponded to designs with 1)homoskedastic errors and covariate independent censoring, 2)heteroskedastic errors and covariate independent censoring, 3) homoskedastic errors and covariate dependent censoring, and 4) heteroskedastic errors and covariate dependent censoring. Theoretically, only the MD estimator introduced here is consistent in all designs, and the only estimator which is consistent in design 4.

HKP and Ying et al.(1995) estimators are consistent under designs 1 and 2, [7], whereas the Buckley-James estimator is inconsistent when the errors are heteroskedastic as is the case in designs 2 and 4.

The results indicate that the estimation method proposed here perform relatively well. For some designs the MD estimator exhibits large values of RMSE for 50 observations, but otherwise appears to be converging at the root $-n$ rate.

As might be expected, the MD estimator, which do not impose homoskedasticity of the error terms, is superior to Buckley-James when the errors are heteroskedastic. It generally outperforms HKP and Ying et al.(1995) estimator when the censoring variable depends on the covariates. This is especially the case when the sample size is 200 or larger. HKP preforms surprisingly well for small samples in design 4. However, its inconsistency is clearly reflected here as well as its bias and RMSE does not shrink with the sample size.

4 Conclusions

This paper introduces a new estimation procedure for an AFT model with conditional heteroskedasticity and very general censoring when compared to existing estimators in the

[6]OLS was used as the starting value when implementing this algorithm for the simulation study.

[7]Actually, the Ying et al.(1995) estimator are inconsistent for the regressor independent censoring designs as well, because of the bound on the support of c_i. However, as observations for which estimated values of the survivor function were close to 0 were "trimmed" away, this is not reflected in the simulation results.

literature. The procedure minimized a third order U-process, and did not required the estimation of the censoring variable distribution, nor did it require nonparametric methods and the selection smoothing parameters. The estimator was shown to have desirable asymptotic properties and a simulation study indicated adequate finite sample performance.

The results established in this paper suggest areas for future research. Specifically, the semiparametric efficiency bound for this general censoring model as yet to be derived, and it would be interesting to see how are MD estimator can be modified to attain the bound. Furthermore, it would be useful to see how if the identification methods used here can be modified to identify regression parameters in a panel data model with fixed effects. We leave these possible extensions for future research.

5 References

Billingsley (1995), *Probability and Measure*, New York, NY: Wiley and Sons.

Buckley, J. and I. James (1979), "Linear Regression with Censored Data," *Biometrika*, 66, 429-436.

Dominguez, M. A. and I. N. Lobato (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions", *Econometrica*, 72, 1601-1616.

Honoré, B.E. Khan, S. and J.L. Powell (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 64, 241-278.

Kaplan, E.L. and P. Meier (1958), "Nonparametric Estimation from Incomplete Data," *Journal of the American Statistical Association*, 53, 457-481.

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

Houl, H. (2002), *Weighted Empirical Processes in Dynamic Nonlinear Models* Springer-Verlag, New York, New York.

Koul, H., V. Susarla, and J. Van Ryzin (1981), "Regression Analysis with Randomly Right Censored Data," *Annals of Statistics*, 9, 1276-1288.

Meier, P. (1975), "Estimation of Distribution Functions ¿From Incomplete Data," in J. Gani, ed., *Perspectives in Probability and Statistics*. London: Academic Press.

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

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

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

Rosen, A. (2005), " Structural Estimation of Oligopoly Models", Working Paper.

Ritov, Y. (1990), "Estimation in a Linear Regression Model with Censored Data," *Annals of Statistics*, 18, 303-328.

Sherman, R.P. (1993), "The Limiting Distribution of the Maximum Rank Correlation Estimator", *Econometrica*, 61, 123-137

Sherman, R.P. (1994a), "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator", *Econometric Theory*, 10, 372-395

Sherman, R.P. (1994b), "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators ", *Annals of Statistics*, 22, 439-459

Wang, J.-G. (1987), "A Note on the Uniform Consistency of the Kaplan-Meier Estimator" *Annals of Statistics*, 15, 1313-1316

Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions", *Journal of the American Statistical Association*, 94, 137-145

Ying, Z., S.H. Jung, and L.J. Wei (1995), "Survival Analysis with Median Regression Models," *Journal of the American Statistical Association*, 90, 178-184

A Proof of Theorem 2.1

Here we verify the conditions in Theorem 2.1 in Newey and McFadden(1994). Identification follows from 2.2. Compactness and continuity follow from Assumptions C1 and C3 respectively. It remains to show uniform convergence of the sample objective function to $Q(\cdot)$. To establish this result we will define the following functions:

$$\begin{aligned}\hat{Q}_n(x_i, x_j\beta) &= \hat{H}_1(\beta, x_i, x_j)^2 I[\hat{H}_1(\beta, x_i, x_j) \geq 0] + \hat{H}_0(\beta, x_i, x_j)^2 I[\hat{H}_0(\beta, x_i, x_j) \leq 0] \\ Q_n(x_i, x_j, \beta) &= H_1(\beta, x_i, x_j)^2 I[H_1(\beta, x_i, x_j) \geq 0] + H_0(\beta, x_i, x_j)^2 I[H_0(\beta, x_i, x_j) \leq 0]\end{aligned}$$

To simplify our proof we assume here that the regressor values lie in a compact set. Then by the fact that the terms in the summation in $\hat{H}_1(x_i, x_j, \beta)$ and $\hat{H}_0(x_i, x_j, \beta)$ involve two indicator functions and hence are bounded, we can apply Lemma 2.4 in Newey and McFadden(1994) to conclude that

$$\sup |\hat{Q}_n(\cdot, \cdot, \cdot) - Q_n(\cdot, \cdot, \cdot)| \xrightarrow{p} 0 \tag{A.1}$$

Next, we will establish that

$$\sup_{\beta \in \mathcal{B}} |Q_n(\beta) - Q(\beta)| = o_p(1) \tag{A.2}$$

For this we can apply existing uniform laws of large numbers for $U-$ statistics. Specifically, we can show the r.h.s. of (A.2) is $O_p(n^{-1/2})$ by Corollary 7 in Sherman(1994a) since the functional space index by β is Euclidean for a constant envelope. The Euclidean property follows from example (2.11) in Pakes and Pollard(1989). ■

B Sketch of Proof of Theorem 2.2

As the proposed estimator is defined as the minimizer of a U-process, our proof strategy will be to provide a locally quadratic approximation function of the objective function. We adopt this strategy since the objective function is not smooth in the parameters. Quadratic approximation of objective functions have been provided in, for example, Pakes and Pollard(1989), Sherman(1993,1994a,b) and Newey and McFadden(1994).

Here, we follow the approach in Sherman(1994b). Having already established consistency of the estimator, we will first establish root-n consistency and asymptotic normality. For root-n consistency we will apply theorem 1 of Sherman(1994). Here, let $G_n(\beta)$ denote the sample objective function and let $G(\beta)$ denote the limiting objective function. From Theorem 1 in Sherman(1994b), sufficient conditions for root-n consistency are that

1. $\hat{\beta} - \beta_0 = o_p(1)$
2. There exists a neighborhood of β_0 and a constant $\kappa > 0$ such that $\mathcal{G}(\beta) - \mathcal{G}(\beta_0) \geq \kappa\|\beta - \beta_0\|^2$ for all β in this neighborhood.
3. Uniformly over $o_p(1)$ neighborhoods of β_0

$$\mathcal{G}_n(\beta) = \mathcal{G}(\beta) + O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2)$$

Once root-n consistency has been established we can apply Theorem 2 in Sherman(1994b) to attain asymptotic normality. A sufficient condition is that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of β_0,

$$\mathcal{G}_n(\beta) - \mathcal{G}_n(\beta_0) = \frac{1}{2}(\beta - \beta_0)'V(\beta - \beta_0) + \frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n + o_p(\frac{1}{n}) \tag{B.1}$$

where W_n converges in distribution to a $N(0, \Omega)$ random vector, and V is positive definite. In this case the asymptotic variance of $\hat{\beta} - \beta_0$ is $V^{-1}\Omega V^{-1}$.

We will turn immediately to (B.1). Here, we will work with the U-statistic decomposition in, for example, Serfling(1980) as our objective function is a third-order U-process. We will first derive an expansion for $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$. We denote that even though $\mathcal{G}_n(\beta)$ is not differentiable in β, $\mathcal{G}(\beta)$ is sufficiently smooth for Taylor expansions to apply as the expectation operator is a smoothing operator. Taking a second order expansion of $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$, we obtain

$$\mathcal{G}(\beta) = \mathcal{G}(\beta_0) + \nabla_\beta \mathcal{G}(\beta_0))'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta^*))(\beta - \beta_0) \tag{B.2}$$

where ∇_β and $\nabla_{\beta\beta}$ denote first and second derivative operators and, and $\beta^a st$ denotes an intermediate value. We note that the first two terms of the right hand side of the above equation are 0, the first by how we defined the objective function, and the second by our identification result in Theorem 1. We will thus evaluate the following result:

$$\nabla_{\beta\beta}\mathcal{G}(\beta^*)) = V + o_p(1) \tag{B.3}$$

To formally show the above result, we first expand $\hat{H}_1(\beta, x_j, x_k)^2$ in the double summation.

$$\hat{H}_1(\beta, x_j, x_k)^2 = \frac{1}{n^2}\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})^2 I[x_j \leq x_i \leq x_k] \tag{B.4}$$

$$+ \frac{1}{n^2}\sum_{l \neq m}(I[v_l \geq x_l'\beta] - \frac{1}{2})(I[v_m \geq x_m'\beta] - \frac{1}{2}) \tag{B.5}$$

$$\cdot (I[v_l \geq x_l'\beta] - \frac{1}{2})I[x_j \leq x_l \leq x_k]I[x_j \leq x_m \leq x_k] \tag{B.6}$$

Similarly, we have:

$$\hat{H}_0(\beta, x_j, x_k)^2 = \frac{1}{n^2}\sum_{i=1}^{n}(\frac{1}{2} - d_i I[v_i \leq x_i'\beta])^2 I[x_j \leq x_i \leq x_k] \tag{B.7}$$

$$+ \frac{1}{n^2}\sum_{l \neq m}(\frac{1}{2} - d_l I[v_l \leq x_l'\beta]) \tag{B.8}$$

$$\cdot (\frac{1}{2} - d_m I[v_m \leq x_m'\beta])I[x_j \leq x_l \leq x_k]I[x_j \leq x_m \leq x_k] \tag{B.9}$$

At this stage, we will concentrate on the $\hat{H}_1(x_j, x_k, \beta)$ half of the objective function and make 2 adjustments to this half of the objective function, both of which will be shown to not affect the limiting distribution theory. The first that the squared terms above in the summation across i can be ignored in the asymptotic theory due to fact we are summing across n terms with the constant being $\frac{1}{n^2}$. The second is that we will replace the indicator function in the objective function

$I[\hat{H}_1(x_j, x_k, \beta) \geq 0]$ with $I[H_1(x_j, x_k, \beta_0) \geq 0]$. This also will not affect the limiting distribution theory and is analogous to results found in Powell(1984) and Khan and Powell(2001). We will thus apply Theorem 2 in Sherman(1994b) to the following 4-th order U-statistic:

$$\frac{1}{n(n-1)}\sum_{j \neq k} I[H_1(x_j, x_k, \beta_0) \geq 0]\frac{1}{n^2}\sum_{l \neq m} I[x_j \leq x_l \leq x_k]I[x_j \leq x_m \leq x_k]\times \tag{B.10}$$

$$(I[v_l \geq x_l'\beta] - \frac{1}{2})(I[v_m \geq x_m'\beta] - \frac{1}{2}) \tag{B.11}$$

We first note that $H_1(x_j, x_k, \beta_0) \geq 0)$ implies that the interval $[x_j, x_k] \subseteq \mathcal{C}$. Thus the other indicators in the above summation imply that both $x_l, x_m \in \mathcal{C}$. This in turn will imply that the expectation of the above summation will be 0 when $\beta = \beta_0$. We will evaluate the expectation as a function of β and expand around β_0.

First, we condition on the regressors x_j, x_k, x_l, x_m, we get

$$I[H_1(x_j, x_k, \beta_0) \geq 0]I[x_j \leq x_l \leq x_k]I[x_j \leq x_m \leq x_k]\times \tag{B.12}$$

$$(S_\epsilon(x_l'(\beta-\beta_0))S_c(x_l'\beta) - \frac{1}{2})(S_\epsilon(x_m'(\beta-\beta_0))S_c(x_m'\beta) - \frac{1}{2}) \tag{B.13}$$

Where $S_\epsilon(\cdot), S_c(\cdot)$ denote the conditional survivor functions of ϵ_i and c_i respectively. Expanding around β_0, we note the first term and the derivative term are 0 since both $x_l, x_m \in \mathcal{C}$. The second derivative term in the expansion is of the form

$$I[H_1(x_j, x_k, \beta_0) \geq 0]I[x_j \leq x_l \leq x_k]I[x_j \leq x_m \leq x_k](\beta-\beta_0)' f_\epsilon(0|x_l) f_\epsilon(0|x_m) x_l x_m' (\beta-\beta_0) \tag{B.14}$$

Next, we take expectations of the above term conditional on x_j, x_k. First, take the expectation with respect to x_l, conditioning on x_j, x_k, ignoring for now the terms involving x_m. This gives the following function of x_j, x_k:

$$G(x_j, x_k) = \int f_\epsilon(0|x_l) x_l I[x_j \leq x_l \leq x_k] f_X(x_l) dx_l \tag{B.15}$$

where $f_X(\cdot)$ denotes the regressor density function. A similar expression can be found when integrating with respect to x_m. Finally, we can apply the same arguments to the other "half" of the objective function involving $H_0(x_j, x_k, \beta_0)$.

Combining all these results we may conclude that form of V in the quadratic approximation in Theorem 2 in Sherman(1994b) is of the form

$$V = 4E[I[[x_j, x_k] \subseteq \mathcal{C}]G(x_j, x_k)G(x_j, x_k)'] \tag{B.16}$$

We next turn attention to the deriving the form of the score term in Theorem 2 in Sherman(1994b). As before, we will work with the first "half" of the objective function, and the 4th order U-process in (B.10). To proceed, we will first replace the term $(I[v_l \geq x_l'\beta] - \frac{1}{2})$ with

$(I[v_l \geq x_l'\beta_0] - \frac{1}{2})$. The resulting remainder term can be shown to be asymptotically negligible uniformly in β in an $O_p(n^{-1/2})$ neighborhood of β_0 using arguments in Pakes and Pollard (1989) and Sherman(1994a,b).

With this replacement, we can look at the form of first order projection terms in the U-statistic decomposition in Sherman(1994b). Recall this involves the sample means of the conditional expectations of each of the four arguments. The conditional expectation given the subscript j, k, m terms are each 0 due to the fact that $x_l \in \mathcal{C}$ and we have replaced β with β_0 in that term. Therefore, all that remains is the conditional expectation of (B.10) (after replacing β with β_0) in the term involving x_l. Here, we can use the same arguments as we did for the Hessian term, this time expanding

$S_\epsilon(x'_m(\beta - \beta_0))S_c(x'_m\beta)$ around $\beta = \beta_0$, conditioning on v_l, x_l.

Here, the first derivative term does not vanish as it did in the Hessian term. To get the form of this first derivative term, define $G(x_j, x_k)$ as before, and now define

$$G(x_j, x_k, x_l) = G(x_j, x_k)I[x_j \leq x_l \leq x_k]$$

from which we define

$$\bar{G}(x_l) = E[G(x_j, x_k, x_l)|x_l]$$

Thus the form of the first order projection terms in the U-statistic decomposition of the first "half" of the objective function is

$$(\beta - \beta_0)'\frac{1}{n}\sum_{l=1}^{n} 2\bar{G}(x_l)(I[v_l \geq x'_l\beta_0] - \frac{1}{2}) \tag{B.17}$$

plus an asymptotically negligible remainder term.

Similar arguments can be used for the other "half" of the objective function involving $H_0(x_j, x_k, \beta)$. Combining both pieces, we get the following representation for the linear term in Theorem 2 in Sherman(1994b)

$$(\beta - \beta_0)'\frac{1}{n}\sum_{l=1}^{n} 2\bar{G}(x_l)(I[v_l \geq x'_l\beta_0] - d_l I[v_l \leq x'_l\beta_0]) \tag{B.18}$$

plus an asymptotically negligible term.

Combining this result with our results for the Hessian term, and applying Theorem 2 in Sherman(1994b), we can conclude that

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{B.19}$$

where $\Omega = E[\delta_{0l}\delta_{0l}']$ with

$$\delta_{0l} = 2\bar{G}(x_l)(I[v_l \geq x_l'\beta_0] - d_l I[v_l \leq x_l'\beta_0])$$

Which established proof of the theorem. $\blacksquare$

TABLE I
Simulation Results for Censored Regression Estimators
CI Censoring, Homosked. Errors

	α				β			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
50 obs.								
MD	0.2424	0.1336	0.5719	0.2090	0.1774	0.1097	0.5033	0.2064
HKP	-0.0032	-0.0261	0.2714	0.1819	-0.0611	-0.0794	0.2728	0.1665
Buckley James	-0.0535	-0.0610	0.1875	0.1225	0.0041	-0.0121	0.1883	0.1353
Ying et al.	-0.0103	-0.0279	0.3480	0.1936	-0.1679	-0.1115	0.3853	0.1922
100 obs.								
MD	0.1136	0.0646	0.3156	0.1367	0.0892	0.0605	0.3151	0.1652
HKP	-0.0081	-0.0336	0.1965	0.1238	-0.0155	-0.0115	0.2091	0.1202
Buckley James	-0.0214	-0.0361	0.1341	0.0911	-0.0104	-0.0061	0.1264	0.0775
Ying et al.	0.0092	0.0188	0.1837	0.1050	-0.0799	-0.0553	0.2122	0.1126
200 obs.								
MD	0.0469	0.0275	0.2141	0.1180	0.0405	0.0292	0.2151	0.1309
HKP	0.0044	-0.0142	0.1305	0.0796	0.0073	-0.0070	0.1429	0.0798
Buckley James	-0.0085	-0.0004	0.1099	0.0699	-0.0010	0.0016	0.0925	0.0537
Ying et al.	-0.0154	-0.0116	0.1368	0.0911	-0.0425	-0.0321	0.1756	0.0797
400 obs.								
MD	0.0078	-0.0006	0.1486	0.0952	0.0119	-0.0008	0.1517	0.1000
HKP	-0.0045	-0.0031	0.1046	0.0598	0.0022	0.0111	0.0978	0.0669
Buckley James	-0.0168	-0.0065	0.0684	0.0490	-0.0083	-0.0131	0.0600	0.0491
Ying et al.	-0.0058	-0.0197	0.1145	0.0624	-0.0421	-0.0431	0.1204	0.0743

TABLE II

Simulation Results for Censored Regression Estimators

CI Censoring, Heterosked. Errors

	α				β			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
50 obs.								
MD	0.0659	0.0433	0.2490	0.1091	0.0612	0.0286	0.2484	0.0865
HKP	-0.0457	-0.0342	0.2708	0.1496	-0.0581	-0.0219	0.2315	0.0891
Buckley James	-0.6671	-0.5731	0.8346	0.5731	-0.9244	-0.7440	1.2189	0.7440
Ying et al.	-0.0614	-0.0561	0.4602	0.1862	-0.2161	-0.1226	0.6423	0.2120
100 obs.								
MD	0.0273	0.0175	0.1655	0.0853	0.0178	0.0075	0.1406	0.0673
HKP	-0.0237	-0.0210	0.1792	0.1138	-0.0204	-0.0210	0.1237	0.0606
Buckley James	-0.5309	-0.4637	0.6263	0.4637	-0.7941	-0.6857	0.9151	0.6857
Ying et al.	-0.0499	-0.0230	0.2698	0.1139	-0.0632	-0.0366	0.2554	0.1173
200 obs.								
MD	-0.0038	-0.0057	0.1261	0.0786	-0.0007	0.0012	0.0973	0.0533
HKP	-0.0119	-0.0116	0.1051	0.0737	-0.0069	-0.0041	0.0676	0.0367
Buckley James	-0.4420	-0.4401	0.4951	0.4401	-0.7528	-0.6889	0.8166	0.6889
Ying et al.	-0.0310	-0.0161	0.1820	0.0934	-0.0431	-0.0251	0.1826	0.0774
400 obs.								
MD	-0.0154	-0.0098	0.0905	0.0596	-0.0095	-0.0072	0.0587	0.0364
HKP	-0.0234	-0.0265	0.0852	0.0556	-0.0140	-0.0145	0.0492	0.0284
Buckley James	-0.4215	-0.4396	0.4450	0.4396	-0.7430	-0.7116	0.7709	0.7116
Ying et al.	-0.0302	-0.0550	0.1864	0.1143	-0.0180	-0.0511	0.2243	0.0866

TABLE III
Simulation Results for Censored Regression Estimators
CD Censoring, Homosked. Errors

	α				β			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
50 obs.								
MD	-0.0383	-0.0099	0.4130	0.2339	0.0182	0.0162	0.4734	0.2345
HKP	-0.0775	-0.0766	0.2398	0.1517	0.0187	0.0049	0.2784	0.1950
Buckley James	-0.0686	-0.0830	0.2241	0.1452	0.0170	0.0284	0.2046	0.1391
Ying et al.	0.0038	-0.0143	0.2750	0.1733	0.3062	0.2712	0.4041	0.2926
100 obs.								
MD	-0.0000	0.0062	0.2625	0.1488	-0.0139	-0.0105	0.2999	0.1785
HKP	-0.0527	-0.0330	0.1643	0.1091	0.0294	0.0425	0.1844	0.1359
Buckley James	-0.0290	-0.0262	0.1428	0.1057	0.0028	0.0104	0.0992	0.0641
Ying et al.	-0.0051	-0.0070	0.1714	0.1208	0.3451	0.3247	0.3830	0.3247
200 obs.								
MD	-0.0017	0.0007	0.2041	0.1245	-0.0106	-0.0110	0.2022	0.1199
HKP	-0.0242	-0.0333	0.1174	0.0831	0.0872	0.0719	0.1453	0.0845
Buckley James	-0.0214	-0.0208	0.0825	0.0595	-0.0022	0.0042	0.0749	0.0518
Ying et al.	0.0505	0.0658	0.1635	0.0948	0.3569	0.3381	0.3916	0.3381
400 obs.								
MD	0.0011	0.0030	0.1616	0.0982	-0.0129	-0.0086	0.1489	0.0927
HKP	-0.0127	-0.0097	0.0767	0.0510	0.0811	0.0782	0.1084	0.0808
Buckley James	0.0024	-0.0016	0.0623	0.0420	0.0062	0.0005	0.0583	0.0452
Ying et al.	0.0308	0.0224	0.1079	0.0633	0.4111	0.4304	0.4316	0.4304

TABLE IV
Simulation Results for Censored Regression Estimators
CD Censoring, Heterosked. Errors

	α				β			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
50 obs.								
MD	-0.0054	0.0000	0.3986	0.1732	0.0286	0.0160	0.5462	0.1904
HKP	0.0962	0.0895	0.3109	0.1947	0.1075	0.0923	0.2556	0.1490
Buckley James	0.1418	-0.0683	1.2248	0.4135	-0.8735	-0.6956	1.4538	0.6956
Ying et al.	-0.6507	-0.4609	2.0226	0.5182	-0.2543	-0.2327	1.2432	0.3681
100 obs.								
MD	0.0005	-0.0021	0.2583	0.1259	-0.0380	-0.0115	0.3956	0.1284
HKP	0.1282	0.1649	0.2467	0.1913	0.1295	0.1169	0.2120	0.1436
Buckley James	0.1910	0.1542	0.6758	0.3917	-0.6270	-0.5537	0.8489	0.5537
Ying et al.	-0.4324	-0.4731	0.6741	0.4948	-0.1263	-0.1504	0.4982	0.2620
200 obs.								
MD	-0.0048	0.0033	0.1997	0.1075	-0.0001	0.0050	0.2073	0.0745
HKP	0.1790	0.1780	0.2411	0.1780	0.1699	0.1733	0.2119	0.1733
Buckley James	0.2958	0.2106	0.6305	0.3071	-0.5496	-0.5003	0.7303	0.5318
Ying et al.	-0.3295	-0.4709	0.5915	0.4872	-0.1500	-0.1923	0.6122	0.2216
400 obs.								
MD	-0.0053	-0.0069	0.1478	0.0822	-0.0046	-0.0036	0.1152	0.0531
HKP	0.1911	0.1884	0.2192	0.1884	0.1607	0.1571	0.1810	0.1571
Buckley James	0.4339	0.3341	0.6524	0.3341	-0.5209	-0.5642	0.6425	0.5642
Ying et al.	-0.4116	-0.4891	0.4869	0.4930	-0.1054	-0.1815	0.3708	0.2025