Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities*

Shakeeb Khan† and **Elie Tamer**‡

December 2006

ABSTRACT. Under a conditional quantile restriction, randomly censored regression models can be written in terms of *conditional moment inequalities*. We study the identified features of these moment inequalities with respect to the regression parameters. These inequalities restrict the parameters to a set. We then show regular point identification can be achieved under a set of interpretable sufficient conditions. Our results generalize existing work on randomly censored models in that we allow for covariate dependent censoring, endogenous censoring and endogenous regressors. We then provide a simple way to convert conditional moment inequalities into unconditional ones while preserving the informational content. Our method obviates the need for nonparametric estimation, which would require the selection of smoothing parameters and trimming procedures. Maintaining the point identification conditions, we propose a quantile minimum distance estimator which converges at the parametric rate and has an asymptotically normal distribution. A small scale simulation study and an application using drug relapse data demonstrate satisfactory finite sample performance.

Key Words: Conditional Moment Inequality models, quantile minimum distance, covariate dependent censoring, heteroskedasticity, endogeneity.

---

*We thank K. Hirano, B. Honoré, J. Powell, and A. Rosen as well as seminar participants at Arizona, Cornell, Duke, OSU, NYU, Princeton, Syracuse, UIUC, University of Virginia and the 2005 World Congress meetings of the Econometric Society for many helpful comments. Both authors also gratefully acknowledge support from the National Science Foundation. Any errors are our own.

†Department of Economics, Duke University, 213 Social Sciences Building, Durham, NC 27708. Phone: (919) 660-1873.

‡Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Phone: (847) 491-8218.

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

$$y = x'\beta_0 + \epsilon \tag{1.1}$$

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*[1] (or AFT) model where y, typically the logarithm of survival time, is right censored at varying censoring points due usually either to data collection limitations or competing risks.

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. Under very weak assumptions on this joint distribution, one strand of the literature gives up on point (or unique) identification of β_0 and provides methods that estimate sets of parameters that are consistent with the assumptions imposed and the observed data. The bounds approach, explained in Manski and Tamer (2002), emphasizes robustness of results and clarity of assumptions imposed at the cost of estimating sets of parameters that might be large. Honore and Lleras-Muney(2006) derive bounds for a *competing risk* model, which is related to the randomly censored regression model above. The more common approach to studying the model above starts with a set of assumptions that guarantee point identification and then provides consistent estimators for the parameter vector of interest under these assumptions. Work in this area includes[2] the papers by Buckley and James (1979), Powell (1984), Koul, Susarla, and Ryzin (1981), Ying, Jung, and Wei (1995), Yang (1999), Buchinsky and Hahn (2001) Honoré, Khan, and Powell (2002) and, more recently Portnoy (2003) and Cosslett (2004). 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. Examples of assumptions which may be regarded as too strong are homoskedastic errors, censoring variables that are independent of the regressors and /or error terms, strong support conditions on the censoring variable which rule out fixed censoring, and exogenous regressors.

[1] An alternative class of models used in duration analysis is the (Mixed) Proportional Hazards Model. See Khan and Tamer(2005) and the references therein for recent developments in those models.

[2] None of these papers impose that c_i itself is restricted to behave like a mixed proportional hazards or accelerated failure time model, as was assumed in point identification results in Heckman and Honore(1989) and Abbring and van den Berg(2003). In many settings it is difficult to justify any modelling of the censoring process, and inconsistent result generally arise when this process is misspecified.

In this paper we aim to generalize existing work in many dimensions. In particular, we wish to construct an estimation procedure allows for endogenous regressors, the censoring variable to be endogenous (i.e. correlated with the error terms) and depend on the covariates in an arbitrary way, and also permits the error terms to be conditionally heteroskedastic.

Our approach to attaining such generalizations will be to write the above censored regression model as a conditional moment *inequality* model, of the form

$$E[m(z;\beta_0)\big|x] \leq 0 \tag{1.2}$$

where $m(.)$ is a known function (up to β_0), (z, x) are observed. This class of models has been studied recently in econometrics. See Manski and Tamer (2002), Chernozhukov, Hong, and Tamer (2002), Andrews, Berry, and Jia (2003), and Ho, Ishii, Pakes, and Porter (2005). Usually, in conditional moment equality models, if those moment conditions are satisfied uniquely at a given parameter, then one can obtain a consistent estimator of this parameter by considering a sample analog of an unconditional moment condition with an appropriate weight function. The choice of weight functions in moment equality models is motivated by efficiency gains (Dominguez and Lobato (2004) is an exception[3]) and not consistency.

Here, with moment inequality models, one needs to be careful about transforming conditional moment inequalities into unconditional ones since this might entail a loss of identification. This comes from the fact that, $E[m(z;\beta_0)|x)] \leq 0 \; \forall x \; a.e. \; \Rightarrow \; E[w(x)m(z;\beta_0)] \leq 0$ for an appropriate nonnegative weight function, but generally $E[w(x)m(z;\beta_0)] \leq 0 \not\Rightarrow E[m(z;\beta_0)|x)] \leq 0 \;\; x \, a.e.$

In one way to deal with this, Manski and Tamer (2002) estimated the conditional moment condition non-parametrically in a first step. This is not practically attractive since x might be multidimensional and nonparametric estimation requires choosing smoothing parameters and trimming procedures. On the other hand, Ho, Ishii, Pakes, and Porter (2005) use an unconditional version of the above moment inequality as the basis for inference. Ho, Ishii, Pakes, and Porter (2005) and Chernozhukov, Hong, and Tamer (2002) are concerned mainly with building confidence regions in this class of models when the parameter is not point identified (or partially identified).
In this paper, we use insights from Bierens (1990) and Chen and Fan (1999) where a conditional moment equality is transformed into an unconditional one in testing problems while preserving power. Although their weight functions do not generally apply to models with inequality restrictions, we will show that a variant of their approach that uses a more "localized" weight functions can be used for conducting inference on the parameters of interest. In

[3]We thank Adam Rosen for bringing this to our attention.

particular, we transform the conditional moment condition to an unconditional one using a class of *two sided* indicator functions. We show under sufficient conditions that the estimator based on this transformation is consistent and we derive its large sample distribution. While we illustrate this method in detail in the context of the randomly censored regression model, our approach can be used for any point identified conditional moment inequality model.

The next section describes the censored model studied in this paper in detail, and establishes the resulting moment inequalities. We then transform the randomly censored model with conditional inequality restriction into one with unconditional moment inequalities which will motivate our proposed minimum distance estimation procedure. We then establish the asymptotic properties for the proposed procedure, specifying sufficient regularity conditions for root-n consistency and asymptotic normality. Section 5 explains how to modify the proposed procedure to an i.v. (instrumental variable) type estimator with the availability of instruments. Section 6 explores the relative finite sample performance of the estimator in two ways- section 6.1 reports results from a simulation study, and 6.2 applies the estimator to explore a comparison of two courses of treatment for drug abuse. Section 7 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 Inequality Conditions in Randomly Censored Regression Models

Throughout the rest of this paper we will be concerned with inference on the k dimensional parameter vector β_0 in the model

$$y_i = x_i'\beta_0 + \epsilon_i \tag{2.1}$$

where x_i is a k-dimensional vector of covariates, β_0 is the unknown parameter of interest, and ϵ_i denotes the unobserved error term. Complications arise due to the censoring of the outcome y_i. In particular, we observe the random vector (x_i, v_i, d_i) such that

$$v_i = \min(y_i, c_i)$$

$$d_i = I(y_i < c_i)$$

where v_i is a scalar variable and d_i is a binary variable that indicates whether an observation is censored or not. The random variable c_i denotes the censoring variable that is only observed for censored observations, and $I[\cdot]$ denotes an indicator function, taking the value 1 if its argument is true and 0 otherwise. In the absence of censoring, $x_i'\beta_0 + \epsilon_i$ would be equal to the observed dependent variable, which in the accelerated failure time model context will usually be the log of survival time. In the censored model, the log-survival time is only partially observed. Another example of the above model can be a Roy/competing risk model where y can be denoted as the negative of wage in sector 1 and c is the negative of wage in sector 2 and one observes y for worker i in sector 1 if and only if $y_i > c_i$.

Next, we describe a set of assumptions that we use for inference on β_0. We first start with a conditional median assumption.

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

$$y_i = x_i'\beta_0 + \epsilon_i$$

This assumption restricts the conditional median of $\epsilon|x$. Our model is thus based on quantile restrictions on durations, which are similar to assumptions made in the quantile regression literature; see Koenker and Bassett (1978). The median restriction here is without loss of generality and our results apply for any quantile. In the case of censoring, our median restriction is similar to one used in other models in the literature- e.g. Powell (1984), Honoré, Khan, and Powell (2002) and Ying, Jung, and Wei (1995): it permits general forms of heteroskedasticity, and is weaker than the independence assumption $\epsilon_i \perp x_i$ as was imposed in Buckley and James (1979), Yang (1999), Portnoy (2003). This assumption alone, in the presence of random censoring, provides inequality restrictions on a set of appropriately defined functions. Let the functions $\tau_1(x_i, \beta)$ and $\tau_2(x_i, \beta)$ be defined as:

$$\begin{aligned}
\tau_1(x_i, \beta) &= E[I[v_i \geq x_i'\beta] \big| \, x_i] - \frac{1}{2} \\
\tau_0(x_i, \beta) &= E[(1 - d_i) + d_i I[v_i > x_i'\beta] \big| \, x_i] - \frac{1}{2} = \frac{1}{2} - E[d_i I[v_i \leq x_i'\beta]|x_i]
\end{aligned}$$

We can show that the above functions, when evaluated at the true parameter, satisfy *inequality restrictions.* This is described in the following lemma.

Lemma 2.1 *At the truth ($\beta = \beta_0$), and on S_X, the support*[4] *of x, the following holds:*

$$\forall x_i \in S_X, \;\; \tau_1(x_i, \beta_0) \leq 0 \leq \tau_0(x_i, \beta_0) \tag{2.2}$$

[4]Throughout we will be assuming that the regressors (with the exception of the constant corresponding to the intercept term in β) will be continuously distributed on S_X. This assumption is only made for notational convenience.

proof: First, we have for τ_1:

$$\begin{aligned}
\tau_1(x_i, \beta_0) &= E[I[v_i \geq x_i'\beta_0] \big| \, x_i] - \frac{1}{2} \\
&= E[\min(y_i, c_i) \geq x_i'\beta_0] \big| \, x_i] - \frac{1}{2} \\
&= P(\epsilon_i \geq 0; c_i \geq x\beta_0 \big| \, x_i] - \frac{1}{2} \\
&\leq P(\epsilon \geq 0 \big| \, x_i) = \frac{1}{2}
\end{aligned}$$

where the inequality follows from the fact that $\{\epsilon_i \geq 0; c_i \geq x\beta_0\} \subset \{\epsilon_i \geq 0\}$. Now, for τ_0:

$$\begin{aligned}
\tau_0(x_i, \beta_0) &= \frac{1}{2} - E[d_i I[v_i \leq x_i'\beta_0]|x_i] \\
&= \frac{1}{2} - P(y_i \leq c_i, y_i \leq x_i'\beta_0|x_i] \\
&= \frac{1}{2} - P(\epsilon_i \leq c_i - x_i'\beta_0, \epsilon_i \leq 0|x_i] \\
&\geq 0
\end{aligned}$$

where the inequality follows using similar arguments as above. ■

As we can see, the randomly censored regression model can be written as a conditional moment inequalities model. With only assumption **A.1**, the model provides the above inequality restrictions that hold at the true parameter β_0. So, these inequalities can be used to construct the *set* of parameters Θ_I that cannot be rejected as the truth. In general and under only **A.1**, this set is not a singleton. However, this set is robust to any kind of correlation between c and x and also c and ϵ and so allows for general types of censoring that can be random and can be endogenous, or dependent on y (or ϵ) conditional on x. The statistical setup is exactly one of competing risks where the risks are dependent, or a Roy model setup. We know there, that a general dependent competing risk model is not (point) identified- see, e.g. Petersen (1975). So, additional assumptions are needed to shrink the set to a point.

It is possible to come up with sufficient point identification conditions in some censored regression models. Heuristically, one looks for conditions under which for every $b \neq \beta_0$, we can find a set of positive measure for x, where one of the two inequalities above is reversed. For example, in the case of fixed (right) censoring at zero ($c = 0$), the set Θ_I shrinks to a point if the set of x's for which $x\beta_0$ is negative has positive mass and xx' is full rank on this set. This was essentially shown by Powell (1983). We extend this intuition to general setups where the censoring is allowed to be random and correlated with the regressors and error terms.

Let $\beta \neq \beta_0$ and let $x\delta = x\beta - x\beta_0$. Then, we have

$$\tau_1(x_i, \beta) = P(\epsilon \geq x_i'\delta; c \geq x_i'\beta|x) - \frac{1}{2} \tag{2.3}$$

$$\tau_0(x_i, \beta) = \frac{1}{2} - P(\epsilon \leq x'\delta; \epsilon \leq c - x'\beta_0|x) \tag{2.4}$$

So, a sufficient condition for point identification is for (2.3) to be positive for some x, *or* for (2.4) to be negative. One needs to find such x's *for every* $b \neq \beta_0$. A sufficient condition for this is the following assumption.

A2 The subset

$$\mathcal{C} = \{x_i \in S_X : \Pr(c_i \geq x_i'\beta_0|x_i) = 1\}$$

does not lie in a proper linear subspace of $\mathbf{R}^k$

A2 imposes relative support conditions on the censoring variable and the index. Specifically, it requires the regressor values for which the lower support point of the censoring distribution, (which we permit to vary with the regressor values) exceeds the index value.

This is a sufficient condition for identification and it is related to a notion of "regular identification" since the condition is necessary (but not sufficient) for root n estimation of β_0. For example, when (y, x, c) are jointly normal, we see that $\mathcal{C}$ has measure 0. This type of "identification at infinity" in the jointly normal model is delicate and typically leads to slower rates of convergence, analogous to results in Andrews and Schafgans (1998).

We note also that this condition easily allows for the fixed censoring case, and it reduces to the condition in Powell (1984) and Honoré, Khan, and Powell (2002). Our ability to accommodate the fixed censoring case is in contrast with other work which imposes relative support conditions- e.g. Ying, Jung, and Wei (1995) and Koul, Susarla, and Ryzin (1981). We also note that A2 does not impose relative support conditions on the latent error term, ϵ_i. This is in contrast to the condition in Koul, Susarla, and Ryzin (1981) which requires that the censoring variable c_i exceeds the support of $x_i'\beta_0 + \epsilon_i$, effectively getting identification off of the regressor values where censoring cannot occur.

Remark 2.1 *Neither A1 or A2 impose restrictions on the statistical dependence between the censoring variable, the regressors, and the latent error term. For example statistical independence between ϵ_i and (x_i, c_i) is imposed in Buckley James (1975) and Yang(1999), Portnoy(2003), and independence between c_i and (x_i, ϵ_i) was imposed in Yang (1999), Honoré,* *Khan, and Powell (2002)* [5]. *This is important in competing risks setups since assuming independence between durations is strong. In Roy model economies, independence is ruled out since one's skill in one sector is naturally correlated to his/her skill in another*[6] *sector. Our conditions above permit conditional heteroskedasticity, covariate dependent censoring, and even* endogenous censoring *(c_i dependent on ϵ_i), which is more general than the conditional independence condition $c_i \perp \epsilon_i | x_i$. Thus we can see that in the sense of our assumptions on the relationship between the censoring variable, the regressors, and the error terms, our assumptions are weaker than existing work on the censored AFT model.*

The next lemma provides a set of *inequality* restrictions that only hold at β_0. Those inequalities are in terms of the observed variables (v_i, d_i, x_i).

Lemma 2.2 *Let assumptions* **A1-A2** *hold. For each $\beta \neq \beta_0$*

$$P\big(x_i \in S_X: \quad \tau_1(x_i, \beta) > 0 \ \mathit{or} \ \tau_0(x_i, \beta) < 0\big) > 0 \tag{2.5}$$

and the parameter of interest β_0 is point identified.

Lemma 2.1 states that at the true value the function τ_1 is negative and τ_0 is positive everywhere on the support of x_i, while lemma 2.2 states that at $\beta \neq \beta_0$, either τ_1 is strictly positive *somewhere* on the support of x_i or τ_0 is strictly negative somewhere on the support of x_i. Hence, under A2 (and A1), the true parameter β_0 is point identified.

Proof of Lemma 2.2: Note equations (2.3) and (2.4) and consider first the case when $x\delta < 0$. Then for x satisfying condition A2 for β,

$$\begin{aligned} x\delta < 0 \implies (2.3) \quad &\geq \quad P(\epsilon \geq x_i'\delta; c \geq x_i'\beta_0 | x) - \frac{1}{2} \\ &= \quad P(\epsilon \geq x_i'\delta | x) - \frac{1}{2} \\ &> \quad 0 \end{aligned}$$

[5] Both of these papers suggest methods to allow for the censoring variable to depend on the covariates. These methods involve replacing the Kaplan-Meier procedure they require with a conditional Kaplan Meier which will require the choice of smoothing parameters to localize the Kaplan-Meier procedure, as well as trimming functions and tail behavior regularity conditions. Furthermore, they do not suggest on how to permit the censoring to be endogenous.

[6] In that literature, one models the process for c also, typically assuming that $c_i = z_i'\gamma_0 + \nu_i$, where exclusion restrictions (some variable in z are not in x) and support or continuity condition deliver identification of β_0 (and γ_0). See Heckman and Honoré (1988).

where the equality follows the fact that x belongs to the special set postulated in A2, and the strict inequality follows from 0 being the *unique* conditional median.
Now for the case when $x\delta > 0$. Consider x satisfying the condition in A2. Hence for that x, we have

$$x\delta > 0 \implies (2.4) \quad < 0$$

where the inequality follows from the fact that $x\delta$ is positive and $c \geq x\beta_0$. ■

Remark 2.2 *Condition* **A2** *is sufficient for the model to point identify the parameter, but if A2 is considered too strong in some settings, then one can maintain only assumption A1 and consistently estimate the set of parameters which include the truth, β_0, using the estimator we propose below (this estimator is "adaptive" in the sense that it estimates the identified set, which under A2 is the singleton β_0). Another avenue would be to impose more assumptions that guarantee point identification, such as independence between c and ϵ as in Honoré, Khan and Powell (2002). Of course, the former method is robust to such independence assumptions at the cost of obtaining set estimates and having to conduct set based inference. So, in some empirical settings, it is plausible to maintain independence between c and ϵ. These include some statistical experiments where the censoring is random and is set independently of the process that generated the outcome (usually the censoring is part of the experimental design). In other settings, especially in economic applications, independence and even conditional independence can be suspect, especially when the censoring can be affected by unobservables that also affect outcomes and so maintaining these assumption would lead to inconsistent estimates (that might not fall in the identified set). So, if one does not insist on point identification, using only A1 delivers a set of parameters and hence there, work is needed to study the topological properties of the set such as whether the set is finite, convex, etc*

Remark 2.3 *Our identification result in Lemmas 2.1 and 2.2 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, Jung, and Wei (1995) which is only based on the function $\tau_1(\cdot,\cdot)$, and consequently requires reweighing the data using the Kaplan-Meier estimator. As alluded to previously, this imposes support conditions which can rule out, among other things, fixed censoring, and does not allow for covariate dependent censoring, unless one uses the conditional Kaplan Meier estimator of Beran (1981). Reweighing the data by the conditional Kaplan Meier is complicated since it involves smoothing parameters and trimming procedures, and furthermore, it does not address the problem of endogenous censoring.*

3 Estimation: Transforming the Conditional Moment Inequalities Model

In this section, we propose an objective function that can be used to conduct inference on the parameter β_0. We study the large sample properties of the extrema of this objective function in the case of point identification, i.e. when A2 above holds. In cases where A2 does not hold, one can use set estimation methods such as those in Chernozhukov, Hong, and Tamer (2002) using the same objective function.

Since our identification results are based on *conditional* moment inequalities holding for all x, one might be tempted to use similar moments that are unconditional[7] on x. However, in general, this strategy might yield a loss of information, i.e., the implied model is not able to point identify β_0. So, to ensure identification, an estimator must preserve the information contained in the conditional inequalities, i.e., ensure that the inequalities hold *for all* x, a.e. Our estimation procedure will have to differ from frameworks used to translate a conditional moment model (based on equality constraints) into an unconditional moment model while ensuring global identification of the parameters of interest. To avoid estimating conditional distributions (which involve smoothing parameters), we extend insights from works by Bierens (1990), Stute (1986), Koul (2002), Chen and Fan (1999) and recently Dominguez and Lobato (2004) and transform the conditional moment inequalities into unconditional ones while preserving the informational content.

We first define the following functions of β, 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 let $H_1(.)$ and $H_2(.)$ be defined as follows:

$$H_1(\beta, t_1, t_2) = E\left\{\tau_1(x_i;\beta)I[t_1 \leq x_i \leq t_2]\right\} = E\left\{[I[v_i \geq x_i'\beta] - \frac{1}{2}]I[t_1 \leq x_i \leq t_2]\right\} \quad (3.1)$$

$$H_0(\beta, t_1, t_2) = E\left\{\tau_0(x_i;\beta)I[t_1 \leq x_i \leq t_2]\right\} = E\left\{[\frac{1}{2} - d_i I[v_i \leq x_i'\beta]]I[t_1 \leq x_i \leq t_2]\right\} \quad (3.2)$$

where the above inequality $t_1 \leq x_i$ is to be taken componentwise. From fundamental properties of expectations,[8] conditional moment conditions can be related to unconditional moment

[7]For example, $E[m(y;x)|x] \leq 0 \Rightarrow E[w(x)m(y;x)] \leq 0$, where $w(x)$ is an appropriate positive weight function.

[8]See for example Shiryaev (1984) page 185.

conditions with indicator functions as above comparing regressor values to all vectors on the support of x. As we will see below, this translates into estimation procedures involving a third order U-process. The crucial point to notice about the functions H_0, H_1 is that although they preserve the information contained in τ_0 and τ_1 above, they are not conditional on x. This means that our procedures will not involve estimating conditional probabilities, and hence there will be no need for choosing smoothing parameters or employing trimming procedures.

In general, one can transform a model with inequality moment condition such as

$$E[m(z;\beta_0)|x] \le 0$$

into an informationally equivalent unconditional model as

$$H(\beta_0, x_1, x_2) = E[m(z;\beta_0)1[x_1 \le x \le x_2] \le 0$$

for all x_1, x_2 where x_1, x_2, x are iid. 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:

$$Q(\beta) = E_{x_j,x_k}\big[H_1(\beta, x_j, x_k)I[H_1(\beta, x_j, x_k) \ge 0] - H_0(\beta, x_j, x_k)I[H_0(\beta, x_j, x_k) \le 0]\big] \quad (3.3)$$

The main identification result is stated in the following lemma:

Lemma 3.1 *Under Assumptions A1, $Q(\beta)$ is minimized at the set Θ_I, where*

$$\Theta_I = \{\beta \in R^k : \tau_1(x,\beta) \le 0 \le \tau_0(x,\beta) \ a.e.\, x\}$$

In addition, if A2 holds, then $Q(\beta)$ is minimized uniquely at $\beta = \beta_0$.

Proof: We prove the point identification result under A1 and A2. We first show that

$$Q(\beta_0) = 0 \quad (3.4)$$

To see why note this follows directly from the previous lemmas which established that $\tau_1(x_i,\beta_0)I[\tau_1(x,\beta_0) \ge 0] = \tau_0(x_i,\beta_0)I[\tau_0(x_i,\beta_0) \le 0] = 0$ 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^*, $\max(\tau_1(x,\beta)I[\tau_1(x,\beta) \ge 0], -\tau_0(x,\beta)I[\tau_0(x,\beta) \le 0]) > 0$ for any $\beta \ne \beta_0$. 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 $\beta \neq \beta_0$. ■
Moreover, note here that we are implicitly assuming that all the regressors in x are continuously distributed. In case some take discrete values, we can rewrite the moment inequality into a vector of inequality restrictions conditional on the continous regressors where each vector is evaluated at a value of the discrete regressors. Hence, assuming that we only have continuous regressors is without loss of generality and is done here for simplicity.

4 Consistency and Asymptotic Normality

Having shown global identification, we propose an estimation procedure, which is based on the analogy principle, and thus 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:

$$\widehat{H}_1(\beta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})I[x_j \leq x_i \leq x_k] \tag{4.1}$$

$$\widehat{H}_0(\beta, x_j, x_k) = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2} - d_i I[v_i \leq x_i'\beta])I[x_j \leq x_i \leq x_k] \tag{4.2}$$

Then, our estimator $\hat{\beta}$ of β_0 is defined as follows:

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} \widehat{Q}_n(\beta) \tag{4.3}$$

$$= \arg\min_{\beta \in \mathcal{B}} \frac{1}{n(n-1)} \sum_{j \neq k} \Big\{ \hat{H}_1(\beta, x_j, x_k) I[\hat{H}_1(\beta, x_j, x_k) \geq 0] \tag{4.4}$$

$$- \hat{H}_0(\beta, x_j, x_k) I[\hat{H}_0(\beta, x_j, x_k) \leq 0] \Big\}$$

where $\mathcal{B}$ is the parameter space to be defined below.

Remark 4.1 *We note the above objective is similar to a standard LAD/median regression objective function, since for a random variable z, we can write $|z| = zI[z \geq 0] - zI[z \leq 0]$. The difference lies in the fact that our objective function "switches" from $\hat{H}_1(\cdot, \cdot, \cdot)$ to $\hat{H}_0(\cdot, \cdot, \cdot)$ when moving from the positive to the negative region. Switching functions is what*

permits us to allow for general forms of censoring.
Note also that the above estimation procedure minimizes a (generated) second order U-process. We note also that analogously to existing rank estimators, (e.g., Han (1987), Khan and Tamer (2005)), this provides us with an estimation procedure without the need to select smoothing parameters or trimming procedures. However, our estimator is more computationally involved than the aforementioned rank estimators, as the functions inside the double summation have to be estimated themselves, effectively resulting in our objective function being a third order U-process.

We next turn attention to the asymptotic properties of the estimator in (4.3). To ensure root-n consistency, we will strengthen Assumption A2 to require the set $\mathcal{C}$ to have positive measure:

A2' The matrix $E[I[x_i \in \mathcal{C}]x_i x_i']$ is of full rank.

We begin by establishing consistency under the following assumptions.

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

C2 We have an iid sample $(d_i, v_i, x_i')'$, $i = 1, \ldots, n$.

C3 $Q(\beta) \equiv E[H_1(\beta, x_j, x_k)I[H_1(x_j, x_k, \beta) \geq 0] - E[H_0(\beta, x_j, x_k)I[H_1(x_j, x_k, \beta) \leq 0]$ is continuous at $\beta = \beta_0$.

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

Theorem 4.1 *Under Assumptions A1,A2', and C1-C3,*

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

For 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 ϵ_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 regressors x_i and censoring values $\{c_i\}$ satisfy

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

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 4.2 *Under Assumptions A1,A2', C1-C3, and D1-D3*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{4.5}$$

where we define V and Ω are as follows. Let

$$\mathcal{C} = \{x : P(c_i \geq x_i'\beta_0 | x_i = x) = 1\}$$

Adopting the notation $I_{ijk} = I[x_j \leq x_i \leq x_k]$, define the function

$$G(x_j, x_k) = I[[x_j, x_k] \subseteq \mathcal{C}] \int f_\epsilon(0|x_i) x_i I_{ijk} f_X(x_i) dx_i \tag{4.6}$$

where $f_X(\cdot)$ denotes the regressor density function. The Hessian matrix is

$$V = 2E[G(x_j, x_k)G(x_j, x_k)'] \tag{4.7}$$

Next we define the outer score term Ω. Let

$$\delta_{0i} = E[G(x_j, x_k) I_{ijk} | x_i](I[v_i \geq x_i'\beta_0] - d_i I[v_i \leq x_i'\beta_0]) \tag{4.8}$$

so we can define the outer score term Ω as

$$\Omega = E[\delta_{0i}\delta_{0i}'] \tag{4.9}$$

To conduct inference, one can either adopt the bootstrap or consistently estimate the variance matrix, using a "plug-in" estimator for the separate components. As is always the case with median based estimators, smoothing parameters will be required to estimate the error conditional density function, making the bootstrap a more desirable approach.

We conclude this section by commenting on computational issues. The above estimator involves optimizing a third order U-statistic. Significant computational time can be reduced if one uses a "split sample" approach (see Honoré and Powell (1994) for an example) would result in an estimator that minimizes a second order U-process that is much simpler computationally, though less efficient. For the problem at hand, the split sample version of our proposed estimator would minimize an objective function of the form:

$$\begin{aligned}
\hat{\beta}_{SS} &= \arg\min_{\beta\in\mathcal{B}} \hat{Q}_{SSn}(\beta) &(4.10)\\
&= \arg\min_{\beta\in\mathcal{B}} \frac{1}{n}\sum_{j=1}^{n}\Big\{\hat{H}_1(\beta,x_j,x_{n-j+1})I[\hat{H}_1(\beta,x_j,x_{n-j+1})\geq 0] &(4.11)\\
&- \hat{H}_0(\beta,x_j,x_{n-j+1})I[\hat{H}_0(\beta,x_j,x_{n-j+1})\leq 0]\Big\}
\end{aligned}$$

5 Endogenous Regressors and Instrumental Variables

In this section we illustrate how the estimation procedure detailed in the previous section can be modified to permit consistent estimation of β_0 when the regressors x_i as well as the censoring variable c_i are endogenous. Our identification strategy now requires one to have a vector of instrumental variables z_i.

The semiparametric literature has seen recent developments in estimating censored regression models when the regressors are endogenous. For fixed censoring models, instrumental variable approaches have been proposed in Hong and Tamer (2003) and Lewbel (2000), whereas a control function approach has been proposed in Blundell and Powell (2004), Blundell and Powell (2003), for nonlinear models including censored regression. However, these estimators are not applicable in the random censoring case, even in the case when the censoring variable is distributed independently of the covariates. Furthermore, they all require the selection of multiple smoothing parameters and trimming procedures.

Here we propose an estimator for the randomly censored regression model with endogenous regressors. This problem arises in a variety of settings in duration analysis. For example, in labor economics, if the dependent variable is unemployment spell length, an explanatory variable such as the amount of training received while unemployed could clearly be endogenous. Another example, studied more often in the biostatistics literature, is when the dependent variable is time to relapse for drug abuse and the endogenous explanatory variable is a course of treatment.

To estimate β_0 in this setting we assume the availability of a vector of instrumental variables z_i which will be defined through the following assumptions. Here, our sufficient conditions for identification are:

A'1 median$(\epsilon_i|z_i) = 0$.

A'2 The subset of the support of instruments

$$\mathcal{C}_{Z0} = \{z_i : P(x_i'\beta_0 \leq c_i|z_i) = 1\}$$

has positive measure. Furthermore, for each $\delta \neq 0$, either of the following subsets of $\mathcal{C}_{Z0}$ also has positive measure:

$$\mathcal{C}_{Z0-} = \{z_i \in \mathcal{C}_{Z0} : P(x_i'\delta \leq 0|z_i) = 1\}$$

$$\mathcal{C}_{Z0+} = \{z_i \in \mathcal{C}_{Z0} : P(x_i'\delta \geq 0|z_i) = 1\}$$

Remark 5.1 *Before outlining an estimation procedure, we comment on the meaning of the above conditions.*

1. *Condition A'1 is analogous to the usual condition of the instruments being uncorrelated with the error terms.*

2. *Condition A'2 details the relationship between the instruments and the regressors. It is most easily satisfied when the exogenous variable(s) has a support that is relatively large when compared to the support of the endogenous variable(s).*[9] *Empirical settings where this support condition arises is in the treatment effect literature, where the endogenous variable is a binary treatment variable, or a binary compliance variable. In the latter case an example of an instrumental variable is also a binary variable indicating treatment assignment if it is done so randomly- see for example Bijwaard and Ridder (2005) who explore the effects of selective compliance to re-employment experiments on unemployment duration.*

Our IV limiting objective function is of the form:

$$Q_{IV}(\beta) = E\left[H_1^*(\beta, z_j, z_k)I[H_1^*(\beta, z_j, z_k) \geq 0] - H_0^*(\beta, z_j, z_k)I[H_0^*(\beta, z_j, z_k) \leq 0]\right] \quad (5.1)$$

[9] Lewbel(2000) also requires an exogenous variable with relatively large support. However, his conditions are stronger than those imposed here in the sense that he assumes the exogenous variable has to have large support when compared to both the endogenous variable(s) *and* the error term, effectively relying on identification at infinity. In the censored setting, his assumption corresponds to obtaining identification from the region of exogenous variable(s) space where the data is uncensored with probability 1.

where here

$$H_1^*(\beta, t_1, t_2) = E[(I[v_i \geq x_i'\beta] - \frac{1}{2})I[t_1 \leq z_i \leq t_2]] \tag{5.2}$$

$$H_0^*(\beta, t_1, t_2) = E[(\frac{1}{2} - d_i I[v_i \leq x_i'\beta])I[t_1 \leq z_i \leq t_2]] \tag{5.3}$$

And our proposed IV estimator minimizes the sample analog of $Q_{IV}(\beta)$, using

$$\hat{H}_1(\beta, z_j, z_k) = \frac{1}{n}\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})I[z_j \leq z_i \leq z_k] \tag{5.4}$$

$$\hat{H}_0(\beta, z_j, z_k) = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2} - d_i I[v_i \leq x_i'\beta])I[z_j \leq z_i \leq z_k] \tag{5.5}$$

The asymptotic properties of this estimator are based on regularity conditions analogous to those in the previous section and we omit the details here.

6 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 in two ways, first by reporting results of a small-scale Monte Carlo study, and then considering an empirical illustration. Specifically, we first study the effects of two courses of treatment for drug abuse on the time to relapse ignoring potential endogeneity, and then we control for selective compliance to treatment using our proposed i.v. estimator.

6.1 Simulation Results

The model used in this simulation study is

$$y_i = \min\{\alpha_0 + x_{1i}\beta_0 + x_{2i}\gamma_0 + \epsilon_i, c_i\} \tag{6.1}$$

where the regressor x_{1i}, x_{2i} were chi-squared, 1 degree of freedom, and standard normal respectively. The true values α_0, β_0, γ_0 of the parameters are -0.5, -1, and 1, respectively. We considered two types of censoring- covariate independent censoring, where c_i was distributed standard normal, and covariate dependent censoring, where we set $c_i = -x_{1i}^2 - x_{2i}$.[10]

We assumed the error distribution of ϵ_i was standard normal. In addition, we simulated designs with heteroskedastic errors as well: $\epsilon_i = \sigma(x_i) \cdot \eta_i$, with η_i having a standard normal distribution and $\sigma(x_i) = \exp(0.5 * x_i)$. For these designs, the overall censoring probabilities vary between 40% and 50%. For each replication of the model, the following estimators were calculated[11]:

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

b) The randomly censored LAD introduced in Honoré, Khan, and Powell (2002), referred to as HKP.

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

d) The estimator proposed by Ying, Jung, and Wei (1995), referred to as YJW;

Both YJW and MD estimators were computed using the Nelder Meade simplex algorithm.[12] The randomly-censored least absolute deviations estimator (HKP) was computed using the iterative Barrodale-Roberts algorithm described by Buchinsky(1995)[13]; 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-V, 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, 4) heteroskedastic errors and covariate dependent censoring, 5) endogenous censoring. Theoretically, only the MD estimator introduced here is consistent in all designs, and the only estimator which is consistent in designs 4 and 5,.

[10]We note that for this design our set $\mathcal{C}$ defined in Assumption **A2** does not have positive measure, violating the sufficient condition for regular identification of our estimator. Nonetheless, as we see in the simulation tables our estimator performs relatively well in finite samples.

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

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

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

HKP and YJW estimators are consistent under designs 1 and 2, 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 YJW estimator when the censoring variable depends on the covariates. This is especially the case when the errors are heteroskedastic, as in this design, the proposed estimator is the only estimator which performs reasonably well. Table 5 indicates that MD is the only consistent estimator for the endogenous censoring design, as it is the only estimators whose values of bias and RMSE decline with sample size.

6.2 Empirical Example: Drug Relapse Duration

We apply the minimum distance procedure to the drug relapse data set used in Hosmer and Lemeshow (1999), who study the effects of various variables, on time to relapse. Those not relapsing before the end of the study are regarded as censored. Similar data were used in Portnoy (2003).

The data is from the University of Massachusetts Aids Research Unit Impact Study. Specifically, the data set is from a 5-year (1989-1994) period comprising of two concurrent randomized trials of residential treatment for drug abuse. The purpose of the original study was to compare treatment programs of different planned durations designed to prevent drug abuse and to also determine whether alternative residential treatment approaches are variable in effectiveness. One of the sites, referred to here as site A randomly assigned participants to 3- and 6-month modified therapeutic communities which incorporated elements of health education and relapse prevention. Here clients were taught how to recognize "high-risk" situations that are triggers to relapse, and taught the skills to enable them to cope with these situations without using drugs. In the other site, referred to here as site B, participants were randomized to either a 6-month or 12-month therapeutic community program involving a highly structured life-style in a communal living setting. This data set contains complete records of 575 subjects.

Here, we use the log of relapse time as our dependent variable, and the following six independent variables: SITE(drug treatment site B=1, A=0), IV(an indicator variable taking the value 1 if subject had recent IV drug use at admission), NDT(number of previous drug abuse treatments), RACE(white(0) or "other"(1)), TREAT (randomly assigned type of treatment, 6 months(1) or 3 months(0)), FRAC (a proxy for compliance, defined as the fraction of length of stay in treatment over length of assigned treatment). Table V reports results for the 4 estimators used in the simulation study as well as estimators of two parametric models- the Weibull and Log-Logistic. Standard errors are in parentheses.

Qualitatively, all estimators deliver similar results in the sense that the signs of the coefficients are the same. However there are noticeable differences in the values of these estimates, as well as their significance[14]. For example, the Weibull estimates are noticeably different from all other estimates, including the other parametric estimator, in most categories, showing a larger (in magnitude) IV effect, which is statistically insignificant for many of the semiparametric estimates, and a smaller TREAT effect. The semiparametric estimators differ both from the parametric estimators as well as each other. The proposed minimum distance estimator, consistent under the most general specifications compared to the others, yields a noticeably smaller (in magnitude) SITE effect, and with the exception of the HKP estimator, a larger TREAT effect.

We extend our empirical study by applying the IV extension of the proposed minimum distance estimator to the same data set. The explanatory variable length of stay (LOS) could clearly be endogenous because of "selective compliance". Specifically, those who comply more with treatment (i.e. have larger values of LOS) may not be representative of the people assigned treatment, in which case the effect of an extra day of treatment would be overstated by estimation procedures which do not control for this form of endogeneity. Given the random assignment of the type of treatment, the treatment indicator (TREAT) is a natural choice (see, e.g. Bloom (1984)) of an instrumental variable as it is correlated with LOS.

We consider estimating a similar model to one considered above, now modelling the relationship between the log of relapse time and the explanatory variables IV, RACE, NDT, SITE, and LOS. Table VI reports results from 4 estimation procedures: 1) ordinary least squares (OLS) 2) two stage least squares (2SLS) using TREAT as an instrument for LOS 3) our proposed minimum distance estimator (MD) and 4) our proposed extension to allow for endogeneity (MDIV) using TREAT as an instrument for LOS. We note that OLS and 2SLS are not able to take into account the random censoring in our data set. Nonetheless the

[14]Reported standard errors for the 4 semiparametric estimators were obtained from the bootstrap, using 575 samples (obtained with replacement) of 575 observations.

results from OLS and 2SLS are similar to MD and MDIV respectively. Most importantly both procedures do indeed indicate selective compliance to treatment as the estimated coefficient on LOS is larger for OLS and MD that it is for 2SLS and MDIV.

7 Conclusions

This paper introduces a new estimation procedure for models that are based on moment inequality conditions. The procedure is applied to estimate parameters for an AFT model with a very general censoring relationship when compared to existing estimators in the 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 of smoothing parameters and trimming procedures. The estimator was shown to have desirable asymptotic properties and both a simulation study and application using drug relapse data 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 has yet to be derived, and it would be interesting to see how are MD estimator can be modified to attain the bound in this specific model, as well as other models based on moment inequality conditions. We leave these possible extensions for future research.

References

ANDREWS, D., S. BERRY, AND P. JIA (2003): "On Placing Bounds on Parameters of Entry Games in the Presence of Multiple Equilibria," Working Paper.

ANDREWS, D. W. K., AND M. SCHAFGANS (1998): "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65, 487–517.

BERAN, R. (1981): "Nonparametric Regeression with Randomly Censored Survival Data," Technical Report, California - Berkeley.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

BIJWAARD, G., AND G. RIDDER (2005): "Correcting for Selective Compliance in a Reemployment Bonus Experiment," Working Paper, forthcoming in Journal of Econometrics.

BLOOM, H. (1984): "Accounting for No-shows in Experimental Evaluation Designs," *Evaluation Review*, 8, 225–246.

BLUNDELL, R., AND J. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Econometrics*, ed. by M. Dewatripont, L. Hanses, and S. Turnsovsky, pp. 312–357. Cambridge University Press.

——— (2004): "Endogeneity in Binary Response Models," *Review of Economic Studies*, 73.

BUCHINSKY, M., AND J. HAHN (2001): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.

BUCKLEY, J., AND I. JAMES (1979): "Linear Regression with Censored Data," *Biometrika*, 66, 429–436.

CHEN, X., AND Y. FAN (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91(2), 373–401.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2002): "Inference in Incomplete Econometric Models," Department of Economics, Princeton University.

COSSLETT, S. (2004): "Efficient Semiparametric Estimation of Censored and Truncated Regressions via a Smoothed Self-Consistency Equation," *Econometrica*, 72, 1277–1293.

DOMINGUEZ, M., AND I. LOBATO (2004): "Consitent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601–1615.

HAN, A. (1987): "The The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 303-316.

Ho, K., J. Ishii, A. Pakes, and J. Porter (2005): "Moment Inequalities and Their Applications," Working Paper.

Hong, H., and E. Tamer (2003): "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71(3), 905–932.

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

Honoré, B., and J. Powell (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64, 241–278.

Hosmer, D., and S. Lemeshow (1999): *Survival Analysis: Regression Modelling of Time to Event Data*. Wiley, New York, NY.

Khan, S., and E. Tamer (2005): "Partial Rank Estimation of Transformation Models with General forms of Censoring," Working Paper, forthcoming Journal of Econometrics.

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

Koul, H. (2002): *Weighted Empirical Processes Dynamic Nonlinear Models*. Springer-Verlag, New York, NY.

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

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145–177.

Manski, C., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

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

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

Serfling, R. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York, NY.

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

——— (1994a): "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators," *Annals of Statistics*, 22, 439–459.

——— (1994b): “U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator,” *Econometric Theory*, 10, 372–395.

Shiryaev, A. (1984): *Probbaility*. Springer-Verlag, New York, NY.

Stute, W. (1986): “On Almost Sure Convergence of Conditional Empirical Distribution Functions,” *Annals of Statistics*, 14(3), 891–901.

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. Jung, and L. Wei (1995): “Survival Analysis with Median Regression Models,” *Journal of the American Statistical Association*, 90, 178–184.

A Proof of Theorem 4.1

Heuristically, to establish consistency, we need identification, compactness, continuity and uniform convergence (See for example Theorem 2.1 in Newey and McFadden(1994)). Identification follows from Lemma 3.1. 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 to ease notation, we will show that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}\hat{H}_1(x_j,x_k,\beta)I[\hat{H}_1(x_j,x_k,\beta)\geq 0]-Q_1(\beta)\right|=o_p(1) \tag{A.1}$$

where

$$Q_1(\beta)=E[H_1(x_j,x_k,\beta)I[H_1(x_j,x_k,\beta)\geq 0]] \tag{A.2}$$

noting that identical arguments can be used for the component of the objective function involving $\hat{H}_0(x_j,x_k,\beta)$. To show (A.1) we will first show that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}\hat{H}_1(x_j,x_k,\beta)I[\hat{H}_1(x_j,x_k,\beta)\geq 0]-H_1(x_j,x_k,\beta)I[H_1(x_j,x_k,\beta)\geq 0]\right| \tag{A.3}$$

is $o_p(1)$. To show (A.3) is $o_p(1)$, we first replace $I[\hat{H}_1(x_j,x_k,\beta)\geq 0]$ with $I[H_1(x_j,x_k,\beta)\geq 0]$. Since the indicator function is bounded, we will next attempt to show that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}\hat{H}_1(x_j,x_k,\beta)-H_1(x_j,x_k,\beta)\right|=o_p(1) \tag{A.4}$$

To do so, we expand $\hat{H}_1(x_j,x_k,\beta)$, which recall involved a summation of observations denoted by subscript i. The term

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}(I[v_i\geq x_i'\beta]-\frac{1}{2})I[x_j\leq x_i\leq x_k]-H_1(x_j,x_k,\beta)$$

is a mean 0 third order U-process. Consequently, by applying Corollary 7 in Sherman(1994a), this term is uniformly $o_p(1)$. It remains to show that replacing $\hat{H}_1(x_j,x_k,\beta)$ with $H_1(x_j,x_k,\beta)$ inside the indicator function yields an asymptotically uniformly negligible remainder term. Specifically, we will show that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}I[\hat{H}_1(x_j,x_k,\beta)\geq 0]-I[H_1(x_j,x_k,\beta)\geq 0]\right|=o_p(1) \tag{A.5}$$

To do so, we will show that $I[\hat{H}_1(x_j, x_k, \beta) \geq 0]I[H_1(x_j, x_k, \beta) < 0]$ is $o_p(1)$ uniformly in x_j, x_k, β. We fix the values of x_j, x_k and decompose $I[H(x_j, x_k, \beta) < 0]$ as

$$I[H_1(x_j, x_k, \beta) < 0] = I[H_1(x_j, x_k, \beta) \leq -\delta_n] + I[H_1(x_j, x_k, \beta) \in (-\delta_n, 0)] \tag{A.6}$$

where $\delta_n = \frac{1}{\log n}$. We will first focus on the term

$$I[\hat{H}_1(x_j, x_k, \beta) \geq 0]I[H_1(x_j, x_k, \beta) \leq -\delta_n] \tag{A.7}$$

The probability that the above product of indicators is positive, recalling that we are fixing x_j, x_k, is less than or equal to the probability that

$$\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})I_{ijk} - H_1(x_j, x_k, \beta) > -nH_1(x_j, x_k, \beta) \tag{A.8}$$

with $H_1(x_j, x_k, \beta) \leq -\delta_n$, where above $I_{ijk} = I[x_j \leq x_i \leq x_k]$. The probability of the above term is less than or equal to the probability that

$$\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2})I_{ijk} - H_1(x_j, x_k, \beta) > n\delta_n \tag{A.9}$$

To which we can apply Hoeffding's inequality, see, e.g. Pollard(1984), to bound above by

$\exp(-2n\delta_n^2)$.

Note this bound is independent of x_j, x_k, β, and converges to 0 at the rate n^{-2}, establishing the uniform (across x_j, x_k, β) convergence of $I[\hat{H}(x_j, x_k, \beta) \geq 0]I[H(x_j, x_k, \beta) \leq -\delta_n]$ to 0. We next show the uniform convergence of

$$\frac{1}{n(n-1)}\sum_{i \neq j} I[\hat{H}_1(x_j, x_k, \beta) \geq 0]I[H_1(x_j, x_k, \beta) \in (-\delta_n, 0)] \tag{A.10}$$

for which it will suffice to establish the uniform convergence of:

$$\frac{1}{n(n-1)}\sum_{i \neq j} I[H_1(x_j, x_k, \beta) \in (-\delta_n, 0)] \tag{A.11}$$

Subtracting $E[I[H_1(x_j, x_k, \beta) \in (-\delta_n, 0)]]$ from the above summation we can again apply Corollary 7 in Sherman(1994a) to conclude that this term is uniformly in β $o_p(1)$. The expectation $E[I[H(x_j, x_k, \beta) \in (-\delta_n, 0)]]$ is uniformly $o_p(1)$ by applying the dominated convergence theorem. Combining all our results we conclude that (A.3) holds.

Next, we will establish that

$$\sup_{\beta \in \mathcal{B}} \left| \frac{1}{n(n-1)} \sum_{i \neq j} H_1(x_j, x_k, \beta)I[H_1(x_j, x_k, \beta) \geq 0] - Q_1(\beta) \right| = o_p(1) \tag{A.12}$$

For this we can apply existing uniform laws of large numbers for centered $U-$ processes. Specifically, we can show the r.h.s. of (A.12) 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 Proof of Theorem 4.2

Having shown consistency, our proof strategy will be to approximate the objective function $\widehat{Q}_n(.)$, locally in a neighborhood of β_0, by a an appropriate quadratic in β function. The approximation needs to allow for the fact that this objective function is not smooth in β. Quadratic approximation of objective functions have been provided in, for example, Pakes and Pollard(1989), and Sherman (1994a), (1994b),(1993) among others. First we establish root-n consistency. For root-n consistency we will apply Theorem 1 of Sherman (1994b). Keeping our notation deliberately close to Sherman(1994b), here we denote our sample objective function $\hat{Q}_n(\beta)$ by $\mathcal{G}_n(\beta)$ and denote our limiting objective function $Q(\beta)$ by $\mathcal{G}(\beta)$. From Theorem 1 in Sherman(1994b), sufficient conditions for rates of convergence 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

$$\mathcal{G}_n(\beta) = \mathcal{G}(\beta) + O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + O_p(n^{-1}) \tag{B.1}$$

which suffices for $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$. In what follows throughout the rest of our proofs, it will prove convenient to subtract the following term, which does not depend on β, from our objective function:

$$\frac{1}{n(n-1)} \sum_{j \neq k} \hat{H}_1(x_j, x_k, \beta_0) I[H_1(x_j, x_k, \beta_0) \geq 0] - \hat{H}_0(x_j, x_k, \beta_0) I[H_0(x_j, x_k, \beta_0) \leq 0] \tag{B.2}$$

We note that since β does not enter (B.2), the value of the estimator is not affected by including this additional term. We also note that expectation of this term conditional on x_j, x_k is 0.

To show the second condition, we will first derive an expansion for $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$. We note 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 and the smoothness conditions in Assumptions **D2**, **D4 (i)**. Taking a second order expansion of $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$, we obtain

$$\mathcal{G}(\beta) = \mathcal{G}(\beta_0) + \nabla_\beta \mathcal{G}(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta^*)(\beta - \beta_0) \tag{B.3}$$

where ∇_β and $\nabla_{\beta\beta}$ denote first and second derivative operators and β^* 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 later formally show that

$$\nabla_{\beta\beta}\mathcal{G}(\beta_0) = V \tag{B.4}$$

and V is invertible by Assumption **A2'**, so we have

$$(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta_0)(\beta - \beta_0) > 0 \tag{B.5}$$

$\nabla_{\beta\beta}\mathcal{G}(\beta)$ is also continuous at $\beta = \beta_0$ by Assumptions **C3** and **D4 (i)**, so there exists a neighborhood of β_0 such that for all β in this neighborhood, we have

$$(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta)(\beta - \beta_0) > 0 \tag{B.6}$$

which suffices for the second condition to hold.

To show the third condition, our first step is to replace the indicator functions in the objective function, $I[\hat{H}_1(x_j, x_k, \beta) \geq 0], I[\hat{H}_0(x_j, x_k, \beta) \leq 0]$, with the functions $I[H_1(x_j, x_k, \beta) \geq 0], I[H_0(x_j, x_k, \beta) \leq 0]$ respectively, and derive the corresponding representation. (We will deal with the resulting remainder term from this replacement shortly.) We expand the terms $\hat{H}_1(x_j, x_k, \beta)$ and $\hat{H}_0(x_j, x_k, \beta)$, first, exclusively dealing with the first expansion since the second is similar. This results in the third order U-process:

$$\frac{1}{n(n-1)(n-2)} \sum_{j \neq k \neq l} I[H_1(x_j, x_k, \beta) \geq 0](I[v_l \geq x_l'\beta] - \frac{1}{2})I_{ljk} \equiv \frac{1}{n(n-1)(n-2)} \sum_{j \neq k \neq l} m(z_j, z_k, z_l) \tag{B.7}$$

where $z_i = (x_i, v_i)$, $I_{ljk} = I[x_j \leq x_l \leq x_k]$. B.7 is a third order U-statistic and we analyze its properties by representing it as a projection plus a degenerate U-process- see, e.g. Serfling (1980). Note that the unconditional expectation corresponds to the first "half" of the limiting objective function, which recall here we denoted by $\mathcal{G}(\beta)$. We will evaluate representations for expectations conditional on each of the three arguments, minus the unconditional expectation. In particular, first we write B.7 as:

$$B.7 = P_n Pm(z_j, ., .) + P_n Pm(., z_k, .) + P_n P(., ., z_l) + U_n h \tag{B.8}$$

where $P_nPm(z_j, ., .)$ is equal to $\frac{1}{n}\sum_j P_{k,l}m(z_j, z_k, z_l)$ and U_nh is a degenerate U-statistic (see Sherman (1994a)). Hence, to get the first order term in B.1, we need to take the first order expansion of the projection terms in B.8.

We first turn attention to the expectation conditional on the third argument l. This summation will be of the form

$$\frac{1}{n}\sum_{l=1}^{n}(I[v_l \geq x_l'\beta] - \frac{1}{2})E\left[I[H_1(x_j, x_k, \beta) \geq 0]I_{ljk}|\, x_l\right] \tag{B.9}$$

To get the $O_p(\frac{1}{\sqrt{n}})$ term in B.1, we will take a mean value expansion of the term inside the above summation around $\beta = \beta_0$. Note by our normalization (i.e the subtraction of (B.2) from the objective function), the initial replacement of β with β_0 yields a term that is $o_p(\frac{1}{n})$ uniformly in β in $o_p(1)$ neighborhoods of β_0. Notice also that $I[H_1(x_j, x_k, \beta_0) \geq 0] \cdot I_{ljk} = 1$ implies that $[x_j, x_k] \subset \mathcal{C}$, as is x_l; we next evaluate

$$\nabla_\beta E\left[I[H_1(x_j, x_k, \beta) \geq 0]I_{ljk}|x_l\right] \tag{B.10}$$

at $\beta = \beta_0$. Here by definition of H_1, we have:

$$H_1(x_j, x_k, \beta) = \int I[x_j \leq u \leq x_k]\left\{P\big(c \geq u'\beta; \epsilon \geq u'(\beta - \beta_0)|u\big) - \frac{1}{2}\right\} f_X(u)du \tag{B.11}$$

When integrating over the different values of x_j, x_k, we decompose the set of values into those satisfying the interval $[x_j, x_k]$ contained in $\mathcal{C}$ and those that do not. We do this because $H_1(x_j, x_k, \beta_0) < 0$ in the latter case and we are outside of the range of integration. Two applications of the dominated convergence theorem on the subset of values where $[x_j, x_k]$ is contained in $\mathcal{C}$ yields that (B.10) is of the form:

$$E\left[G(x_j, x_k)\right] = E\left[I[[x_j, x_k] \subseteq \mathcal{C}]\int f_{\epsilon|X}(0|u)I_{ujk}uf_X(u)du\right]$$

This is so since $\nabla_\beta P\big(c \geq u'\beta; \epsilon \geq u'(\beta - \beta_0)|u\big) = -f_{\epsilon|x}(0)$ on the set $\mathcal{C}$. So combining this expansion term with (B.9), yields

$$\frac{1}{n}\sum_{l=1}^{n} I[x_l \in \mathcal{C}](I[v_l \geq x_l'\beta_0] - \frac{1}{2})E\left[G(x_j, x_k)\right]'(\beta - \beta_0) \tag{B.12}$$

Next we evaluate

$$\frac{1}{n}\sum_{l=1}^{n} \nabla_\beta(I[v_l \geq x_l'\beta] - \frac{1}{2})E\left[I[H_1(x_j, x_k, \beta_0) \geq 0]I_{ljk}|\, x_l\right] \tag{B.13}$$

This term will cancel out with the corresponding derivative term from the $H_0(\cdot, \cdot, \cdot)$ "half" of the objective function, completing the representation of the linear term in our expansion, which remains as is in (B.12). We note that using similar arguments, along with the law of large numbers, it follows that the remainder term in the mean value expansion of $(I[v_l \geq x_l'\beta] - \frac{1}{2})E[I[H_1(x_j, x_k, \beta) \geq 0]]$ yields a term that is $o_p(\|\beta - \beta_0\|^2)$. Also, we note that the expectation of this term, which must be subtracted from our $U-$statistic decomposition can be shown to be negligible using similar arguments.

As far as the other two projections in B.8, it can be shown that the expectation conditional j, when combined with the corresponding conditional expectation of the other "half" of the objective function, has the representation:

$$o_p(\|\beta - \beta_0\|^2) \tag{B.14}$$

as does the expectation conditional on the second argument, indexed by k. So, the $O_p(||\beta - \beta_0||/\sqrt{n})$ term on B.1 is B.12 and since it has expectation (across v_l, x_l) of 0, and finite variance it is bounded in probability by a CLT. So we have established that so far the sample objective function can be represented as the limiting objective function plus

$$O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + o_p(\frac{1}{n}) \tag{B.15}$$

Finally, we note the higher order terms in the projection theorem are $o_p(n^{-1})$ uniformly for β in $o_p(1)$ neighborhoods of β_0 using arguments similar to those in Theorem 3 in Sherman(1993). So by Theorem 1 in Sherman(1994b), we are able to show, that it can be expressed as the limiting first "half" of the objective function plus a remainder term that is (uniformly in $o_p(1)$ neighborhoods of β_0),

$$O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + o_p(\frac{1}{n}) \tag{B.16}$$

Collecting terms, we conclude that the infeasible estimator, which replaced $I[\hat{H}_1(x_j, x_k, \beta) \geq 0]$, $I[\hat{H}_0(x_j, x_k, \beta) \geq 0]$ with $I[H_1(x_j, x_k, \beta) \geq 0]$, $I[H_0(x_j, x_k, \beta) \geq 0]$ respectively, is $O_p(n^{-1/2})$.

To derive a rate of convergence for the actual estimator, we will derive a rate for:

$$\sum_{j \neq k} (I[\hat{H}_1(x_j, x_k, \beta) \geq 0] - I[H_1(x_j, x_k, \beta) \geq 0])\hat{H}_1(x_j, x_k, \beta) \tag{B.17}$$

as well as the second "half" involving $H_0(x_j, x_k, \beta)$. To establish the negligibility of (B.17), we will first show that

$$\frac{1}{n(n-1)} \sum_{j \neq k} I[\hat{H}_1(x_j, x_k, \beta) < 0] I[H_1(x_j, x_k, \beta) \geq 0] = o_p(\log n n^{-1/2}) \tag{B.18}$$

uniformly in β in $o_p(1)$ neighborhoods of β_0. To do so, we decompose

$$I[\hat{H}_1(x_j, x_k, \beta) < 0] = I[\hat{H}_1(x_j, x_k, \beta) < -\delta_n] + I[\hat{H}_1(x_j, x_k, \beta) \in [\delta_n, 0)] \tag{B.19}$$

where δ_n is a sequence of positive numbers converging to 0 at the rate $\log n/\sqrt{n}$. We aim to show each of the above indicator functions multiplied by $I[H_1(x_j, x_k, \beta) \geq 0]$ corresponds to an $o_p(n^{-1/2})$ term, uniformly in β in $o_p(1)$ neighborhoods of β_0. First, dealing with

$$\frac{1}{n(n-1)} \sum_{j \neq k} I[\hat{H}_1(x_j, x_k, \beta) < -\delta_n] I[H_1(x_j, x_k, \beta) \geq 0] \tag{B.20}$$

For any β, we will show that (conditioning on x_j, x_k),

$$P(\hat{H}_1(x_j, x_k, \beta) < -\delta_n) \to 0 \tag{B.21}$$

whenever $H_1(x_j, x_k, \beta) \geq 0$. We note the above probability can be expressed as:

$$P(\sum_{i=1}^{n} (\frac{1}{2} - I[v_i \geq x_i'\beta]) I_{ijk} + H_1(x_j, x_k, \beta) > n(\delta_n + H_1(x_j, x_k, \beta))) \tag{B.22}$$

which is less than or equal to

$$P(\sum_{i=1}^{n} (\frac{1}{2} - I[v_i \geq x_i'\beta]) I_{ijk} + H_1(x_j, x_k, \beta) > n\delta_n) \tag{B.23}$$

since $H(x_j, x_k, \beta) \geq 0$. Note we can apply Hoeffding's inequality (see, e.g. Pollard(1984), page 191) to the above probability since $(\frac{1}{2} - I[\cdot])$ is bounded between $\frac{-1}{2}$ and $\frac{1}{2}$. The resulting exponential bound for the above probability is $\exp(-2n\delta_n^2)$ which converges to 0 at the rate n^{-2}. We next show that

$$P(\hat{H}_1(x_j, x_k, \beta) \in [-\delta_n, 0)) \to 0 \tag{B.24}$$

for β uniformly in an $o_p(1)$ neighborhood of β_0 and $H_1(x_j, x_k, \beta) \geq 0$. We decompose the above probability into

$$P(\hat{H}_1(x_j, x_k, \beta) \in [\delta_n, 0)), 0 \leq H_1(x_j, x_k, \beta) \leq 2\delta_n) \tag{B.25}$$

$$+P(\hat{H}_1(x_j, x_k, \beta) \in [-\delta_n, 0)), H_1(x_j, x_k, \beta) > 2\delta_n) \tag{B.26}$$

The first piece of the decomposition is less than or equal to

$$P(|H_1(x_j, x_k, \beta)| \leq 2\delta_n)(\delta_n) \tag{B.27}$$

uniformly in β in $o_p(1)$ neighborhoods of β_0 where the equality uses Assumption **D4(i)**. The second probability can be bounded above by:

$$P(|\hat{H}_1(x_j, x_k, \beta) - H_1(x_j, x_k, \beta)| \geq \delta_n) \tag{B.28}$$

which can be written as

$$P(|\sum_{i=1}^{n}(I[v_i \geq x_i'\beta] - \frac{1}{2}) - H_1(x_j, x_k, \beta)| \geq n\delta_n) \tag{B.29}$$

To which we can apply the corollary to Hoeffding's inequality (see, e.g. Pollard(1984)) to conclude that the above probability is bounded above by $exp(-2n\delta_n^2)$ which converges to 0 at the rate n^{-2}. This establishes (B.18). Similar arguments can be used to show that

$$\frac{1}{n(n-1)}\sum_{j\neq k} I[\hat{H}_1(x_j, x_k, \beta) \geq 0]I[H_1(x_j, x_k, \beta) < 0]_p(\log n n^{-1/2}) \tag{B.30}$$

uniformly for β in $o_p(1)$ neighborhoods of β_0. We can now get the desired rate of convergence of our estimator, $O_p(n^{-1/2})$, by extending the above arguments to establish the following interaction term

$$\frac{1}{n(n-1)}\sum_{j\neq k} \hat{H}_1(x_j, x_k, \beta)I[\hat{H}_1(x_j, x_k, \beta) \geq 0]I[H_1(x_j, x_k, \beta) < 0] \tag{B.31}$$

is $O_p(n^{-1})$. We proceed as before, letting δ_n denote a sequence of positive numbers that is $O(\sqrt{\log n/n})$. Again we will decompose the indicator $I[\hat{H}_1(x_j, x_k, \beta) \geq 0]$ into the sum of two indicators:

$$I[\hat{H}_1(x_j, x_k, \beta) \geq \delta_n] + I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n)] \tag{B.32}$$

Using Hoeffding's inequality as before we can conclude that

$$\frac{1}{n(n-1)}\sum_{j\neq k} \hat{H}_1(x_j, x_k, \beta)I[\hat{H}_1(x_j, x_k, \beta) \geq \delta_n]I[H_1(x_j, x_k, \beta) < 0] \tag{B.33}$$

is $O_p(n^{-2})$. Next we establish the rate of convergence of

$$\frac{1}{n(n-1)}\sum_{j\neq k} \hat{H}_1(x_j, x_k, \beta)I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n)]I[H_1(x_j, x_k, \beta) < 0] \tag{B.34}$$

which is less than or equal to

$$\frac{1}{n(n-1)}\delta_n\sum_{j\neq k} I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n)]I[H_1(x_j, x_k, \beta) < 0] \tag{B.35}$$

Note we can now replace $I[H_1(x_j, x_k, \beta < 0]$ with $I[H_1(x_j, x_k, \beta) \in [-\delta_n, 0)]$, and the remainder term is $O_p(n^{-2})$ using the same exponential bounds as before. So it remains to establish a rate of convergence for

$$\frac{1}{n(n-1)}\delta_n\sum_{j\neq k} I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n)]I[H_1(x_j, x_k, \beta) \in [-\delta_n, 0)] \tag{B.36}$$

We note the above expression is $o_p(\frac{1}{n})$ uniformly in β in $o_p(1)$ neighborhoods of β_0 as we have already established that

$I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n)]$ is $O_p(\delta_n)$ uniformly in x_j, x_k and similar arguments can be used to show that

$I[H_1(x_j, x_k, \beta) \in [-\delta_n, 0)]$ is (uniformly) $O_p(\delta_n)$ as well.

Consequently, we conclude that $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$. We can now turn attention to establishing asymptotic normality of the estimator.

Now that 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,

$$\mathcal{G}_n(\beta) - \mathcal{G}_n(\beta_0) = \frac{1}{2}(\beta - \beta_0)'V(\beta - \beta_0) + \frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n + o_p(\frac{1}{n}) \tag{B.37}$$

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 to (B.37). Here, we will again 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 $\mathcal{G}(\beta_0)$, since $\mathcal{G}(\beta)$ is related to the limiting objective function. 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 by Assumptions **D2**, **D4(i)**. Taking a second order expansion of $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$, we obtain

$$\mathcal{G}(\beta) = \mathcal{G}(\beta_0) + \nabla_\beta \mathcal{G}(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta^*)(\beta - \beta_0) \tag{B.38}$$

where ∇_β and $\nabla_{\beta\beta}$ denote first and second derivative operators and, and β^* 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 show the following result:

$$\nabla_{\beta\beta}\mathcal{G}(\beta^*) = V + o_p(1) \tag{B.39}$$

The form of the matrix V is as the second derivative with respect to β of the following function evaluated at $\beta = \beta_0$.

$$E[H_1(x_j, x_k, \beta)I[H_1(x_j, x_k, \beta) \geq 0]] - E[H_0(x_j, x_k, \beta)I[H_0(x_j, x_k, \beta) \leq 0]] \tag{B.40}$$

Note that by definition:

$$H_1(x_j,x_k,\beta) = \int_{x_j}^{x^k} \left(P\big(c \geq u'\beta; \epsilon \geq u'(\beta-\beta_0)|u\big) - \frac{1}{2}\right) f_X(u)du$$

$$H_0(x_j,x_k,\beta) = \int_{x_j}^{x^k} \left(\frac{1}{2} - \int^{u'(\beta-\beta_0)} \int_{e+x\beta_0} f_{(c,\epsilon)|u}(c,e)dcde\right) f_X(u)du$$

So, the second derivative is:

$$V = E\left[\nabla_{\beta\beta}H_1 I_1 + 2\nabla_\beta H_1 \nabla_\beta I_1 + H_1 \nabla_{\beta\beta} I_1 - \nabla_{\beta\beta}H_0 I_0 - 2\nabla_\beta H_0 \nabla_\beta I_0 - H_0 \nabla_{\beta\beta} I_0\right] \quad \text{(B.41)}$$

where $I_1 = I[H_1(x_j,x_k;\beta_0) \geq 0]$ and similarly for I_0. The above expression for V can be simplified. For example, we have $\nabla_{\beta\beta}H_1(x_j,x_k;\beta) - \nabla_{\beta\beta}H_0(x_j,x_k;\beta) = 0$ on the set $\mathcal{C}$.

Next, notice that by a simple integration by parts argument,

$$E[\nabla_\beta H_1 \nabla_\beta I_1 + H_1 \nabla_{\beta\beta} I_1] = 0 \quad \text{(B.42)}$$

and similarly for its H_0 part. Hence, what remains is

$$\begin{aligned} V &= E\left[\nabla_\beta H_1 \nabla_\beta I_1 - \nabla_\beta H_0 \nabla_\beta I_0\right] && \text{(B.43)} \\ &= 2E\left[1[[x_j,x_k] \subset \mathcal{C}] \int_{x_j}^{x_k} x f_\epsilon(0|x)dF_x \int_{x_j}^{x_k} x' f_\epsilon(0|x)dF_x\right] && \text{(B.44)} \\ &= 2E[I[[x_j,x_k] \subseteq \mathcal{C}]G(x_j,x_k)G(x_j,x_k)'] && \text{(B.45)} \end{aligned}$$

We next turn attention to the deriving the form of the outer product of the score term in Theorem 2 in Sherman(1994b). Note this was basically done in our arguments showing root-n consistency. This involves the conditional expectation, conditioning on each of the three arguments in the third order process, subtracting the unconditional expectation. We first condition on the first argument, denoted by the subscript j. Note here we are taking the expectation of the term $I[v_l \geq x_l'\beta] - \frac{1}{2}$ as well as $\frac{1}{2} - d_l I[v_l \leq x_l'\beta]$, so using the same arguments as we did for the unconditional expectation, the average of this conditional expectation is $O_p(\|\beta-\beta_0\|^2)/\sqrt{n}$, and thus asymptotically negligible for β in $O_p(n^{-1/2})$ neighborhoods of β_0. The same applies to the expectation conditional on the second argument of the third-order U-process, denoted by the subscript k.

We therefore turn attention to expectation conditional on the third argument, denoted by the subscript l. Here we proceed as before when showing root-n consistency, expanding

$$I[H_1(x_j,x_k,\beta) \geq 0]I[v_l \geq x_l'\beta] - \frac{1}{2} \quad \text{(B.46)}$$

around $\beta = \beta_0$. Recall this yielded the mean 0 process:

$$\frac{1}{n}\sum_{l=1}^{n} E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - \frac{1}{2}) \tag{B.47}$$

plus a negligible remainder term. Consequently, using the same arguments of half of the objective function involving $H_0(\cdot,\cdot,\cdot)$ we can express the linear term in our expansion (used to derive the form of the outer score term) as:

$$\frac{1}{n}\sum_{l=1}^{n} E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - d_l I[v_i \leq x_l'\beta_0])'(\beta - \beta_0) + o_p(n^{-1}) \tag{B.48}$$

which corresponds to

$$\frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n \tag{B.49}$$

$$W_n \Rightarrow N(0, E[\delta_{0l}\delta_{0l}']) \tag{B.50}$$

where

$$\delta_{0l} = E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - d_l I[v_i \leq x_l'\beta_0]) \tag{B.51}$$

This completes a representation for the linear term in the U-statistic representation. The remainder term, involving second and third order U-processes (see, e.g. equation (5) in Sherman(1994b), can be shown to be asymptotically negligible (specifically it is $o_p(n^{-1})$ uniformly in β in an $O_p(n^{-1/2})$ neighborhood of β_0 using Lemma 2.17 in Pakes and Pollard (1989) and Sherman(1994b) Theorem 3.

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

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{B.52}$$

where

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

Which establishes the proof of the theorem. ■

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	0.2411	0.1159	0.6098	0.1963	-0.1617	-0.0695	0.7169	0.1948
HKP	-0.0322	-0.0356	0.2746	0.2182	0.0009	0.0051	0.1600	0.1249
Buckley James	-0.0296	-0.0437	0.2124	0.1461	0.0025	-0.0049	0.1245	0.0693
YJW	-0.2163	-0.2288	0.3106	0.2534	0.0789	0.0655	0.1555	0.1158
100 obs.								
MD	0.1059	0.0594	0.2979	0.1232	-0.0824	-0.0408	0.2781	0.1106
HKP	-0.0111	-0.0112	0.1744	0.1375	0.0035	0.0017	0.0948	0.0754
Buckley James	-0.0235	-0.0273	0.1299	0.0918	0.0061	0.0052	0.0794	0.0535
YJW	-0.1524	-0.1441	0.2283	0.1839	0.0527	0.0370	0.1058	0.0783
200 obs.								
MD	0.0438	0.0368	0.1711	0.0935	-0.0235	-0.0126	0.1328	0.0735
HKP	0.0015	-0.0040	0.1332	0.1051	-0.0011	-0.0023	0.0686	0.0545
Buckley James	-0.0210	-0.0240	0.1048	0.0738	0.0083	0.0091	0.0525	0.0347
YJW	-0.0934	-0.0776	0.1655	0.1293	0.0300	0.0276	0.0696	0.0532
400 obs.								
MD	0.0095	0.0003	0.1056	0.0626	-0.0097	-0.0088	0.0808	0.0492
HKP	-0.0048	-0.0142	0.0945	0.0736	0.0042	0.0032	0.0492	0.0383
Buckley James	-0.0047	-0.0048	0.0725	0.0492	0.0002	-0.0009	0.0369	0.0253
YJW	-0.0391	-0.0358	0.1077	0.0835	0.0129	0.0116	0.0467	0.0358

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	0.3083	0.1576	0.9230	0.2911	0.0156	-0.0072	0.9642	0.3599
HKP	2.2080	0.0252	31.5952	2.5406	-0.8822	-0.0768	10.4137	1.3964
Buckley James	1.9799	0.6611	5.5605	0.6679	-1.9703	-0.9228	4.3843	0.9757
YJW	-0.1381	-0.1663	0.4222	0.3352	-0.0257	0.0435	0.6394	0.4794
100 obs.								
MD	0.1642	0.0751	0.5633	0.1605	-0.0231	-0.0055	0.6197	0.2626
HKP	4.8267	-0.0131	86.1774	5.0571	-0.8983	-0.0609	14.1936	1.2946
Buckley James	2.5409	0.8277	8.3142	0.8277	-2.4717	-1.0941	7.1011	1.0941
YJW	-0.1266	-0.1028	0.3794	0.2621	0.0354	0.0556	0.5035	0.3685
200 obs.								
MD	0.0619	0.0456	0.2194	0.1245	-0.0025	0.0088	0.2847	0.1673
HKP	0.0665	0.0133	0.4674	0.2301	-0.0520	0.0012	0.4418	0.3236
Buckley James	4.6481	1.2122	34.5088	1.2122	-3.8331	-1.4037	22.0628	1.4037
YJW	-0.0788	-0.0676	0.2301	0.1713	0.0352	0.0285	0.3415	0.2543
400 obs.								
MD	0.0201	0.0072	0.1340	0.0808	-0.0134	-0.0165	0.2055	0.1378
HKP	-0.0038	-0.0206	0.1888	0.1431	0.0180	0.0158	0.2979	0.2347
Buckley James	4.9056	1.5762	21.8564	1.5762	-4.2223	-1.7588	16.1604	1.7588
YJW	-0.0454	-0.0493	0.1714	0.1323	0.0282	0.0294	0.2697	0.2017

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	0.2318	0.1273	0.6043	0.2937	0.2704	0.0679	1.0035	0.3673
HKP	0.1195	0.0230	0.6267	0.3883	-0.2123	-0.1960	0.6009	0.4385
Buckley James	-0.0390	-0.0477	0.2693	0.1737	-0.0021	0.0007	0.2846	0.1394
YJW	0.8236	0.6180	1.3345	0.8776	-2.2428	-1.9628	2.7286	2.2450
100 obs.								
MD	0.1151	0.0609	0.3790	0.1681	0.1739	0.0960	0.5459	0.2419
HKP	0.1042	0.0457	0.3656	0.2526	-0.1984	-0.1837	0.3914	0.3039
Buckley James	-0.0355	-0.0459	0.1779	0.1312	0.0050	-0.0004	0.1779	0.0787
YJW	0.7568	0.7295	2.5556	0.9934	-2.0231	-1.9468	3.1495	2.2189
200 obs.								
MD	0.0561	0.0342	0.2189	0.1207	0.1246	0.0740	0.3580	0.1947
HKP	0.0799	0.0602	0.2199	0.1700	-0.1817	-0.1778	0.2753	0.2247
Buckley James	-0.0187	-0.0265	0.1216	0.0788	0.0057	0.0032	0.1071	0.0420
YJW	0.8143	0.7990	0.9926	0.8206	-1.9930	-1.9612	2.0831	1.9930
400 obs.								
MD	0.0314	0.0215	0.1579	0.0998	0.0642	0.0391	0.2268	0.1114
HKP	0.0625	0.0590	0.1525	0.1202	-0.1811	-0.1859	0.2368	0.2013
Buckley James	-0.0061	-0.0107	0.0838	0.0581	0.0016	0.0013	0.0620	0.0234
YJW	0.9022	0.8839	0.9728	0.9029	-2.0317	-1.9860	2.0777	2.0317

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	0.3221	0.1483	0.8822	0.3214	0.4476	0.1939	1.6561	0.4751
HKP	17.3471	0.4300	153.2266	17.4901	-4.7159	-1.0175	27.6868	4.8081
Buckley James	1.7253	0.3113	5.6605	0.3716	-2.3871	-0.9728	5.5402	0.9790
YJW	0.9745	0.5735	2.5513	1.0331	-2.7026	-2.1897	4.2208	2.7121
100 obs.								
MD	0.1452	0.0492	0.4698	0.2012	0.2691	0.1049	0.7333	0.2933
HKP	20.1083	0.3676	163.0567	20.1872	-4.4281	-0.8931	24.1675	4.4517
Buckley James	2.3266	0.4856	7.4144	0.4856	-3.3150	-1.1578	9.8721	1.1578
YJW	0.7854	0.6658	3.6963	1.0380	-2.3131	-2.1072	4.7598	2.5239
200 obs.								
MD	0.0674	0.0442	0.2425	0.1382	0.1570	0.0872	0.4477	0.2430
HKP	13.3405	0.3622	101.3273	13.3764	-3.2912	-0.9041	16.3618	3.2944
Buckley James	4.6914	0.6757	28.3437	0.6757	-5.4411	-1.3982	32.3954	1.3982
YJW	0.9670	0.7625	3.5172	0.9707	-2.4190	-2.1862	4.7660	2.4190
400 obs.								
MD	0.0294	0.0142	0.1827	0.1045	0.1253	0.0862	0.3388	0.1779
HKP	83.9411	0.2946	663.8925	83.9501	-11.1658	-0.8572	76.6133	11.1658
Buckley James	6.4436	1.3839	30.7388	1.3839	-5.8856	-2.1991	19.2469	2.1991
YJW	0.8527	0.8572	0.9087	0.8534	-2.2082	-2.1888	2.2396	2.2082

TABLE V

Simulation Results for Censored Regression Estimators

End. Censoring, Homosked. Errors

	α				β			
	Mean Bias	Med. Bias	RMSE	MAD	Mean Bias	Med. Bias	RMSE	MAD
50 obs.								
MD	0.3221	0.1483	0.8822	0.3214	0.4476	0.1939	1.6561	0.4751
HKP	-0.2837	-0.3124	1.0604	0.6562	-0.1277	-0.1108	0.3055	0.1899
Buckley James	-0.6154	-0.6246	0.9891	0.7096	-0.2267	-0.2375	0.3080	0.2419
YJW	1.3391	1.1141	2.2510	1.1936	1.6087	1.6710	1.9920	1.6710
100 obs.								
MD	0.1452	0.0492	0.4698	0.2012	0.2691	0.1049	0.7333	0.2933
HKP	-0.4087	-0.3632	0.8266	0.5193	-0.1684	-0.1586	0.2587	0.1689
Buckley James	-0.6412	-0.6571	0.8280	0.6647	-0.2331	-0.2285	0.2723	0.2316
YJW	1.3479	1.0610	2.2394	1.1183	1.5984	1.6145	1.9851	1.6145
200 obs.								
MD	0.0674	0.0442	0.2425	0.1382	0.1570	0.0872	0.4477	0.2430
HKP	-0.4616	-0.4471	0.6359	0.4801	-0.1907	-0.1877	0.2241	0.1877
Buckley James	-0.6280	-0.6255	0.7268	0.6255	-0.2310	-0.2346	0.2505	0.2346
YJW	10.5304	1.3098	175.7172	1.3183	3.4390	1.1062	39.9520	1.1190
400 obs.								
MD	0.0294	0.0142	0.1827	0.1045	0.1253	0.0862	0.3388	0.1779
HKP	-0.4656	-0.4236	0.5817	0.4236	-0.1906	-0.1809	0.2125	0.1809
Buckley James	-0.6304	-0.6229	0.6765	0.6229	-0.2310	-0.2321	0.2405	0.2321
YJW	4.2454	1.4062	26.2483	1.4534	1.8082	-0.7231	6.7396	0.7727

TABLE VI
Empirical Study of Drug Relapse Data

	Weibull	Log Log.	LAD	YJW	HKP	Buck. James	Min. Dis.
INT	4.8350	3.8752	3.8368	3.8620	3.4108	3.7295	3.6130
	(0.1187)	(0.1110)	(0.1070)	(0.1457)	(0.3533)	(0.1310)	(0.1411)
SITE	-0.4866	-0.5254	-0.4926	-0.4952	-0.3547	-0.5559	-0.2578
	(0.1040)	(0.0938)	(0.0959)	(0.0944)	(0.1352)	(0.1016)	(0.1176)
IV	-0.3673	-0.1835	-0.1769	-0.1683	-0.1192	-0.1277	-0.0919
	(0.0985)	(0.0862)	(0.0876)	(0.0896)	(0.0943)	(0.0908)	(0.1007)
NDT	-0.0243	-0.0209	-0.0119	-0.0122	-0.0164	-0.0186	-0.0393
	(0.0078)	(0.007)	(0.0075)	(0.0062)	(0.0065)	(0.0071)	(0.0065)
RACE	0.2964	0.3288	0.3393	0.3292	0.4411	0.3413	0.4050
	(0.1073)	(0.0952)	(0.0948)	(0.1131)	(0.1147)	(0.0990)	(0.0987)
TREAT	0.4215	0.6114	0.6243	0.6075	0.7605	0.6120	0.7952
	(0.0905)	(0.0839)	(0.0820)	(0.0919)	(0.1451)	(0.0847)	(0.1221)
FRAC	1.1543	1.468	1.2488	1.2357	1.6790	1.5732	1.5412
	(0.0990)	(0.0839)	(0.0798)	(0.0938)	(0.3038)	(0.0869)	(0.1123)

TABLE VII
Empirical Study of Selective Compliance using Drug Relapse Data

	OLS	2SLS	MD	MDIV
INT	4.3726	4.3090	4.0869	4.5012
	(0.0807)	(0.2141)	(0.1827)	(0.1047)
IV	-0.1783	-0.1863	-0.1219	-0.1457
	(0.0777)	(0.0812)	(0.0668)	(0.0551)
RACE	0.2840	0.2651	0.3087	0.3391
	(0.0836)	(0.0879)	(0.0461)	(0.0645)
NDT	-0.0171	-0.0177	-0.0313	-0.0285
	(0.0067)	(0.0071)	(0.0171)	(0.0082)
SITE	-0.4187	-0.2354	-0.3123	-0.2382
	(0.0833)	(0.1230)	(0.0932)	(0.1183)
LOS	0.0086	0.0050	0.0114	0.0067
	(0.0005)	(0.0018)	(0.0013)	(0.0009)