Minimum Distance Estimation of Randomly Censored Regression Models with Endogeneity

Shakeeb Khan*
University of Rochester

Elie Tamer
Northwestern University

March 2006

Abstract

This paper proposes minimum distance estimation procedures for the slope coefficients and location parameter in randomly censored regression models that are used in duration and competing risk models. The proposed procedure generalizes existing work in terms of weakening the restrictions imposed on the distribution of the error term and the censoring variable. Examples of such generalizations include allowing for conditional heteroskedasticity, covariate dependent censoring, endogenous regressors and endogenous censoring. The estimator is shown to converge at the parametric rate with an asymptotic normal distribution. A small scale simulation study and applications using drug relapse and re-employment bonus data demonstrate satisfactory finite sample performance.

JEL Classification: C14, C25, C13.

Key Words: Accelerate failure time model, covariate dependent censoring, minimum distance, heteroskedasticity, endogenous regressors, endogenous censoring.

*Corresponding author. S. Khan, Department of Economics, University of Rochester, Rochester, NY, 14627, USA. (585)-275-5781, skhan@troi.cc.rochester.edu. We thank B. Honoré and J. Powell as well as seminar participants at University of Arizona, Cornell, Duke, Princeton, Syracuse, University of Virginia, and the 2005 World Congress meetings of the Econometric Society for many helpful comments. Both authors also gratefully acknowledge support from the NSF. Any errors are our own.
1 Introduction

Much of the recent econometrics, statistics, and biostatistics literature has been concerned with distribution-free estimation of the parameter vector $\beta_0$ in the linear regression model

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

where the dependent variable $y$ is subject to censoring that can potentially be random. For example, in the duration literature, this model is known as the accelerated failure time\(^1\) (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.

Duration models have received a great deal of attention in both the applied and theoretical econometrics literature. This is because many time-to-event variables are of interest to researchers conducting empirical studies in labor economics, development economics, public finance and finance. For example, the time-to-event of interest may be the length of a welfare spell, the time between purchases of a particular good, and time intervals between child births, to name a few. (See Van den Berg(2001) for a recent survey on duration models and their applications.)

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 $(x, \epsilon, c)$ where $c$ is the censoring variable. Work in this area includes the papers by Buckley and James (1979), Powell(1984), Koul et al.(1981), Ying et al.(1995), Yang(1999), Honoré et al.(2002) and, more recently Portnoy(2003) and Cosslett(2004), among many others. Unfortunately, each of the estimation methods introduced in the literature impose some assumption which may be considered too strong and not reasonably characterized by the data, and furthermore, the proposed methods will yield inconsistent estimators of the parameters of interest if these assumptions do not hold. Examples of assumptions which may be regarded as too strong are homoskedastic errors, censoring variables that are independent of the regressors, strong support conditions on the censoring variable which rule out fixed censoring, and exogenous regressors.

This paper adds to this important literature in different dimensions. We propose an estimation procedure for the censored regression model which does not require any of these strong conditions. Specifically, it permits conditional heteroskedasticity in the data, permits

---

\(^1\)An alternative class of models used in duration analysis is the (Mixed) Proportional Hazards Model. See Khan and Tamer(2005) and the references therein for recent developments in those models.
the censoring variable to depend on the covariates in an arbitrary way, can be applied to both fixed and randomly censored data. Our proposed minimum distance estimator does not require nonparametric estimation of the censoring variable or error distribution, and consequently does not require the selection of smoothing parameters nor trimming procedures.

The following section describes the censored model studied in this paper in detail, and introduces the proposed minimum distance estimation procedure. It also compare this procedure to others recently proposed in the literature. Section 2 establishes the asymptotic properties for the proposed procedure, specifying sufficient regularity conditions for root-$n$ consistency and asymptotic normality. Section 3 explains how to modify the proposed procedure to an i.v. type estimator with the availability of instruments. Section 4 explores the relative finite sample performance of the estimator in two ways: section 4.1 reports results from a simulation study, and 4.2 applies the estimator to explore a comparison of two courses of treatment for drug abuse. Section 5 concludes by summarizing results and discussing areas for future research. A mathematical appendix provides the details of the proofs of the asymptotic theory results.

2 Model and Minimum Distance Estimation Method

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

\[
\begin{align*}
    v_i &= \min(y_i, c_i) = \min(x'_i\beta_0 + \epsilon_i, c_i) \\
    d_i &= I[x'_i\beta_0 + \epsilon_i < c_i]
\end{align*}
\]  

(2.1)  

(2.2)

where $\beta_0$ denotes the unknown $k$-dimensional parameter vector of interest, $c_i$ denotes the censoring variable that is only observed for censored observations, and $\epsilon_i$ denotes the unobserved error term. $I[\cdot]$ 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 AFT context usually is the log of the survival time. In the censored model, the log-survival time is only partially observed.

The parameter of interest is $\beta_0$. We provide an identification result that will be based
primarily on the following conditions, which characterize the censoring and error term behavior:

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

A2 \( c_i \perp \epsilon_i|x_i \)

A3 The matrix \( E[I[x'_i\beta_0 \leq c_i|x_i] \) is invertible.

The first of the above conditions is the conditional median assumption imposed previously in the literature—e.g. Honoré et al. (2002) and Ying et al. (1995). It permits very general forms of heteroskedasticity, and is weaker than the assumption \( \epsilon_i \perp x_i \) as was imposed in Buckley and James (1979), Yang (1999), Portnoy (2003). The second condition allows the censoring variable to depend on the regressors in an arbitrary way, and is weaker than the condition \( c_i \perp x_i \) imposed in Honoré et al. (2002), Ying et al. (1995)\(^2\). Thus we can see that by permitting both conditional heteroskedasticity and covariate dependent censoring, our assumptions are weaker than existing work on the censored AFT model. Furthermore the third condition imposes weaker support conditions on the censoring variable when compared to existing work in the literature. For example, the estimator in Koul et al. (1981) requires the support of the censoring variable to be sufficiently large, thereby ruling out fixed censoring. The estimator in Ying et al. (1995) requires the censoring variable to exceed the index \( x'_i\beta_0 \) with probability 1, thus often ruling out the fixed censoring case as well.

We now describe our procedure for identifying \( \beta_0 \). Before doing so, we introduce functions of the data we will be using to establish an identification result that estimation will be based on. These functions will be shown to satisfy inequality conditions which will be the basis for our identification strategy. For any possible parameter value \( \beta \), define the functions of \( x_i \)

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

and let

\[
\tau_0(x_i, \beta) = E[(1 - d_i) + d_iI[v_i > x'_i\beta]|x_i] - \frac{1}{2} = \frac{1}{2} - E[d_iI[v_i \leq x'_i\beta]|x_i]
\]

\(^2\)Both of these papers suggest methods to allow for the censoring variable to depend on the covariates, which involve replacing the Kaplan-Meier procedure they require with a conditional Kaplan-Meier. This will require the choice of smoothing parameters to localize the Kaplan-Meier procedure, as well as trimming functions and tail behavior regularity conditions.
The following lemma establishes the inequality condition which will motivate an objective function that our estimator will be based on.

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

\[
g(x_i, \beta) = \tau_1(x_i, \beta)I[\tau_1(x_i, \beta) \geq 0] - \tau_0(x_i, \beta)I[\tau_0(x_i, \beta) \leq 0]
\]

(2.6)

then under Assumptions A1-A3,

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

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

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

**Proof:** First we note

\[
P(v_i \geq x'_i \beta_0 | x_i) = P(\epsilon_i \geq 0 | x_i)P(c_i \geq x'_i \beta_0 | x_i) = \frac{1}{2}P(c_i \geq x'_i \beta_0 | x_i)
\]

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

\[
d_i v_i = d_i (x'_i \beta_0 + \epsilon_i) \leq x'_i \beta_0 + \epsilon_i
\]

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

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

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

\[
\tau_1(x^*, \beta) = P(\epsilon_i \geq x^* \delta | x_i = x^*) - \frac{1}{2} > 0.
\]

by A2. If \( x^* \delta > 0 \), first note since for any \( x_i \),

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

So at \( c^*, x^* \), where \( x^* \beta_0 < c^* \), \( x^* \delta > 0 \), the expected value of the above expression is greater than \( \frac{1}{2} \). Consequently \( \tau_0(x^*, \beta) < 0 \). Therefore, we have established \( g(x^*, \beta) \neq 0 \) for any \( \beta \neq \beta_0 \) and any \((c^*, x^*) \in S_{cx} \), which has positive measure. \( \blacksquare \)
Remark 2.1 Our identification result in Lemma 2.1 uses available information in the two functions $\tau_1(\cdot, \cdot)$ and $\tau_0(\cdot, \cdot)$. We can contrast this with the procedure in Ying et al. (1995) which is only based on the function $\tau_1(\cdot)$, and consequently requires to reweight the data using the Kaplan Meier (1958) estimator. As alluded to previously, this imposes strong support conditions on the censoring variable does not allow for covariate dependent censoring, unless one uses the conditional Kaplan Meier estimator in Beran (1981) to reweight the data, which introduces the complication of selecting smoothing parameters and trimming procedures.

Lemma 2.1 establishes a conditional inequality moment condition which we aim to base our estimator for $\beta_0$ on. Recent work on conditional moment equality conditions includes Donald et al. (2003), and Dominguez and Lobato (2004)\(^3\).

Our identification result and estimation procedure will be distinct 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. Our model is based on a set of conditional moment inequalities that are satisfied uniquely for all $x$ at the truth. To ensure global identification from a set of conditional moment inequalities, we provide a procedure that takes account of moment inequalities and preserve global point identification\(^4\).

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

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

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

where above the inequality $t_1 \leq x_i$, corresponds to each component of the two vectors. From Billingsley (1995, Theorem 16.10) a conditional moment condition can be related to unconditional moment conditions with indicator functions as above comparing regressor

\(^3\)Similar results are also given in Koul (2002) and Stute (1996).

\(^4\)Rosen (2005) uses a similar $U$-process approach to estimate models that are set identified.
values to all vectors in \( \mathbb{R}^k \). Our identification result will need to make use of two distinct vectors in \( \mathbb{R}^k \). As we will see below, this will translate into estimation procedures involving higher order (second or third) U-processes.

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

\[
Q(\beta) = E[H_1(\beta, x_j, x_k)I[H_1(\beta, x_j, x_k) \geq 0] - H_0(\beta, x_j, x_k)I[H_0(\beta, x_j, x_k) \leq 0]]
\] (2.9)

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

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

**Proof:** We first show that

\[
Q(\beta_0) = 0
\] (2.10)

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

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

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

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

\[ \hat{\beta} = \arg \min_{\beta \in B} \hat{Q}_n(\beta) \] (2.13)

\[ = \arg \min_{\beta \in B} \frac{1}{n(n-1)} \sum_{j \neq k} \left\{ \hat{H}_1(\beta, x_j, x_k) I[\hat{H}_1(\beta, x_j, x_k) \geq 0] \right. \] 

\[ \left. - \hat{H}_0(\beta, x_j, x_k) I[\hat{H}_0(\beta, x_j, x_k) \leq 0] \right\} \] (2.14)

**Remark 2.2** We note the above objective is similar to a standard LAD objective function, since for a random variable \( z \), we can write \( |z| = zI[z \geq 0] - zI[z \leq 0] \). The difference lies in the fact that our objective function “switches” from \( \hat{H}_1(\cdot, \cdot, \cdot) \) to \( \hat{H}_0(\cdot, \cdot, \cdot) \) when moving from the positive to the negative region. Switching functions is what permits us to allow for general forms of censoring.

**Remark 2.3** As we can see, the above estimation procedure minimizes a (generated) second order U-process. We note that, analogously to existing rank estimators, (e.g., Han(1987), Cavanagh and Sherman(1997), Khan and Tamer(2005)), this provides us with an estimation procedure without the need to select smoothing parameters or trimming procedures. However, we note our proposed procedure 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. To that end, we point out that a “split sample” approach (see, e.g. Newey and Powell(1990) and Honoré and Powell(1994)) would result in an estimator that minimizes a second order U-process that is much simpler computationally, though less efficient. For the problem at hand, the split sample version of our proposed estimator would minimize an objective function of the form:
\[ \hat{\beta}_{SS} = \arg \min_{\beta \in \mathcal{B}} Q_{SSn}(\beta) \]  
\[ = \arg \min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^{n} \left\{ \hat{H}_1(\beta, x_j, x_{n-j+1}) I[\hat{H}_1(\beta, x_j, x_{n-j+1}) \geq 0] - \hat{H}_0(\beta, x_j, x_{n-j+1}) I[\hat{H}_0(\beta, x_j, x_{n-j+1}) \leq 0] \right\} \]

We next turn attention to the asymptotic properties of the estimator in (2.13). We begin by establishing consistency under the following assumptions.

**C1** The parameter space \( \mathcal{B} \) is a compact subset of \( \mathbb{R}^k \).

**C2** The sample vector \((d_i, v_i, x'_i)'\) is i.i.d.

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

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

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

\[ \hat{\beta} \overset{p}{\to} \beta_0 \]

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

**D1** \( \beta_0 \) is an interior point of the parameter space \( \mathcal{B} \).

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

**D3** The censoring values \( c_i \) are distributed independently of \( \epsilon_i \) conditionally on the regressors. We note this weak restriction permits both conditional heteroskedasticity and covariate dependent censoring.
D4 The regressors $x_i$ and censoring values $\{c_i\}$ satisfy

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

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

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

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

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

(2.17)

where we define $V$ as follows. Let $C$ denote the subset of $\mathcal{X}$ where for all $x \in C$, $P(c_i \geq x_i'\beta_0|x_i = x) = 1$. Adopting the notation $I_{ijk} = I[x_j \leq x_i \leq x_k]$, define the function

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

(2.18)

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

$$
V = 2E[G(x_j, x_k)G(x_j, x_k)']
$$

(2.19)

Next we define the outer score term $\Omega$. Let

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

(2.20)

so we can define the outer score term $\Omega$ as

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

(2.21)

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.
3 Endogenous Regressors, Endogenous Censoring 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 $\beta_0$ when the regressors $x_i$ and/or the censoring variable $c_i$ may be endogenous and one has a vector of instrumental variables $z_i$.

The semiparametric literature has seen recent developments in estimating censored regression models when the regressors are endogenous. Instrumental variable approaches have been propose in Hong and Tamer(2003) and Lewbel(2000), whereas a control function approach has been proposed in Blundell and Powell(2005). However, none of these estimators are 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 censoring and/or endogenous regressors. Allowing for endogenous censoring brings the model more in line with widely studied models like selection models and the Roy model (see, e.g. Heckman and Honore(1990)). Endogenous regressors can arise 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 $\beta_0$ in this setting we assume the availability of a vector of instrumental variables $z_i$. Here, our sufficient conditions for identification are:

**A’1** $\text{median}(\epsilon_i|z_i) = 0$.

**A’2** The subset of the support of instruments

$C_{Z0} = \{z_i : P(x_i' \beta_0 \leq c_i|z_i)\}$

has positive measure, as does either of the following subsets of $C_{Z0}$:

$C_{Z0-} = \{z_i \in C_{Z0} : P(x_i' \delta \leq 0|z_i) = 1 \ \forall \delta \neq 0\}$

$C_{Z0+} = \{z_i \in C_{Z0} : P(x_i' \delta \geq 0|z_i) = 1 \ \forall \delta \neq 0\}$
A’3 $c_i \perp e_i|z_i$.

**Remark 3.1** Before outlining an estimation procedure, we comment on the meaning of the above conditions.

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

2. Condition A’2 details the relationship between the instruments and the regressors. It is most easily satisfied when the exogenous variable(s) has a support that is relatively large when compared to the support of the endogenous variable(s). 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.

3. Condition A’3 is weaker than A3 in the sense that correlation between the censoring variable and the errors conditional on the regressors is now permitted if a suitable instrumental variable is available. This is analogous to the usual exclusion restriction imposed in selection models, where in this context we have an explanatory variable that effects the censoring process but not the latent dependent variable location function. We note that Assumptions A’1 and A’3 by themselves are sufficient for identification of $\beta_0$ in a model with endogenous censoring but exogenous regressors.

Our IV limiting objective function is of the form:

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

(3.1)

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, this corresponds to obtaining identification from the region of exogenous variable(s) space where the data is uncensored with probability 1.
where here

\[
H_1^*(\beta, t_1, t_2) = E[(I[v_i \geq x_i'\beta] - \frac{1}{2})I[t_1 \leq z_i \leq t_2]] \\
H_0^*(\beta, t_1, t_2) = E[(\frac{1}{2} - d_iI[v_i \leq x_i'\beta])I[t_1 \leq z_i \leq t_2]]
\]

(3.2) (3.3)

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

\[
\hat{H}_1(\beta, z_j, z_k) = \frac{1}{n} \sum_{i=1}^{n} (I[v_i \geq x_i'\beta] - \frac{1}{2})I[z_j \leq z_i \leq z_k] \\
\hat{H}_0(\beta, z_j, z_k) = \frac{1}{n} \sum_{i=1}^{n} (\frac{1}{2} - d_iI[v_i \leq x_i'\beta])I[z_j \leq z_i \leq z_k]
\]

(3.4) (3.5)

The asymptotic properties of this estimator are based on the following regularity conditions:

C'1 \( \beta_0 \) lies in the interior of \( B \), a compact subset of \( \mathbb{R}^k \).

C'2 The vector \((d_i, v_i, x_i', z_i')\) is i.i.d.

C'3 The objective function

\[
Q_{IV}(\beta) \equiv E[H_1^*(\beta, z_j, z_k)I[H_1^*(\beta, z_j, z_k) \geq 0] - H_0^*(\beta, z_j, z_k)I[H_0^*(\beta, z_j, z_k) \leq 0]]
\]

is continuous at \( \beta = \beta_0 \).

C'4 Conditional on the instrumental variables, the error terms are continuously distributed, with conditional density function denoted by \( f(\epsilon|z) \), which in addition to satisfying the conditional median 0 condition, is bounded, Lipschitz continuous in \( \epsilon \), and bounded away from 0 for \( \epsilon \) in a neighborhood of 0, uniformly across \( z_i \).

C'5 The censoring variables, regressors and instrumental variables also satisfy the following conditions:

(i) \( P(|c_i - x_i'\beta| \leq d|z_i) = O(d) \) uniformly in all values of \( z_i \) and \( \|\beta - \beta_0\| < \eta_0 \) for some \( \eta_0 > 0 \)

(ii) \( E[I[c_i - x_i'\beta_0 > \eta_0]x_i|x_i'z_i] \) is invertible for all values of \( z_i \) and some \( \eta_0 > 0 \).
Under the above conditions, we can establish the limiting distribution theory for the IV estimator. The arguments of the proof are completely analogous those used to prove the previous limiting distribution theory result and are omitted.

**Theorem 3.1** Under Assumptions A’1-A’3, C’1-C’4

$$\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}_Z\Omega_Z V^{-1}_Z)$$  \hspace{1cm} (3.6)

where we define $V_Z$ as follows. Let $C_Z$ denote the subset of $Z$ where for all $z \in C_Z$, $P(c_i \geq x_i' \beta_0 | z_i = z) = 1$. Define the function

$$G_Z(z_j, z_k) = \int f(0 | z_i) x_i I[z_j \leq z_i \leq z_k] f_Z(z_i) dz_i$$  \hspace{1cm} (3.7)

where $f_Z(\cdot)$ denotes the instrumental variable density function. Then

$$V_Z = 2E[I[[z_j, z_k] \subseteq C] G(z_j, z_k) G(z_j, z_k)']$$  \hspace{1cm} (3.8)

Next we define the outer score term $\Omega$. Now define the function

$$G^*_Z(z_j, z_k, z_i) = G_Z(z_j, z_k) I[z_j \leq z_i \leq z_k]$$

from which we define

$$\bar{G}_Z(z_i) = E[G^*_Z(z_j, z_k, z_i) | z_i]$$

The outer score term is defined as

$$\Omega_Z = E[\delta_0 \delta_0']$$

with

$$\delta_0 = \bar{G}(z_i)(I[v_i \geq x_i' \beta_0] - d_i I[v_i \leq x_i' \beta_0])$$

4 **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 endogenity, and then we control for selective compliance to treatment using our proposed i.v. estimator.
4.1 Simulation Results

The model used in this simulation study is

\[ y_i = \min\{\alpha_0 + x_{1i}\beta_0 + x_{2i}\gamma_0 + \epsilon_i, c_i\} \quad (4.1) \]

where the regressor \( x_{1i}, x_{2i} \) were chi-squared, 1 degree of freedom, and standard normal respectively. The true values \( \alpha_0, \beta_0, \gamma_0 \) 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 \( c_i = -x_{1i}^2 - x_{2i} \).

We assumed the error distribution of \( \epsilon_i \) was standard normal. In addition, we simulated designs with heteroskedastic errors as well:

\[ \epsilon_i = \sigma(x_i) \cdot \eta_i, \]

with \( \eta_i \) having a standard normal distribution and \( \sigma(x_i) = \exp(0.5 \cdot x_i) \). For these designs, the overall censoring probabilities vary between 40% and 50%. For each replication of the model, the following estimators were calculated\(^6\):

a) The minimum distance least absolute deviations (MD) estimator introduced in this paper in (2.13) and its split sample variant (MDSS) in (2.15).

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

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

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

Both the Ying et al. (1995) and MD, MDSS estimators were computed using the Nelder Meade simplex algorithm.\(^7\) The randomly-censored least absolute deviations estimator (HKP) was computed using the iterative Barrodale-Roberts algorithm described by Buchinsky (1995)\(^8\); 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.

\(^6\)The simulation study was performed in GAUSS and C++. Codes for the estimators introduced in this paper are available from the authors upon request.

\(^7\)OLS, LAD, and true parameter values were used in constructing the initial simplex for the results reported.

\(^8\)OLS was used as the starting value when implementing this algorithm for the simulation study.
The results of 401 replications of these estimators for each design, with sample sizes of 50, 100, 200, and 400, are summarized in Tables I-IV, which report the mean bias, median bias, root-mean-squared error, and mean absolute error. These 4 tables corresponded to designs with 1) homoskedastic errors and covariate independent censoring, 2) heteroskedastic errors and covariate independent censoring, 3) homoskedastic errors and covariate dependent censoring, and 4) heteroskedastic errors and covariate dependent censoring. Theoretically, only the MD estimator introduced here is consistent in all designs, and the only estimator which is consistent in design 4.

HKP and Ying et al.(1995) estimators are consistent under designs 1 and 2, 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 methods proposed here perform relatively well. Throughout, the MDSS has a larger RMSE than that of the MD, indicating the relative computational ease of the MDSS comes at the expense of a loss of efficiency compared to the MD. 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 and MDSS estimators, which do not impose homoskedasticity of the error terms, is superior to Buckley-James when the errors are heteroskedastic. They generally outperform HKP and Ying et al.(1995) estimator when the censoring variable depends on the covariates. This is especially the case when the errors are heteroskedastic, as in this design, the proposed estimators are the only ones which performs reasonably well.

### 4.2 Empirical Example

#### 4.2.1 Drug Relapse Duration

We apply the minimum distance procedure to the drug relapse data set used in Hosmer and Lameshow(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.

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

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

Qualitatively, all estimators deliver similar results in the sense that the signs of the coefficients are the same. However there are noticeable differences in the values of these estimates, as well as their significance. 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.

---

9Reported standard errors for the 4 semiparametric estimators were obtained from the bootstrap, using 575 samples (obtained with replacement) of 575 observations.
Given the random assignment of the type of treatment, the treatment indicator (TREAT) is a natural choice (see, e.g. Bloom(1984)) of an instrumental variable as it is correlated with LOS.

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

5 Conclusions

This paper introduces a new estimation procedure for an AFT model with conditional heteroskedasticity and general censoring 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. A less efficient, though computationally simple split sample estimator was also proposed.

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


A  Proof of Theorem 2.1

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

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

where

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

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

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

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

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

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

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

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

$$\sup_{\beta \in B} \left| \frac{1}{n(n-1)} \sum_{i \neq j} I[\hat{H}_1(x_j, x_k, \beta) \geq 0] - I[H_1(x_j, x_k, \beta) \geq 0] \right| = o_p(1) \quad (A.5)$$
To do so, we will show that $I[\hat{H}_1(x_j, x_k, \beta) \geq 0 | I[H_1(x_j, x_k, \beta) < 0]$ is $o_p(1)$ uniformly in $x_j, x_k, \beta$. We fix the values of $x_j, x_k$ and decompose $I[H(x_j, x_k, \beta) < 0]$ as

$$I[H_1(x_j, x_k, \beta) < 0] = I[H_1(x_j, x_k, \beta) \leq -\delta_n] + I[H_1(x_j, x_k, \beta) \in (-\delta_n, 0)]$$  \hspace{1cm} (A.6)

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

$$I[\hat{H}_1(x_j, x_k, \beta) \geq 0 | I[H_1(x_j, x_k, \beta) \leq -\delta_n]$$  \hspace{1cm} (A.7)

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

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

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

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

To which we can apply Hoeffding’s inequality, see, e.g. Pollard(1984), to bound above by $\exp(-2n\delta_n^2)$.

Note this bound is independent of $x_j, x_k, \beta$, and converges to 0 at the rate $n^{-2}$, establishing the uniform (across $x_j, x_k, \beta$) convergence of

$$I[\hat{H}(x_j, x_k, \beta) \geq 0 | I[H(x_j, x_k, \beta) \leq -\delta_n] to 0. We next show the uniform convergence of

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

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

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

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

Next, we will establish that

$$\sup_{\beta \in B} \left| \frac{1}{n(n-1)} \sum_{i \neq j} H_1(x_j, x_k, \beta) I[H_1(x_j, x_k, \beta) \geq 0] - Q_1(\beta) \right| = o_p(1)$$  \hspace{1cm} (A.12)
For this we can apply existing uniform laws of large numbers for centered $U-$ processes. Specifically, we can show the r.h.s. of (A.12) is $O_p(n^{-1/2})$ by Corollary 7 in Sherman(1994a) since the functional space index by $\beta$ is Euclidean for a constant envelope. The Euclidean property follows from example (2.11) in Pakes and Pollard(1989).

\[\Box\]

### B Proof of Theorem 2.2

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

Here, we follow the approach in Sherman(1994b). Having already established consistency of the estimator, we will first establish root-$n$ consistency and asymptotic normality. For root-$n$ consistency we will apply Theorem 1 of Sherman(1994b). We will actually apply this theorem twice, first establishing a slower than root-$n$ consistency result and then root-$n$ consistency. Keeping our notation deliberately as close as possible to Sherman(1994b), here we denote our sample objective function $\hat{Q}_n(\beta)$ by $G_n(\beta)$ and denote our limiting objective function $Q(\beta)$ by $G(\beta)$. From Theorem 1 in Sherman(1994b), sufficient conditions for rates of convergence are that

1. $\hat{\beta} - \beta_0 = o_p(1)$

2. There exists a neighborhood of $\beta_0$ and a constant $\kappa > 0$ such that $G(\beta) - G(\beta_0) \geq \kappa \|\beta - \beta_0\|^2$ for all $\beta$ in this neighborhood.

3. Uniformly over $o_p(1)$ neighborhoods of $\beta_0$

\[
G_n(\beta) = G(\beta) + O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + O_p(\epsilon_n)
\]

which suffices for $\hat{\beta} - \beta_0 = O_p(\max(\epsilon^{1/2}, n^{-1/2}))$ To show the second condition, we will first derive an expansion for $G(\beta)$ around $G(\beta_0)$. We denote that even though $G_n(\beta)$ is not differentiable in $\beta$, $G(\beta)$ is sufficiently smooth for Taylor expansions to apply as the expectation operator is a smoothing operator and the smoothness conditions in Assumptions D2, D4 (i). Taking a second order expansion of $G(\beta)$ around $G(\beta_0)$, we obtain

\[
G(\beta) = G(\beta_0) + \nabla_\beta G(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta} G(\beta_0)(\beta - \beta_0) \quad (B.1)
\]
where $\nabla_{\beta}$ and $\nabla_{\beta\beta}$ denote first and second derivative operators and $\beta^*$ 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

$$\nabla_{\beta\beta}G(\beta_0) = V$$

and $V$ is positive definite by Assumption A3, so we have

$$(\beta - \beta_0)'\nabla_{\beta\beta}G(\beta_0)(\beta - \beta_0) > 0$$

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

$$(\beta - \beta_0)'\nabla_{\beta\beta}G(\beta)(\beta - \beta_0) > 0$$

which suffices for the second condition to hold.

To show the third condition, our first step is to replace the indicator functions in the objective function,

$I[\hat{H}_1(x_j, x_k, \beta) \geq 0], I[\hat{H}_0(x_j, x_k, \beta) \leq 0]$, with true functions

$I[H_1(x_j, x_k, \beta) \geq 0], I[H_0(x_j, x_k, \beta) \leq 0]$, and derive the corresponding representation. (We will deal with the resulting remainder term from this replacement shortly.) We expand the terms $\hat{H}_1(x_j, x_k, \beta), \hat{H}_0(x_j, x_k, \beta)$, first exclusively dealing with the first expansion, resulting in the third order $U$-process:

$$\frac{1}{n(n-1)(n-2)} \sum_{j \neq k \neq l} I[H_1(x_j, x_k, \beta) \geq 0](I[v_l \geq x'_l] - \frac{1}{2})I_{ijk}$$

where here $I_{ijk} = I[x_j \leq x_i \leq x_k]$. Turning attention to (B.5), we will apply projection theorems to representation of degenerate $U$-processes- see, e.g. Serfling (1980). We note the unconditional expectation corresponds to the first "half" of the limiting objective function, which recall here we denoted by $G(\beta)$. We will evaluate representations for expectations conditional on each of the three arguments, minus the unconditional expectation. We first turn attention to the expectation conditional on the third argument in the $U$-statistic which is denoted above by the subscript $l$. This summation will be of the form

$$\frac{1}{n} \sum_{i=1}^{n} (I[v_l \geq x'_l] - \frac{1}{2})E[I[H_1(x_j, x_k, \beta) \geq 0]|I_{ijk}] x_i]$$

(B.6)
We will take a mean value expansion of \( E[I[H_1(x_j, x_k, \beta) \geq 0]] \) (B.7) around \( \beta = \beta_0 \). Note by our normalization the initial replacement of \( \beta \) with \( \beta_0 \) yields a term that is identically 0. Notice also that \( I[H_1(x_j, x_k, \beta_0) \geq 0] \cdot I_{ijk} = 1 \) implies that \( x_l \in C \), as is \([x_j, x_k] \); we next evaluate

\[
\nabla_\beta E[I[H_1(x_j, x_k, \beta) \geq 0]]
\]

at \( \beta = \beta_0 \). Here we note we can write:

\[
H_1(x_j, x_k, \beta) = \int I[x_j \leq u \leq x_k]S_{\epsilon[u]}(u'\beta) \left\{ S_{\epsilon[u]}(u'(\beta - \beta_0)) - \frac{1}{2} \right\} f_X(u)du
\]

and

\[
E[I[H_1(x_j, x_k, \beta) \geq 0]] = \int_0^\infty H(x_j, x_k, \beta)dF_{H_1(x_j, x_k, \beta)}
\]

when integrating over the different values of \( x_j, x_k \), here we decompose the set of values into those satisfying the interval \([x_j, x_k] \) contained in \( C \) and those that do not. We do this because \( H_1(x_j, x_k, \beta_0) < 0 \) in the latter case and we are outside of the range of integration. Two applications of the dominated convergence theorem on the subset of values where \([x_j, x_k] \) is contained in \( C \) yields that (B.8) is of the form:

\[
E[G(x_j, x_k)]
\]

where recall

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

so combining this expansion term with (B.6), yields

\[
\frac{1}{n} \sum_{l=1}^n I[x_l \in C](I[v_l \geq x_l'\beta] - \frac{1}{2})E[G(x_j, x_k)]'(\beta - \beta_0)
\]

plus a remainder term involving the second derivative of \( E[I[H_1(x_j, x_k, \beta) \geq 0]] \) with respect to \( \beta \).

\[^{10}\]An alternative proof strategy, which yields the same result is to replace the non-smooth function \( I[v_l \geq x_l'\beta] - \frac{1}{2} \) with its expectation conditional on \( x_l \). Since this conditional expectation is smooth we can then apply the expansion to the product of this (smooth) conditional expectation and the (smooth) function \( E[I[H_1(x_j, x_k, \beta) \geq 0]] \). Such an approach can be shown to be valid using analogous empirical process theory arguments to those used in, say, Andrews(1994).
Recall that in our $U$–statistic decomposition, we are subtracting the unconditional expectation from the conditional expectation, implying that (B.12) has mean 0 after this subtraction. At this stage we can replace $I[v_i \geq x_i' \beta]$ with $I[v_i \geq x_i' \beta_0]$ in (B.12), and by locally uniform central limit theorems (see, e.g, example 2.11 Pakes and Pollard(1989)), the resulting remainder term is uniformly $o_p(\|\beta - \beta_0\|/\sqrt{n})$.

Note that the above term (B.12) with $I[v_i \geq x_i' \beta_0]$ replacing $I[v_i \geq x_i' \beta]$ has expectation (across $v_i, x_i$) of 0, and finite variance, so a CLT can be applied to the above summation, so this piece will correspond to the $O_p(\|\beta - \beta_0\|/\sqrt{n})$ term in Theorem 1 in Sherman(1994b). By the law of large numbers, the remainder term in the mean value expansion of $E[H_1(x_j, x_k, \beta) \geq 0]$ yields a term that is $o_p(\|\beta - \beta_0\|^2)$.

So we have established that so far the sample objective function can be represented as the limiting objective function plus

$$O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + o_p(\frac{1}{n}) \quad (B.13)$$

Using the same arguments, it can be shown that the expectation conditional on the first term, indexed by subscript $j$ has the representation:

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

as does the expectation conditional on the second argument, indexed by $k$. Finally, we note the higher order terms in the projection theorem are $o_p(n^{-1})$ uniformly for $\beta$ in $o_p(1)$ neighborhoods of $\beta_0$ using arguments similar to those in Theorem 3 in Sherman(1993). So by Theorem 1 in Sherman(1994b), we are able to show, (focusing exclusively on the first "half" of the sample objective function for simplicity), 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 $\beta_0$),

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

Collecting terms would enable us to apply Theorem 1 of Sherman(1994b) to conclude that the infeasible estimator, which replaced $I[\hat{H}_1(x_j, x_k, \beta) \geq 0], I[\hat{H}_0(x_j, x_k, \beta) \geq 0]$ with $I[H_1(x_j, x_k, \beta) \geq 0], I[H_0(x_j, x_k, \beta) \geq 0]$ respectively, is $O_p(n^{-1/2})$. To derive a rate of convergence for the actual estimator, we will derive a rate for :

$$\sum_{j \neq k} (I[\hat{H}_1(x_j, x_k, \beta) \geq 0] - I[H_1(x_j, x_k, \beta) \geq 0]) \hat{H}_1(x_j, x_k, \beta) \quad (B.16)$$

26
as well as the second "half" involving \( H_0(x_j, x_k, \beta) \). We note that since \( \hat{H}_1(x_j, x_k, \beta) \) is uniformly bounded in \( \beta, x_j, x_k \), to establish the negligibility of (B.16), we will formally show that

\[
\frac{1}{n(n-1)} \sum_{j \neq k} I[\hat{H}_1(x_j, x_k, \beta) < 0]I[H_1(x_j, x_k, \beta) \geq 0] = o_p(\log nn^{-1/2}) \quad (B.17)
\]

uniformly in \( \beta \) in \( o_p(1) \) neighborhoods of \( \beta_0 \). To do so, we decompose

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

(B.18)

where \( \delta_n \) is a sequence of positive numbers converging to 0 at the rate \( \log n/\sqrt{n} \). We aim to show each of the above indicator functions multiplied by \( I[H_1(x_j, x_k, \beta) \geq 0] \) corresponds to an \( o_p(n^{-1/2}) \) term, uniformly in \( \beta \) in \( o_p(1) \) neighborhoods of \( \beta_0 \). First, dealing with

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

(B.19)

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

\[
P(\hat{H}_1(x_j, x_k, \beta) < -\delta_n) \to 0
\]

(B.20)

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

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

(B.21)

which is less than or equal to

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

(B.22)

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

\[
P(\hat{H}_1(x_j, x_k, \beta) \in [-\delta_n, 0)) \to 0
\]

(B.23)

for \( \beta \) uniformly in an \( o_p(1) \) neighborhood of \( \beta_0 \) and \( H_1(x_j, x_k, \beta) \geq 0 \). We decompose the above probability into

\[
P(\hat{H}_1(x_j, x_k, \beta) \in [\delta_n, 0)), |H_1(x_j, x_k, \beta)| \leq 2\delta_n)
\]

(B.24)
\[+ P(\hat{H}_1(x_j, x_k, \beta) \in [-\delta_n, 0)), |H_1(x_j, x_k, \beta)| > 2\delta_n) \quad (B.25)\]

The first piece of the decomposition is less than or equal to
\[P(|H_1(x_j, x_k, \beta)| \leq 2\delta_n) = O(\delta_n) \quad (B.26)\]
uniformly in \(\beta\) in \(o_p(1)\) neighborhoods of \(\beta_0\) where the equality uses Assumption \textbf{D4(i)}. The second probability can be bounded above by:
\[P(|\hat{H}_1(x_j, x_k, \beta) - H_1(x_j, x_k, \beta)| \geq \delta_n) \quad (B.27)\]
which can be written as
\[P(\sum_{i=1}^{n} (I[v_i \geq x_i^\prime \beta] - \frac{1}{2}) - H_1(x_j, x_k, \beta) | \geq n\delta_n) \quad (B.28)\]
To which we can apply the corollary to Hoeffding’s inequality (see, e.g. Pollard(1984)) to conclude that the above probability is bounded above by \(\exp(-2n\delta_n^2)\) which converges to 0 at the rate \(n^{-2}\). This establishes (B.17). Similar arguments can be used to show that
\[\frac{1}{n(n-1)} \sum_{j \neq k} I[\hat{H}_1(x_j, x_k, \beta) \geq 0] I[H_1(x_j, x_k, \beta) < 0] = O_p(\log nn^{-1/2}) \quad (B.29)\]
uniformly for \(\beta\) in \(o_p(1)\) neighborhoods of \(\beta_0\), which establishes (B.16) converges to 0. Combining all our terms, the rate of convergence is the rate of the slowest term which as of now is \(O_p(\delta_n)\), from which we can apply Theorem 1 of Sherman(1994b) to conclude the estimator converges at the rate \(O_p(\delta_n^{1/2})\). We can use this intermediate rate to get the desired rate of \(O_p(n^{-1/2})\) by establishing the following interaction term
\[\frac{1}{n(n-1)} \sum_{j \neq k} \hat{H}_1(x_j, x_k, \beta) I[\hat{H}_1(x_j, x_k, \beta) \geq 0] I[H_1(x_j, x_k, \beta) < 0] \quad (B.30)\]
is \(o_p(n^{-1})\). We proceed as before, letting \(\delta_n\) denote a sequence of positive numbers that is \(O(\sqrt{\log n/n})\). Again we will decompose the indicator \(I[\hat{H}_1(x_j, x_k, \beta) \geq 0]\) into the sum of two indicators:
\[I[\hat{H}_1(x_j, x_k, \beta) \geq \delta_n] + I[\hat{H}_1(x_j, x_k, \beta) \in [0, \delta_n]]\]
Using Hoeffding’s inequality as before we can conclude that
\[\frac{1}{n(n-1)} \sum_{j \neq k} \hat{H}_1(x_j, x_k, \beta) I[\hat{H}_1(x_j, x_k, \beta) \geq \delta_n] I[H_1(x_j, x_k, \beta) < 0]\]
is $O_p(n^{-2})$. Next we establish the rate of convergence of

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

which is less than or equal to

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

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

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

We note the above expression is $o_p(\frac{1}{n})$ uniformly in $\beta$ in $o_p(\sqrt{n})$ neighborhoods of $\beta_0$ as we have already established that

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

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

Consequently, we may conclude that

$\hat{\beta} - \beta_0 = O_p(n^{-1/2})$ by Theorem 1 in Sherman(1994b). We can now turn attention to establishing asymptotic normality of the estimator.

Now that root-$n$ consistency has been established we can apply Theorem 2 in Sherman(1994b) to attain asymptotic normality. A sufficient condition is that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\beta_0$,

$$\mathcal{G}_n(\beta) - \mathcal{G}_n(\beta_0) = \frac{1}{2}(\beta - \beta_0)'V(\beta - \beta_0) + \frac{1}{\sqrt{n}}(\beta - \beta_0)'W_n + o_p(\frac{1}{n}) \quad (B.31)$$

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.31). Here, we will again work with the $U$-statistic decomposition in, for example, Serfling(1980) as our objective function is a third order $U$-process. We will first derive an expansion for $\mathcal{G}(\beta)$ around $\mathcal{G}(\beta_0)$, since $\mathcal{G}(\beta)$ is related to the limiting objective function. We denote that even though $\mathcal{G}_n(\beta)$ is not differentiable in $\beta$, $\mathcal{G}(\beta)$ is sufficiently
smooth for Taylor expansions to apply by Assumptions D2, D4(i). Taking a second order expansion of \( G(\beta) \) around \( G(\beta_0) \), we obtain

\[
G(\beta) = G(\beta_0) + \nabla_\beta G(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)'\nabla_{\beta\beta} G(\beta^*)(\beta - \beta_0)
\]  

where \( \nabla_\beta \) and \( \nabla_{\beta\beta} \) denote first and second derivative operators and, and \( \beta^* \) 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:

\[
\nabla_{\beta\beta} G(\beta^*) = V + o_p(1) \tag{B.33}
\]

To formally show the above result, we first expand \( \hat{H}_1(\beta, x_j, x_k) \) in the double summation. As before, we will replace \( \hat{H}_1(x_j, x_k, \beta) \) with \( H_1(x_j, x_k, \beta) \) inside the indicator function, since we have already shown such a replacement results in an asymptotically negligible remainder term.

\[
\hat{H}_1(\beta, x_j, x_k) = \frac{1}{n} \sum_{i=1}^{n} (I[v_i \geq x'_i\beta] - \frac{1}{2})I[x_j \leq x_i \leq x_k] \tag{B.34}
\]

Similarly, we have:

\[
\hat{H}_0(\beta, x_j, x_k) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} - d_i I[v_i \leq x'_i\beta]\right)I[x_j \leq x_i \leq x_k] \tag{B.35}
\]

At this stage, we will concentrate on the \( \hat{H}_1(x_j, x_k, \beta) \) half of the objective function and take the expectation of the following third order \( U \)-process:

\[
\frac{1}{n(n-1)(n-2)} \sum_{j \neq k \neq l} I[H_1(x_j, x_k, \beta) \geq 0]I[x_j \leq x_l \leq x_k](I[v_l \geq x'_l\beta] - \frac{1}{2}) \tag{B.36}
\]

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

First, we condition on the regressors \( x_j, x_k, x_l \), we get
\[ I[H_1(x_j, x_k, \beta) \geq 0]I[x_j \leq x_l \leq x_k] \]  
\[ (S_e(x_i'(\beta - \beta_0))S_c(x_i'\beta) - \frac{1}{2}) \]  
(B.37)

Where \( S_e(\cdot), S_c(\cdot) \) denote the conditional survivor functions of \( \epsilon_i \) and \( c_i \) respectively. Expanding around \( \beta_0 \), noting we have the product of two functions of \( \beta \), we note the first term and the derivative term are 0 since \( x_l \in C \). The second derivative term in the expansion is of the form
\[ (\beta - \beta_0)^2 \cdot I[[x_j, x_k] \subseteq C] \nabla_\beta H_1(x_j, x_k, \beta_0) I_{ijk} f_\epsilon(0|x_i)x_i'(\beta - \beta_0) \]  
(B.38)

Next, we take expectations of the above term conditional on \( x_j, x_k \). This gives the following function of \( x_j, x_k \):
\[ 2I[[x_j, x_k] \subseteq C] \nabla_\beta H_1(x_j, x_k, \beta_0) \int f_\epsilon(0|x_i)x_i I_{ijk} f_X(x_i) dx_i \]  
(B.39)

where \( f_X(\cdot) \) denotes the regressor density function. Finally, we can apply the same arguments to the other ”half” of the objective function involving \( H_0(x_j, x_k, \beta_0) \).

Combining all these results we may conclude that form of \( V \) in the quadratic approximation in Theorem 2 in Sherman(1994b) is of the form
\[ V = 2E[I[[x_j, x_k] \subseteq C]G(x_j, x_k)G(x_j, x_k')] \]  
(B.40)

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 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 \( I[v_l \geq x_i'\beta - \frac{1}{2}] \) as well as \( \frac{1}{2} - d_l I[v_l \leq x_i'\beta] \), 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 \( \beta \) in \( O_p(n^{-1/2}) \) neighborhoods of \( \beta_0 \). The same applies to the expectation conditional on the second argument of the third-order \( U \)-process, denoted by the subscript \( k \).

We therefore turn attention to expectation conditional on the third argument, denoted by the subscript \( l \). Here we proceed as before when showing root-\( n \) consistency, expanding
\[ I[H_1(x_j, x_k, \beta) \geq 0]I[v_l \geq x_i'\beta] - \frac{1}{2} \]
around $\beta = \beta_0$. Recall this yielded the mean 0 process:

$$
\frac{1}{n} \sum_{l=1}^{n} E\left[G(x_j, x_k)I_{ijk}[I[v_l \geq x'_l/\beta_0] - \frac{1}{2}]\right] (B.42)
$$

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

$$
\frac{1}{n} \sum_{l=1}^{n} E\left[G(x_j, x_k)I_{ijk}[I[v_l \geq x'_l/\beta_0] - d_l I[v_i \leq x'_i/\beta_0]]'(\beta - \beta_0) + o_p(n^{-1}) \right] (B.43)
$$

which corresponds to

$$
\frac{1}{\sqrt{n}} (\beta - \beta_0)'W_n \quad (B.44)
$$

where

$$
W_n \Rightarrow N(0, E[\delta_0\delta_0']) (B.45)
$$

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 $\beta$ in an $O_p(n^{-1/2})$ neighborhood of $\beta_0$ 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

$$
\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, V^{-1}\Omega V^{-1}) \quad (B.47)
$$

where

$$
\Omega = E[\delta_0\delta_0']
$$

Which establishes the proof of the theorem. ■
# TABLE I
Simulation Results for Censored Regression Estimators
Cl Censoring, Homosked. Errors

<table>
<thead>
<tr>
<th></th>
<th>Mean Bias</th>
<th>Med. Bias</th>
<th>RMSE</th>
<th>MAD</th>
<th>Mean Bias</th>
<th>Med. Bias</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>100 obs.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.2411</td>
<td>0.1159</td>
<td>0.6098</td>
<td>0.1963</td>
<td>-0.1617</td>
<td>-0.0695</td>
<td>0.7169</td>
<td>0.1948</td>
</tr>
<tr>
<td>MDSS</td>
<td>0.0098</td>
<td>0.0115</td>
<td>0.4787</td>
<td>0.2956</td>
<td>0.0403</td>
<td>0.0679</td>
<td>0.5065</td>
<td>0.2971</td>
</tr>
<tr>
<td>HKP</td>
<td>-0.0322</td>
<td>-0.0356</td>
<td>0.2746</td>
<td>0.2182</td>
<td>0.0009</td>
<td>0.0051</td>
<td>0.1600</td>
<td>0.1249</td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0296</td>
<td>-0.0437</td>
<td>0.2124</td>
<td>0.1461</td>
<td>0.0025</td>
<td>-0.0049</td>
<td>0.1245</td>
<td>0.0693</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.2163</td>
<td>-0.2288</td>
<td>0.3106</td>
<td>0.2534</td>
<td>0.0789</td>
<td>0.0655</td>
<td>0.1555</td>
<td>0.1158</td>
</tr>
<tr>
<td><strong>200 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>400 obs.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.1059</td>
<td>0.0594</td>
<td>0.2979</td>
<td>0.1232</td>
<td>-0.0824</td>
<td>-0.0408</td>
<td>0.2781</td>
<td>0.1106</td>
</tr>
<tr>
<td>MDSS</td>
<td>0.1004</td>
<td>0.0735</td>
<td>0.4848</td>
<td>0.3034</td>
<td>-0.0732</td>
<td>-0.0440</td>
<td>0.5289</td>
<td>0.2988</td>
</tr>
<tr>
<td>HKP</td>
<td>-0.0111</td>
<td>-0.0112</td>
<td>0.1744</td>
<td>0.1375</td>
<td>0.0035</td>
<td>0.0017</td>
<td>0.0948</td>
<td>0.0754</td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0235</td>
<td>-0.0273</td>
<td>0.1299</td>
<td>0.0918</td>
<td>0.0061</td>
<td>0.0052</td>
<td>0.0794</td>
<td>0.0535</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.1524</td>
<td>-0.1441</td>
<td>0.2283</td>
<td>0.1839</td>
<td>0.0527</td>
<td>0.0370</td>
<td>0.1058</td>
<td>0.0783</td>
</tr>
<tr>
<td><strong>200 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>400 obs.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0438</td>
<td>0.0368</td>
<td>0.1711</td>
<td>0.0935</td>
<td>-0.0235</td>
<td>-0.0126</td>
<td>0.1328</td>
<td>0.0735</td>
</tr>
<tr>
<td>MDSS</td>
<td>0.0625</td>
<td>0.0436</td>
<td>0.3541</td>
<td>0.2362</td>
<td>-0.0524</td>
<td>-0.0188</td>
<td>0.3702</td>
<td>0.2066</td>
</tr>
<tr>
<td>HKP</td>
<td>0.0015</td>
<td>-0.0040</td>
<td>0.1332</td>
<td>0.1051</td>
<td>-0.0011</td>
<td>-0.0023</td>
<td>0.0686</td>
<td>0.0545</td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0210</td>
<td>-0.0240</td>
<td>0.1048</td>
<td>0.0738</td>
<td>0.0083</td>
<td>0.0091</td>
<td>0.0525</td>
<td>0.0347</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.0934</td>
<td>-0.0776</td>
<td>0.1655</td>
<td>0.1293</td>
<td>0.0300</td>
<td>0.0276</td>
<td>0.0696</td>
<td>0.0532</td>
</tr>
</tbody>
</table>
### Table II
Simulation Results for Censored Regression Estimators
CI Censoring, Heterosked. Errors

<table>
<thead>
<tr>
<th></th>
<th>Mean Bias</th>
<th>Med. Bias</th>
<th>RMSE</th>
<th>MAD</th>
<th>Mean Bias</th>
<th>Med. Bias</th>
<th>RMSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>50 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.3083</td>
<td>0.1576</td>
<td>0.9230</td>
<td>0.2911</td>
<td>0.0156</td>
<td>-0.0072</td>
<td>0.9642</td>
<td>0.3599</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.0826</td>
<td>-0.0729</td>
<td>0.5084</td>
<td>0.3204</td>
<td>-0.0075</td>
<td>0.0167</td>
<td>0.5475</td>
<td>0.3110</td>
</tr>
<tr>
<td>HKP</td>
<td>2.2080</td>
<td>0.0252</td>
<td>31.5952</td>
<td>2.5406</td>
<td>-0.8822</td>
<td>-0.0768</td>
<td>10.4137</td>
<td>1.3964</td>
</tr>
<tr>
<td>Buckley James</td>
<td>1.9799</td>
<td>0.6611</td>
<td>5.5605</td>
<td>0.6679</td>
<td>-1.9703</td>
<td>-0.9228</td>
<td>4.3843</td>
<td>0.9757</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.1381</td>
<td>-0.1663</td>
<td>0.4222</td>
<td>0.3352</td>
<td>-0.0257</td>
<td>0.0435</td>
<td>0.6394</td>
<td>0.4794</td>
</tr>
<tr>
<td><strong>100 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.1642</td>
<td>0.0751</td>
<td>0.5633</td>
<td>0.1605</td>
<td>-0.0231</td>
<td>-0.0055</td>
<td>0.6197</td>
<td>0.2626</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.0373</td>
<td>-0.0729</td>
<td>0.5731</td>
<td>0.3084</td>
<td>-0.1208</td>
<td>-0.0639</td>
<td>0.6549</td>
<td>0.3006</td>
</tr>
<tr>
<td>HKP</td>
<td>4.8267</td>
<td>-0.0131</td>
<td>86.1774</td>
<td>5.0571</td>
<td>-0.8983</td>
<td>-0.0609</td>
<td>14.1936</td>
<td>1.2946</td>
</tr>
<tr>
<td>Buckley James</td>
<td>2.5409</td>
<td>0.8277</td>
<td>8.3142</td>
<td>0.8277</td>
<td>-2.4717</td>
<td>-1.0941</td>
<td>7.1011</td>
<td>1.0941</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.1266</td>
<td>-0.1028</td>
<td>0.3794</td>
<td>0.2621</td>
<td>0.0354</td>
<td>0.0556</td>
<td>0.5035</td>
<td>0.3685</td>
</tr>
<tr>
<td><strong>200 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0619</td>
<td>0.0456</td>
<td>0.2194</td>
<td>0.1245</td>
<td>-0.0025</td>
<td>0.0088</td>
<td>0.2847</td>
<td>0.1673</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.1048</td>
<td>-0.1172</td>
<td>0.4256</td>
<td>0.2616</td>
<td>-0.0548</td>
<td>-0.0259</td>
<td>0.4175</td>
<td>0.2311</td>
</tr>
<tr>
<td>HKP</td>
<td>0.0665</td>
<td>0.0133</td>
<td>0.4674</td>
<td>0.2301</td>
<td>-0.0520</td>
<td>0.0012</td>
<td>0.4418</td>
<td>0.3236</td>
</tr>
<tr>
<td>Buckley James</td>
<td>4.6481</td>
<td>1.2122</td>
<td>34.5088</td>
<td>1.2122</td>
<td>-3.8331</td>
<td>-1.4037</td>
<td>22.0628</td>
<td>1.4037</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.0788</td>
<td>-0.0676</td>
<td>0.2301</td>
<td>0.1713</td>
<td>0.0352</td>
<td>0.0285</td>
<td>0.3415</td>
<td>0.2543</td>
</tr>
<tr>
<td><strong>400 obs.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0201</td>
<td>0.0072</td>
<td>0.1340</td>
<td>0.0808</td>
<td>-0.0134</td>
<td>-0.0165</td>
<td>0.2055</td>
<td>0.1378</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.1576</td>
<td>-0.1818</td>
<td>0.3356</td>
<td>0.2455</td>
<td>-0.0434</td>
<td>-0.0331</td>
<td>0.3058</td>
<td>0.1881</td>
</tr>
<tr>
<td>HKP</td>
<td>-0.0038</td>
<td>-0.0206</td>
<td>0.1888</td>
<td>0.1431</td>
<td>0.0180</td>
<td>0.0158</td>
<td>0.2979</td>
<td>0.2347</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>-0.0454</td>
<td>-0.0493</td>
<td>0.1714</td>
<td>0.1323</td>
<td>0.0282</td>
<td>0.0294</td>
<td>0.2697</td>
<td>0.2017</td>
</tr>
<tr>
<td></td>
<td>α</td>
<td></td>
<td></td>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Bias</td>
<td>Med. Bias</td>
<td>RMSE</td>
<td>MAD</td>
<td>Mean Bias</td>
<td>Med. Bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.2318</td>
<td>0.1273</td>
<td>0.6043</td>
<td>0.2937</td>
<td>0.2704</td>
<td>0.0679</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDSS</td>
<td>0.2053</td>
<td>0.0007</td>
<td>0.9988</td>
<td>0.3410</td>
<td>-0.4358</td>
<td>-0.0008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HKP</td>
<td>0.1195</td>
<td>0.0230</td>
<td>0.6267</td>
<td>0.3883</td>
<td>-0.2123</td>
<td>-0.1960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0390</td>
<td>-0.0477</td>
<td>0.2693</td>
<td>0.1737</td>
<td>-0.0021</td>
<td>0.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.8236</td>
<td>0.6180</td>
<td>1.3435</td>
<td>0.8776</td>
<td>-2.2428</td>
<td>-1.9628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.1151</td>
<td>0.0609</td>
<td>0.3790</td>
<td>0.1681</td>
<td>0.1739</td>
<td>0.0960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDSS</td>
<td>0.0828</td>
<td>-0.0417</td>
<td>0.8681</td>
<td>0.2500</td>
<td>-0.2057</td>
<td>-0.0089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HKP</td>
<td>0.1042</td>
<td>0.0457</td>
<td>0.3656</td>
<td>0.2526</td>
<td>-0.1984</td>
<td>-0.1837</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0355</td>
<td>-0.0459</td>
<td>0.1779</td>
<td>0.1312</td>
<td>0.0050</td>
<td>-0.0004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.7568</td>
<td>0.7295</td>
<td>2.5556</td>
<td>0.9934</td>
<td>-2.0231</td>
<td>-1.9468</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0561</td>
<td>0.0342</td>
<td>0.2189</td>
<td>0.1207</td>
<td>0.1246</td>
<td>0.0740</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.0972</td>
<td>-0.1007</td>
<td>0.3464</td>
<td>0.1811</td>
<td>-0.0688</td>
<td>-0.1007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HKP</td>
<td>0.0799</td>
<td>0.0602</td>
<td>0.2199</td>
<td>0.1700</td>
<td>-0.1817</td>
<td>-0.1778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0187</td>
<td>-0.0265</td>
<td>0.1216</td>
<td>0.0788</td>
<td>0.0057</td>
<td>0.0032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.8143</td>
<td>0.7990</td>
<td>0.9926</td>
<td>0.8206</td>
<td>-1.9930</td>
<td>-1.9612</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0314</td>
<td>0.0215</td>
<td>0.1579</td>
<td>0.0998</td>
<td>0.0642</td>
<td>0.0391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.1228</td>
<td>-0.1250</td>
<td>0.2572</td>
<td>0.1648</td>
<td>-0.0444</td>
<td>-0.0374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HKP</td>
<td>0.0625</td>
<td>0.0590</td>
<td>0.1525</td>
<td>0.1202</td>
<td>-0.1811</td>
<td>-0.1859</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckley James</td>
<td>-0.0061</td>
<td>-0.0107</td>
<td>0.0838</td>
<td>0.0581</td>
<td>0.0016</td>
<td>0.0013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.9022</td>
<td>0.8839</td>
<td>0.9728</td>
<td>0.9029</td>
<td>-2.0317</td>
<td>-1.9860</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III**
Simulation Results for Censored Regression Estimators
CD Censoring, Homosked. Errors
<table>
<thead>
<tr>
<th></th>
<th>50 obs.</th>
<th>100 obs.</th>
<th>200 obs.</th>
<th>400 obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>β</td>
<td>α</td>
<td>β</td>
</tr>
<tr>
<td></td>
<td>Mean Bias</td>
<td>Med. Bias</td>
<td>RMSE</td>
<td>MAD</td>
</tr>
<tr>
<td>CD Censoring, Heterosked. Errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.3221</td>
<td>0.1483</td>
<td>0.8822</td>
<td>0.3214</td>
</tr>
<tr>
<td>MDSS</td>
<td>0.1539</td>
<td>0.0009</td>
<td>1.0312</td>
<td>0.4271</td>
</tr>
<tr>
<td>HKP</td>
<td>17.3471</td>
<td>0.4300</td>
<td>153.2266</td>
<td>17.4901</td>
</tr>
<tr>
<td>Buckley James</td>
<td>1.7253</td>
<td>0.3113</td>
<td>5.6005</td>
<td>0.3716</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.9745</td>
<td>0.5735</td>
<td>2.5513</td>
<td>1.0331</td>
</tr>
<tr>
<td>100 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.1452</td>
<td>0.0492</td>
<td>0.4698</td>
<td>0.2012</td>
</tr>
<tr>
<td>MDSS</td>
<td>0.0057</td>
<td>-0.1319</td>
<td>0.8533</td>
<td>0.3086</td>
</tr>
<tr>
<td>HKP</td>
<td>20.1083</td>
<td>0.3676</td>
<td>163.0567</td>
<td>20.1872</td>
</tr>
<tr>
<td>Buckley James</td>
<td>2.3266</td>
<td>0.4856</td>
<td>7.4144</td>
<td>0.4856</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.7854</td>
<td>0.6658</td>
<td>3.6963</td>
<td>1.0380</td>
</tr>
<tr>
<td>200 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0674</td>
<td>0.0442</td>
<td>0.2425</td>
<td>0.1382</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.1915</td>
<td>-0.2053</td>
<td>0.4212</td>
<td>0.2600</td>
</tr>
<tr>
<td>HKP</td>
<td>13.3405</td>
<td>0.3622</td>
<td>101.3273</td>
<td>13.3764</td>
</tr>
<tr>
<td>Buckley James</td>
<td>4.6914</td>
<td>0.6757</td>
<td>28.3437</td>
<td>0.6757</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.9670</td>
<td>0.7625</td>
<td>3.5172</td>
<td>0.9707</td>
</tr>
<tr>
<td>400 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>0.0294</td>
<td>0.0142</td>
<td>0.1827</td>
<td>0.1045</td>
</tr>
<tr>
<td>MDSS</td>
<td>-0.2374</td>
<td>-0.2469</td>
<td>0.3393</td>
<td>0.2500</td>
</tr>
<tr>
<td>HKP</td>
<td>83.9411</td>
<td>0.2946</td>
<td>663.8925</td>
<td>83.9501</td>
</tr>
<tr>
<td>Ying et al.</td>
<td>0.8527</td>
<td>0.8572</td>
<td>0.9087</td>
<td>0.8534</td>
</tr>
</tbody>
</table>
TABLE V
Empirical Study of Drug Relapse Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.1187)</td>
<td>(0.1110)</td>
<td>(0.1070)</td>
<td>(0.1457)</td>
<td>(0.3533)</td>
<td>(0.1310)</td>
<td>(0.1797)</td>
</tr>
<tr>
<td>SITE</td>
<td>-0.4866</td>
<td>-0.5254</td>
<td>-0.4926</td>
<td>-0.4952</td>
<td>-0.3547</td>
<td>-0.5559</td>
<td>-0.2578</td>
</tr>
<tr>
<td></td>
<td>(0.1040)</td>
<td>(0.0938)</td>
<td>(0.0959)</td>
<td>(0.0944)</td>
<td>(0.1352)</td>
<td>(0.1016)</td>
<td>(0.0626)</td>
</tr>
<tr>
<td>IV</td>
<td>-0.3673</td>
<td>-0.1835</td>
<td>-0.1769</td>
<td>-0.1683</td>
<td>-0.1192</td>
<td>-0.1277</td>
<td>-0.0919</td>
</tr>
<tr>
<td></td>
<td>(0.0985)</td>
<td>(0.0862)</td>
<td>(0.0876)</td>
<td>(0.0896)</td>
<td>(0.0943)</td>
<td>(0.0908)</td>
<td>(0.0459)</td>
</tr>
<tr>
<td>NDT</td>
<td>-0.0243</td>
<td>-0.0209</td>
<td>-0.0119</td>
<td>-0.0122</td>
<td>-0.0164</td>
<td>-0.0186</td>
<td>-0.0393</td>
</tr>
<tr>
<td></td>
<td>(0.0078)</td>
<td>(0.007)</td>
<td>(0.0075)</td>
<td>(0.0062)</td>
<td>(0.0065)</td>
<td>(0.0071)</td>
<td>(0.0072)</td>
</tr>
<tr>
<td>RACE</td>
<td>0.2964</td>
<td>0.3288</td>
<td>0.3393</td>
<td>0.3292</td>
<td>0.4411</td>
<td>0.3413</td>
<td>0.4050</td>
</tr>
<tr>
<td></td>
<td>(0.1073)</td>
<td>(0.0952)</td>
<td>(0.0948)</td>
<td>(0.1131)</td>
<td>(0.1147)</td>
<td>(0.0990)</td>
<td>(0.0408)</td>
</tr>
<tr>
<td>TREAT</td>
<td>0.4215</td>
<td>0.6114</td>
<td>0.6243</td>
<td>0.6075</td>
<td>0.7605</td>
<td>0.6120</td>
<td>0.7952</td>
</tr>
<tr>
<td></td>
<td>(0.0905)</td>
<td>(0.0839)</td>
<td>(0.0820)</td>
<td>(0.0919)</td>
<td>(0.1451)</td>
<td>(0.0847)</td>
<td>(0.0988)</td>
</tr>
<tr>
<td>FRAC</td>
<td>1.1543</td>
<td>1.468</td>
<td>1.2488</td>
<td>1.2357</td>
<td>1.6790</td>
<td>1.5732</td>
<td>1.5412</td>
</tr>
<tr>
<td></td>
<td>(0.0990)</td>
<td>(0.0839)</td>
<td>(0.0798)</td>
<td>(0.0938)</td>
<td>(0.3038)</td>
<td>(0.0869)</td>
<td>(0.1502)</td>
</tr>
</tbody>
</table>
### Table VI

**Empirical Study of Selective Compliance using Drug Relapse Data**

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