Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities*

Shakeeb Khan[a†] and **Elie Tamer**[b]

[a] *Department of Economics, Duke University , Durham, NC 27708.*

[b] *Department of Economics, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208.*

This Version: August 2008

Abstract

Under a 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 to the parameter vector of interest, and has an asymptotically normal distribution. A small scale simulation study and an application using drug relapse data demonstrate satisfactory finite sample performance.

JEL Classification: C13; C31
Keywords: Conditional Moment Inequality models, quantile minimum distance, covariate dependent censoring, heteroskedasticity, endogeneity.

*We thank K. Hirano, B. Honoré, J. Powell, I. Prucha and A. Rosen as well as seminar participants at many institutions, the 2005 World Congress meetings of the Econometric Society and the 2006 Econometrics in Rio conference for many helpful comments. We also thank an editor and the referees for their comments. Both authors also gratefully acknowledge support from the National Science Foundation. Any errors are our own.

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

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

$$y_i = x_i'\beta_0 + \epsilon_i \tag{1.1}$$

where the dependent variable y_i 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_i, ϵ_i, c_i) where c_i 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[2]. 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[3] 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). Some of the assumptions that these papers use 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.

We aim to recast the model above in a framework that delivers point identification yet

[1]An alternative class of models used in duration analysis is the (Mixed) Proportional Hazards Model. See Khan and Tamer(2007) and the references therein for recent developments in those models.

[2]Recently. Honoré and Lleras-Muney(2006) derive bounds for a *competing risk* model, which is related to the randomly censored regression model above.

[3]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 Honoré (1990) and Abbring and van den Berg(2003). In many settings it is difficult to justify any modelling of the censoring process, and inconsistent results generally arise when this process is misspecified.

weakens many of the assumptions used in the existing literature. In particular, we wish to construct an estimation procedure that 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

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

where $m(.)$ is a known function (up to β_0), (z_i, x_i) 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), Pakes, Porter, Ho, and Ishii (2005), Andrews and Soares (2007), and Rosen (2006). 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 taking 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[4]) and not consistency. Other papers that consider moment inequalities are Andrews and Soares (2007), Canay (2007) and references therein.

With moment inequality models, transforming conditional moment inequalities into unconditional ones is more involved since this might entail a loss of identification. This comes from the fact that, $E[m(z_i;\beta_0)|x_i] \leq 0 \ \forall x_i \ a.e. \Rightarrow E[w(x_i)m(z_i;\beta_0)] \leq 0$ for an appropriate nonnegative weight function, but generally $E[w(x_i)m(z_i;\beta_0)] \leq 0 \not\Rightarrow E[m(z_i;\beta_0)|x_i] \leq 0 \ \ x_i \, a.e.$ More importantly, in models that are partially identified, the choice of the "instrument function" $w(.)$ affects identification in that different weight function can lead to different identification regions (as opposed to methods with equality constraints like GMM, where choice of w generally impacts efficiency, but not point identification). Ideally, one would want a weight function that leads to the identified (or sharp) set.

One approach to handling moment inequalities while preserving the information content is in Manski and Tamer (2002), who estimated the conditional moment condition nonparametrically in a first step. This is not practically attractive since x_i might be multidimensional and nonparametric estimation requires choosing smoothing parameters and trimming procedures. On the other hand, Pakes, Porter, Ho, and Ishii (2005) use an unconditional version of the above moment inequality as the basis for inference. Other papers on moment inequality takes a set of *unconditional* moment conditions as its starting point. In this

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

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 more "localized" weight functions can be used for conducting inference on the parameters of interest. In 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[5] 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), and 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 $\tau_0(x_i, \beta)$ 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.

[5]Implicitly, this assumption requires that this median is unique and hence that the conditional distribution of $\epsilon|x$ is strictly increasing around zero.

Lemma 2.1 *At the truth ($\beta = \beta_0$), and on S_X, the support*[6] *of x, the following holds:*

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

proof: First, we have for τ_1:

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

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

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

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* Θ_I that contains observationally equivalent parameters (including β_0). 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_i and x_i and also c_i and ϵ_i and so allows for general types of censoring that can be random and can be endogenous, or dependent on y_i (conditional on x_i)[7]. The statistical setup is exactly one of competing risks where the risks are allowed to

[6]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, though as pointed out to us by a referee, in the discrete regressor case, both finite sample performance and limiting distribution theory will be affected if strict inequalities are used in our objective function.

be dependent, or a Roy model setup. It is well known that a general dependent competing risk model is not (point) identified- see, e.g. **?**. So, additional assumptions are needed to shrink the set to a point.

It is possible to obtain 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 ($c_i = 0$), the set Θ_I shrinks to a point if the set of x_i's for which $x_i'\beta_0$ is negative has positive mass and $x_i x_i'$ is full rank on this set. This intuition was used to derive similar point identification conditions in Manski and Tamer (2002) and was also essentially shown by Powell (1984). We extend this intuition to our model setup where the censoring is allowed to be random and correlated with the regressors and error terms.

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

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

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

So, a sufficient condition for point identification is for (2.3) to be positive for some x_i, *or* for (2.4) to be negative. One needs to find such x_i'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.

[7] By construction, we have $y_1 \leq y \leq y_2$ where $y_1 = y \times d + (1-d) \times (-\infty)$ and $y_2 = y \times d + (1-d) \times c$. Hence, by A.1, we have $med(y_1|x) \leq x'\beta_0 \leq med(y_2|x)$. Now, define the set $\Theta = \{b : Med(y_1|x) \leq x'b \leq Med(y_2|x)\}$. This set is the *sharp* set, since for any $b \in \Theta$, there exists a $y \in [y_1, y_2]$ such that $Med(y|x) = x'b$ (of course this presumes that Θ is nonempty, since β_0 belongs to it.). These bounds on medians contain exactly the same information as the inequalities in (2.2) above and hence the identified set constructed using (2.2) is sharp.

A2 is a sufficient condition for identification and it is related to the notion of "regular identification" since this condition is necessary (but not sufficient) for root n estimation of β_0. For example, when (y_i, x_i, c_i) are jointly normal, we see that $\mathcal{C}$ has measure 0, but it can be shown that β_0 here is point identified. 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). To estimate β_0 at the regular rate, we strenghten A2 in section 3 below and require that the set $\mathcal{C}$ has positive measure.

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 from values of the regressors where censoring cannot occur.

Remark 2.1 *Assumptions A.1 and A.2 allows for correlation between the censoring variable, the regressors and the latent error terms. On the other hand, statistical independence between ϵ_i and (x_i, c_i) is imposed in Buckley and James (1979) and Yang (1999), Portnoy (2003). Independence between c_i and (x_i, ϵ_i) was imposed in Ying, Jung, and Wei (1995), and Honoré, Khan, and Powell (2002)* [8]*. 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*[9] *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 $c_i \perp \epsilon_i | x_i$. The cost of these generalities is in terms of strong point identification conditions. A.2 requires support restrictions on the censoring variables relative to the index. Naturally, under only assumption A.1, the model identifies a set of parameters that includes β_0.*

[8]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 use 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, it is not clear how to allow endogenous censoring in their setups.

[9]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 (1990).

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* $\beta \neq \beta_0$

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

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: Recalling $\tau_1(x_i,\beta) = P(\epsilon \geq x_i'\delta; c_i \geq x_i'\beta | x_i) - \frac{1}{2}$ and $\tau_0(x_i,\beta) = \frac{1}{2} - P(\epsilon \leq x_i'\delta; \epsilon_i \leq c_i - x_i'\beta_0 | x_i)$ consider first the case when $x_i'\delta < 0$. Then for x satisfying condition A2 for β_0,

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

where (1) follows from x belonging to the special set postulated in A2, and (2) follows from 0 being the *unique* conditional median. Now for the case when $x_i'\delta > 0$. Consider x_i satisfying the condition in A2. Hence for that x_i, we have

$$x_i'\delta > 0 \implies (2.4) \quad < 0$$

where the inequality follows from the fact that $x_i'\delta$ is positive and $c_i \geq x_i\beta_0$. ■

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 way to obtain weaker sufficient* *point identification conditions is require independence between c_i and ϵ_i as in Honoré, Khan, and Powell (2002). In some empirical settings, it is plausible to maintain independence between c_i and ϵ_i. 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 or even conditional independence can be suspect, since censoring can be affected by unobservables that also affect outcomes and so maintaining independence would lead to inconsistent estimates (that might not fall in the identified set).*

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 reweighting 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). Reweighting 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.*

We conclude this section by pointing out that our identification results readily extend to the doubly censored regression model. For this model the econometrician observes the doubly censored sample, which we can express as the pair (v_i, d_i) where

$$d_i = I[c_{1i} < x_i'\beta_0 + \epsilon_i \le c_{2i}] + 2 \cdot I[x_i'\beta_0 + \epsilon_i \le c_{1i}] + 3 \cdot I[c_{2i} < x_i'\beta_0 + \epsilon_i]$$

$$v_i = I[d_i = 1] \cdot (x_i'\beta_0 + \epsilon_i) + I[d_i = 2]c_{1i} + I[d_i = 3]c_{2i}$$

where c_{1i}, c_{2i} denote left and right censoring variables, whose distributions may depend on (x_i, ϵ_i) and who satisfy $P(c_{1i} < c_{2i}) = 1$.

Note we have the following two conditional moment inequalities:

$$E[I[d_i \neq 2]I[v_i \ge x_i'\beta_0]|x_i] - \frac{1}{2} \le 0 \tag{2.6}$$

$$\frac{1}{2} - E[I[d_i \neq 3]I[v_i \le x_i'\beta_0]|x_i] \ge 0 \tag{2.7}$$

from which we can redefine the functions $\tau_0(x_i, \beta), \tau_1(x_i, \beta)$ accordingly. With these functions we can set identify the parameter vector under A1. To get point identification in the double censoring case we would change A2 to:

A2 The subset

$$\mathcal{C}_D = \{x_i \in S_X : \Pr(c_{1i} \leq x_i'\beta_0 \leq c_{2i}|x_i) = 1\}$$

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

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[10] 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 values x_i, one might be tempted to use similar moments that are unconditional[11] on x_i. 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_i, 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

[10] In this and subsequent sections we will focus on the singly censored model. We note from our remark in the previous section that we can easily modify our objective function to conduct inference in the doubly censored model as well.

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

defined as follows:

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

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

where the above inequality $t_1 \le x_i$ is to be taken componentwise. From fundamental properties of expectations,[12] conditional moment conditions can be related to unconditional moment 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_i. 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 an inequality moment condition such as

$$E[m(y_i;\beta_0)|x_i] \le 0 \quad \text{for all} \quad x_i$$

into an informationally equivalent unconditional model as

$$H(\beta_0, x_1, x_2) = E\left[m(y_i;\beta_0)1[x_1 \le x_i \le x_2]\right] \le 0 \quad \text{for all} \quad (x_1, x_2)$$

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:

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

and the following additional assumptions: first 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.

This is a sufficient condition for *regular identification* of β_0. It is slightly stronger than assumption A2 since it requires that $\mathcal{C}$ has positive measure. This for example rules the

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

case in which (y, c, x) are jointly normal. Our estimation approach is still valid (provides consistent estimates for the identified features of the model) even when A2' (or A2) do not hold. Next, we impose the following smoothness conditions on the error distribution, regressors and censoring variable:

A3 The regressors x_i are assumed to have compact support, denoted by $\mathcal{X}$, and are continuously distributed[13] with a density function $f_X(\cdot)$ that is bounded away from 0 on $\mathcal{X}$.

A4 The error terms ϵ_i are absolutely continuously distributed with conditional density function $f_\epsilon(\epsilon \mid x)$ 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.

The main identification result is stated in the following lemma:

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

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

In addition, if A2',A3,A4 hold, then $Q(\beta)$ is minimized uniquely at $\beta = \beta_0$.

Proof: To prove the point identification result under A1,A2',A3-A4 we first show that

$$Q(\beta_0) = 0 \tag{3.4}$$

To see why, note this follows directly from the previous lemmas which established that $\tau_1(x_i, \beta_0)I[\tau_1(x_i, \beta_0) \geq 0] = \tau_0(x_i, \beta_0)I[\tau_0(x_i, \beta_0) \leq 0] = 0$ for all values of x_i on its support. Similarly, as established in that lemma, for each $\beta \neq \beta_0$, there exists a regressor value x^*, such that, under A4, for all x in a sufficiently small neighborhood of x^*, $\max(\tau_1(x, \beta)I[\tau_1(x, \beta) \geq 0], -\tau_0(x, \beta)I[\tau_0(x, \beta) \leq 0]) > 0$. Let $\mathcal{X}_\delta$ denote this neighborhood of x^*, which under A3 can be chosen to be contained in $\mathcal{X}$. Since x_i has the same support across observations, for each $x^{**} \in \mathcal{X}_\delta$, we can find values of x_j, x_k s.t. $x_j \leq x^{**} \leq x_k$, establishing that $Q(\beta) > 0$. ■

[13] This is assumption is made for convenience- discrete regressors with finite supports can be allowed for as well.

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-process 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 are 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 discuss a way to reduce this computational burden at the end of this section.

We next turn attention to the asymptotic properties of the estimator in (4.3).

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 $(d_i, v_i, x_i')'$, $i = 1, \ldots, n$.

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

Theorem 4.1 *Under Assumptions A1,A2',A3,A4, and C1-C2,*

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

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

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

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 4.2 *Under Assumptions A1,A2',A3-A4, C1-C2, and D1-D2*

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

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). This would result in an estimator that minimizes a second order U-process that is much simpler computationally, though less efficient[14]. For the problem at hand, the split sample version of our proposed estimator would minimize an objective function of the form:

$$\hat{\beta}_{SS} = \arg\min_{\beta \in \mathcal{B}} \hat{Q}_{SSn}(\beta) \tag{4.10}$$

$$= \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^{n} \Big\{ \hat{H}_1(\beta, x_j, x_{n-j+1}) I[\hat{H}_1(\beta, x_j, x_{n-j+1}) \geq 0] \tag{4.11}$$

$$- \hat{H}_0(\beta, x_j, x_{n-j+1}) I[\hat{H}_0(\beta, x_j, x_{n-j+1}) \leq 0] \Big\}$$

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

[14]An alternative procedure would be a "resampling" based estimator, where one resamples $O(n)$ regressor values (with replacement) in the construction of the functions $\hat{H}_1(\beta, \cdot, \cdot), \hat{H}_0(\beta, \cdot, \cdot)$. We leave the asymptotic properties of this procedure for future work.

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), **?**, 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:

IV1 $\text{median}(\epsilon_i|z_i) = 0$.

IV2 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 $\delta \neq 0$, at least one 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 IV1 is analogous to the usual condition of the instruments being uncorrelated with the error terms.*

2. *Condition IV2 details the relationship between the instruments and the regressors. It is most easily satisfied when the exogenous variable(s) have support that are relatively large when compared to the support of the endogenous variable(s).*[15] *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:

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

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. The theorem below establishes the limiting distribution theory for this estimator. Its proof is omitted since it follows from virtually identical arguments used to prove the previous theorem.

[15] 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.

Theorem 5.1

$$\sqrt{n}(\hat{\beta}_{IV} - \beta_0) \Rightarrow N(0, V_{IV}^{-1}\Omega_{IV}V_{IV}^{-1}) \tag{5.6}$$

where we define V_{IV} and Ω_{IV} are as follows. Let

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

Adopting the notation $I_{ijk} = I[z_j \leq z_i \leq z_k]$, define the function

$$G(z_j, z_k) = I[[z_j, z_k] \subseteq \mathcal{C}] \int f_\epsilon(0|z_i)E[x_i|z_i]I_{ijk}f_Z(z_i)dz_i \tag{5.7}$$

where $f_Z(\cdot)$ denotes the instrument density function. The Hessian matrix is

$$V_{IV} = 2E[G(z_j, z_k)G(z_j, z_k)'] \tag{5.8}$$

Next we define the outer score term Ω_{IV}. Let

$$\delta_{0i} = E[G(z_j, z_k)I_{ijk}|z_i](I[v_i \geq x_i'\beta_0] - d_iI[v_i \leq x_i'\beta_0]) \tag{5.9}$$

so we can define the outer score term Ω as

$$\Omega_{IV} = E[\delta_{0i}\delta_{0i}'] \tag{5.10}$$

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 regressors 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 $c_i = -x_{1i}^2 - x_{2i}$.[16]

We assumed the error distribution of ϵ_i was standard normal. In addition, we simulated designs with heteroskedastic errors as well: $\epsilon_i = \sigma(x_{2i}) \cdot \eta_i$, with η_i having a standard normal distribution and $\sigma(x_{2i}) = \exp(0.5 * x_{2i})$. We also simulated a design with endogenous censoring where there was one regressor, distributed standard normal, the error was homoskedastic, and the censoring variable was related to η_i as $c_i = (4 + u_i) * \eta_i^2$ where u_i was an independent random variable distributed uniformly on the unit interval.

For these designs, the overall censoring probabilities vary between 40% and 50%. For each replication of the model, the following estimators were calculated[17]:

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.[18] The randomly-censored least absolute deviations estimator (HKP) was computed using the iterative Barrodale-Roberts algorithm described by Buchinsky(1995)[19]; 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 5 tables corresponded to

[16]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.

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

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

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

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,.

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 performs 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 does 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 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 estimator 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 VI 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[20]. 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,

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

SITE, and LOS. Table VII 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 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., and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities using Generalize Moment Selection," Working Paper, Yale University.

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 (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.

Canay, I. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Wisconsin Working Paper.

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.

HECKMAN, J., AND B. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

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.

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2005): "Moment Inequalities and Their Applications," Working Paper.

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.

ROSEN, A. (2006): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," UCL Working Paper.

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 follows from Assumptions C1 and continuity of the objection function follows from the smoothness conditions in A3,A4 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

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

where

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

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

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

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

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}(\hat{H}_1(\beta,x_j,x_k)-H_1(\beta,x_j,x_k))I[H_1(\beta,x_j,x_k)\geq 0]\right|=o_p(1) \tag{A.4}$$

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

$$\frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq k}((I[v_i\geq x_i'\beta]-\frac{1}{2})I[x_j\leq x_i\leq x_k]-H_1(\beta,x_j,x_k))I[H_1(\beta,x_j,x_k)\geq 0]$$

is a mean 0 third order U-process. Consequently, by the i.i.d assumption in C2, and applying Corollary 7 in Sherman(1994a), this term is uniformly $o_p(1)$. It remains to show that replacing $\hat{H}_1(\beta,x_j,x_k)$ with $H_1(\beta,x_j,x_k)$ inside the indicator function yields an asymptotically uniformly negligible remainder term. Specifically, we will show that

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

Noting that $\hat{H}_1(\beta, x_j, x_k)$ is uniformly bounded in β, x_j, x_k, it will suffice to show that

$$\sup_{\beta\in\mathcal{B}}\left|\frac{1}{n(n-1)}\sum_{i\neq j}|I[\hat{H}_1(\beta,x_j,x_k)\geq 0]-I[H_1(\beta,x_j,x_k)\geq 0]|\right|=o_p(1) \tag{A.6}$$

To do so, we will, w.l.o.g., only show that $\frac{1}{n(n-1)}\sum_{i\neq j} I[\hat{H}_1(\beta,x_j,x_k)\geq 0]I[H_1(\beta,x_j,x_k)<0]$ is $o_p(1)$ uniformly in β. To do so owe will add and subtract the expectation of the term inside the double summation, conditional on x_j, x_k. We note that by subtracting this conditional expectation from the double summation, we again have a mean 0 U-process, to which we can again apply Corollary 7 in Sherman(1994a) to conclude this is uniformly $o_p(1)$.

It thus remains to show that

$$\frac{1}{n(n-1)}\sum_{i\neq j}E[I[\hat{H}_1(\beta,x_j,x_k)\geq 0]I[H_1(\beta,x_j,x_k)<0]|x_j,x_k]$$

is $o_p(1)$ uniformly in β.

We condition on values of x_j, x_k and decompose $I[H(\beta,x_j,x_k)<0]$ as

$$I[H_1(\beta,x_j,x_k)<0]=I[H_1(\beta,x_j,x_k)\leq-\delta_n]+I[H_1(\beta,x_j,x_k)\in(-\delta_n,0)] \tag{A.7}$$

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

$$I[\hat{H}_1(\beta,x_j,x_k)\geq 0]I[H_1(\beta,x_j,x_k)\leq-\delta_n] \tag{A.8}$$

The probability (i.e. conditional expectation) that the above product of indicators is positive, recalling that we are conditioning on x_j, x_k, is less than or equal to the probability that

$$\sum_{i=1}^{n}(I[v_i\geq x_i'\beta]-\frac{1}{2})I_{ijk}-H_1(\beta,x_j,x_k)>-nH_1(\beta,x_j,x_k) \tag{A.9}$$

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

$$\sum_{i=1}^{n}(I[v_i\geq x_i'\beta]-\frac{1}{2})I_{ijk}-H_1(\beta,x_j,x_k)>n\delta_n \tag{A.10}$$

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

$$\frac{1}{n(n-1)} \sum_{i \neq j} I[\hat{H}_1(\beta, x_j, x_k) \geq 0] I[H_1(\beta, x_j, x_k) \in (-\delta_n, 0)] \tag{A.11}$$

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

$$\frac{1}{n(n-1)} \sum_{i \neq j} I[H_1(\beta, x_j, x_k) \in (-\delta_n, 0)] \tag{A.12}$$

to 0. Subtracting $E[I[H_1(\beta, x_j, x_k) \in (-\delta_n, 0)]]$ 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 $E[I[H(\beta, x_j, x_k) \in (-\delta_n, 0)]]$ is uniformly $o_p(1)$ by applying the dominated convergence theorem. Combining all our results we conclude that (A.3) is $o_p(1)$.

Next, we will establish that

$$\sup_{\beta \in \mathcal{B}} \left| \frac{1}{n(n-1)} \sum_{i \neq j} H_1(\beta, x_j, x_k) I[H_1(\beta, x_j, x_k) \geq 0] - Q_1(\beta) \right| = o_p(1) \tag{A.13}$$

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.13) 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. $\hat{\beta} - \beta_0 = O_p(\delta_n)$

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 $O_p(\delta_n)$ neighborhoods of β_0

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

which suffices for $\hat{\beta} - \beta_0 = O_p(\max(\epsilon_n^{1/2}, n^{-1/2}))$. With this theorem we will establish root-n consistency in two stages. Having already established consistency we will first set $\delta_n = o(1)$, $\epsilon_n = O(n^{-1/2})$, and show the above three conditions are satisfied. This will imply that the estimator is fourth-root consistent. We will then show the conditions of the theorem are satisfied for $\delta_n = O(n^{-1/4})$ and $\epsilon_n = O(n^{-1})$, establishing root-n consistency.

Regarding the showing the first result of fourth root consistency, in what follows throughout the rest of our proof, it will prove convenient to subtract the following term, which does not depend on β, from our objective function:

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

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 the expectation of this term conditional on x_j, x_k is 0.

To show the second of the three conditions, 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 Assumption **A3**,**A4**,**D2** imply differentiability after taking expectations. 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 **A3** and **D2**, 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, $I[\hat{H}_1(\beta, x_j, x_k) \geq 0], I[\hat{H}_0(\beta, x_j, x_k) \leq 0]$, with the functions $I[H_1(\beta, x_j, x_k) \geq 0], I[H_0(\beta, x_j, x_k) \leq 0]$ respectively, and derive the corresponding representation. (We will deal with the resulting remainder term from this replacement shortly) We expand the terms $\hat{H}_1(\beta, x_j, x_k)$ and $\hat{H}_0(\beta, x_j, x_k)$, first, exclusively dealing with the first expansion since the second is similar. This results in the third order U-process:

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

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:

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

where $P_n Pm(z_j, ., .)$ is equal to $\frac{1}{n}\sum_j P_{k,l} m(z_j, z_k, z_l)$ where $P_{k,l}$ denotes expectation with respect to second and third arguments, and $U_n h$ 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

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

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 $I[H_1(\beta_0, x_j, x_k) \geq 0] \cdot I_{ljk} = 1$ implies that $[x_j, x_k] \subset \mathcal{C}$, as is x_l; we next evaluate

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

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

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

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 $H_1(\beta_0, x_j, x_k) < 0$ in the latter case and the expectation term in (B.10) is 0. 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:

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

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

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

Next we evaluate

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

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 it 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 $(I[v_l \ge x_l'\beta] - \frac{1}{2}) E[I[H_1(\beta, x_j, x_k) \ge 0]]$ 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:

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

as does the expectation conditional on the second argument, indexed by k. So, the $O_p(||\beta - \beta_0||/\sqrt{n})$ term is (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

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

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 $I[\hat{H}_1(\beta,x_j,x_k)\geq 0]$, $I[\hat{H}_0(\beta,x_j,x_k)\geq 0]$ with $I[H_1(\beta,x_j,x_k)\geq 0]$, $I[H_0(\beta,x_j,x_k)\geq 0]$ respectively, is $O_p(n^{-1/2})$.

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

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

as well as the second "half" involving $H_0(\beta,x_j,x_k)$. To establish the negligibility of (B.17), we will first derive a rate of convergence for

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

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

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

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 $\hat{H}_1(\beta,x_j,x_k)$ and $I[H_1(\beta,x_j,x_k)\geq 0]$ corresponds to a negligible term, uniformly in β in $o_p(1)$ neighborhoods of β_0. First, dealing with

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

Noting that since $\hat{H}_1(\beta,x_j,x_k)$ is bounded uniformly in β,x_j,x_k it will suffice to show the negligibility of

$$\frac{1}{n(n-1)}\sum_{j\neq k}I[\hat{H}_1(\beta,x_j,x_k)<-\delta_n]I[H_1(\beta,x_j,x_k)\geq 0] \tag{B.21}$$

As we did before, we will add and subtract the conditional expectation of the term inside the double summation. First dealing with the subtraction, we now have a centered U-process, where the variance of the term in the double summation vanishes at the rate of $o(1)$. Thus, by Theorem 3 in Sherman(1994b) the centered process is uniformly $o_p(n^{-1})$ Now, dealing with the addition of the conditional expectation:

$$\frac{1}{n(n-1)}\sum_{i\neq j} E[I[\hat{H}_1(\beta,x_j,x_k)<-\delta_n]I[H_1(\beta,x_j,x_k)\geq 0]|x_j,x_k] \tag{B.22}$$

we can follow the same arguments as in our consistency proof to conclude that it is uniformly $o_p(n^{-1})$.

It thus remains to show the negligibility of

$$\frac{1}{n(n-1)}\sum_{i\neq j} I[\hat{H}_1(\beta,x_j,x_k)\in[-\delta_n,0)]\hat{H}_1(\beta,x_j,x_k)I[H_1(\beta,x_j,x_k)\geq 0] \tag{B.23}$$

Note the absolute value of the above summation is bounded above by

$$\delta_n\frac{1}{n(n-1)}\sum_{i\neq j} I[\hat{H}_1(\beta,x_j,x_k)\in[-\delta_n,0)]I[H_1(\beta,x_j,x_k)\geq 0] \tag{B.24}$$

In the above double summation, we add the sum of indicators $I[[x_j,x_k]\subset\mathcal{C}]+I[[x_j,x_k]\not\subset\mathcal{C}]$

First we evaluate the rate for

$$\delta_n\frac{1}{n(n-1)}\sum_{i\neq j} I[\hat{H}_1(\beta,x_j,x_k)\in[-\delta_n,0)]I[H_1(\beta,x_j,x_k)\geq 0]I[[x_j,x_k]\subset\mathcal{C}] \tag{B.25}$$

(Similar arguments can be used for the subset $[x_j,x_k]\not\subset\mathcal{C}$, but we omit the details.) We again add and subtract the conditional (on x_j,x_k) expectation of the term inside the above double summation; the centered process (i.e. after subtracting the conditional expectation) is $o_p(n^{-1})$ uniformly in $o_p(1)$ neighborhoods of β_0 after we multiply by δ_n. Now, regarding the expectation

$$E[I[\hat{H}_1(\beta,x_j,x_k)\in[-\delta_n,0)]I[H_1(\beta,x_j,x_k)\geq 0]I[[x_j,x_k]\subset\mathcal{C}]]$$

we can decompose $I[H_1(\beta,x_j,x_k)\geq 0]$ into $I[H_1(\beta,x_j,x_k)=0]+I[H_1(\beta,x_j,x_k)>0]$. We first deal with the term:

$$E[I[\hat{H}_1(\beta,x_j,x_k)\in[-\delta_n,0)]I[H_1(\beta,x_j,x_k)=0]I[[x_j,x_k]\subset\mathcal{C}]]$$

Note that the indicator $I[H_1(\beta,x_j,x_k)=0]$ is 0 unless $\beta=\beta_0$ as $[x_j,x_k]\subset\mathcal{C}$, and note we have

$$E[I[\hat{H}_1(\beta_0,x_j,x_k)\in[-\delta_n,0)]=o_p(n^{-1})$$

Therefore it only remains to establish a rate for

$$E[I[\hat{H}_1(\beta, x_j, x_k) \in [-\delta_n, 0)]I[H_1(\beta, x_j, x_k) > 0]I[[x_j, x_k] \subset \mathcal{C}]]$$

We will simply expand the term $E[I[H_1(\beta, x_j, x_k) > 0]I[[x_j, x_k] \subset \mathcal{C}]]$ as a function of β around β_0- the lead term is 0, and the remainder term is $O(\|\beta - \beta_0\|)$ which is $o_p(1)$. However this will not suffice to conclude that the estimator is root-n consistent, only that is $o(\sqrt{\delta_n})$. But by a second application of Theorem 1 in Sherman(1994b), this time looking in $o(\sqrt{\delta_n})$ neighborhoods of β_0 we can conclude the estimator is indeed $\sqrt{n}$-consistent.

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,

$$\mathcal{G}_n(\beta) - \mathcal{G}_n(\beta_0) = \frac{1}{2}(\beta - \beta_0)'V(\beta - \beta_0) + \frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n + o_p(\frac{1}{n}) \tag{B.26}$$

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.26). 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 note 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 **A4**, **D2**. Taking a second order expansion of $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$, we obtain

$$\mathcal{G}(\beta) = \mathcal{G}(\beta_0) + \nabla_\beta \mathcal{G}(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta}\mathcal{G}(\beta^*)(\beta - \beta_0) \tag{B.27}$$

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:

$$\nabla_{\beta\beta}\mathcal{G}(\beta^*) = V + o_p(1) \tag{B.28}$$

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

$$E[H_1(\beta, x_j, x_k)I[H_1(\beta, x_j, x_k) \geq 0]] - E[H_0(\beta, x_j, x_k)I[H_0(\beta, x_j, x_k) \leq 0]] \tag{B.29}$$

Note that by definition:

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

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

So, the second derivative is:

$$V = E\left[\nabla_{\beta\beta}H_1I_1 + 2\nabla_\beta H_1\nabla_\beta I_1 + H_1\nabla_{\beta\beta}I_1 - \nabla_{\beta\beta}H_0I_0 - 2\nabla_\beta H_0\nabla_\beta I_0 - H_0\nabla_{\beta\beta}I_0\right] \quad \text{(B.30)}$$

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,

$$E[\nabla_\beta H_1\nabla_\beta I_1 + H_1\nabla_{\beta\beta}I_1] = 0 \quad \text{(B.31)}$$

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

$$\begin{aligned} V &= E\left[\nabla_\beta H_1\nabla_\beta I_1 - \nabla_\beta H_0\nabla_\beta I_0\right] && \text{(B.32)} \\ &= 2E\left[1[[x_j, x_k] \subset \mathcal{C}] \int_{x_j}^{x_k} x f_\epsilon(0|x)dF_x \int_{x_j}^{x_k} x' f_\epsilon(0|x)dF_x\right] && \text{(B.33)} \\ &= 2E[I[[x_j, x_k] \subseteq \mathcal{C}]G(x_j, x_k)G(x_j, x_k)'] && \text{(B.34)} \end{aligned}$$

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 $\frac{1}{2} - d_lI[v_l \leq x_l'\beta]$, 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

$$I[H_1(\beta, x_j, x_k) \geq 0]I[v_l \geq x_l'\beta] - \frac{1}{2} \quad \text{(B.35)}$$

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

$$\frac{1}{n}\sum_{l=1}^{n} E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - \frac{1}{2}) \tag{B.36}$$

plus a negligible remainder term. Consequently, using the same arguments of the 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:

$$\frac{1}{n}\sum_{l=1}^{n} E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - d_l I[v_i \leq x_l'\beta_0])'(\beta - \beta_0) + o_p(n^{-1}) \tag{B.37}$$

which corresponds to

$$\frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n \tag{B.38}$$

$$W_n \Rightarrow N(0, E[\delta_{0l}\delta_{0l}']) \tag{B.39}$$

where

$$\delta_{0l} = E[G(x_j, x_k)I_{ljk}](I[v_l \geq x_l'\beta_0] - d_l I[v_i \leq x_l'\beta_0]) \tag{B.40}$$

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

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \tag{B.41}$$

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)