Quantile Regression Under Random Censoring

Bo Honoré
Princeton University
Shakeeb Khan*
University of Rochester
James L. Powell
University of California at Berkeley

First Version: October 1993
This Version: August 1999

Abstract

Censored regression models have received a great deal of attention in both the theoretical and applied econometric literature. Most of the existing estimation procedures for either cross sectional or panel data models are designed only for models with fixed censoring. In this paper, a new procedure for adapting these estimators designed for fixed censoring to models with random censorship is proposed. This procedure is then applied to the CLAD and quantile estimators of Powell(1984,1986a) to obtain an estimator of the regression coefficients under a mild conditional quantile restriction on the error term that is applicable to samples exhibiting fixed or random censoring. The resulting estimator is shown to have desirable asymptotic properties, and performs well in a small scale simulation study.

JEL Classification: C24, C14, C13.

Key Words: censored quantile regression, random censoring, Kaplan-Meier product limit estimator, accelerated failure time model.

---

*Corresponding author. Dept. Economics, University of Rochester, Rochester, NY 14627; e-mail: skhan@troi.cc.rochester.edu. We are grateful to Jambe and Rodchenko for helpful comments on an earlier draft.

1 Introduction

Over the past decade, the censored regression model, known to economists as the Tobit model (after Tobin 1958), has been the object of much attention in the econometric literature on semiparametric estimation. Relaxing the traditional parametric restrictions on the form of the distribution of the underlying error terms, a number of consistent estimators have been proposed which require only weak conditions on these distributions, including: constant conditional quantiles (Powell(1984,1986a); Nawata (1990); Newey and Powell(1990), conditional symmetry (Powell(1986b), Lee(1993a,b), Newey(1991)), and independence of the errors and regressors (Duncan(1986); Fernandez(1986); Honoré and Powell(1993); Horowitz(1986, 1988); and Moon(1989)). These proposed estimators all exploit an assumption that the censoring values for the dependent variable are known for all observations, even those that are not censored; while the typical estimator is constructed under the presumption that the dependent variable is censored to the left at zero, it is generally straightforward to modify it for either right or left censored data (or both) with variable censoring values. Hereafter, in a loose analogy to panel data modelling, we refer to such models as *fixed* censoring models, since the censoring values, though possibly variable, may not be distributed independently from the regressors.

A parallel literature in the statistics and biometrics literature has been concerned with estimation of the parameters of a related model, the regression model with *random* censoring. In this model the dependent variable typically represents the logarithm of a survival time (in which case the regression model corresponds to an *accelerated failure time* duration model), which is right-censored at varying censoring points which are observed only when the observation is censored. In addition, the censoring times are generally (but not always) assumed to be independently distributed of the regressors and error terms. Studies which propose semiparametric methods under random (right) censorship include Miller (1976), Prentice (1978), Buckley and James (1979), Koul, Suslara, and Van Ryzin (1981), Leurgans (1987), and Ritov (1990), among others. These estimation methods typically impose an assumption of independence of the error terms and covariates; those that do not impose independence instead require strong conditions on the censoring distribution which generally rule out censoring at a constant value, as is typical in econometrics.

In this paper we describe a method for adapting estimators proposed for fixed censoring to sampling with random right censorship. We apply this method to the censored least absolute deviations and quantile estimators of Powell (1984, 1986a) to obtain an estimator of the regression coefficients which will be consistent under a relatively-weak quantile restriction on the error terms, and which is equally applicable to samples with constant or random censoring. The following section describes this estimation approach, and compares the modified form of the censored regression quantile estimator to other quantile-based estimators for random censoring that have appeared in the statistics literature. Section 3 gives sufficient conditions to ensure the root n-consistency and asymptotic normality of the proposed estimator, and section 4 analyzes its performance using a simulation study and an empirical example. The final section discusses application of the general estimation method to other censored regression estimators in the econometric literature, and considers whether the assumption of independence of the censoring times and covariates could be relaxed. Proofs of the large-sample results of section 3 are available in a mathematical appendix.

2 The Model and Estimation Method

The object of estimation is the p-dimensional vector of regression coefficients β_0 in a linear latent variable model

$$y_i^* = x_i'\beta_0 + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where y_i^* is the (uncensored and scalar) dependent variable of interest, x_i is an observable p-vector of covariates, and ε_i is an unobserved error term. With right censorship, the latent variable y_i^* is observed only when it is less than some scalar censoring variable c_i; that is, the observed dependent variable y_i is

$$y_i = \min\{y_i^*, c_i\} = \min\{x_i'\beta_0 + \varepsilon_i, c_i\}. \tag{2.2}$$

In a random sample with fixed censoring, n independently-distributed observations on the triple (y_i, c_i, x_i) are assumed to be available; with random censoring, the observations are of the form (y_i, d_i, x_i), where d_i is a binary variable indicating whether the dependent variable is uncensored:

$$d_i = 1\{y_i^* < c_i\} = 1\{x_i'\beta_0 + \varepsilon_i < c_i\}, \tag{2.3}$$

for "$1\{A\}$" the indicator function for the set A.

For samples with fixed censoring, the estimators of β_0 cited in the preceding section often are defined as solutions to minimization problems and/or estimating equations constructed using sample averages of functions of the observable data and unknown parameters, i.e.,

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \rho(y_i, c_i, x_i, \beta) \tag{2.4}$$

or

$$0 \cong \frac{1}{n} \sum_{i=1}^{n} \psi(y_i, c_i, x_i, \hat{\beta}) \tag{2.5}$$

for certain functions $\rho(\cdot)$ or $\psi(\cdot)$. Of course, some estimators involve more complicated minimands / estimating equations, defined using higher-order U-statistics or involving preliminary (nonparametric) estimators of unknown functions, but the analysis of their large sample behavior, though more difficult, follows the same lines as in this simple case. Consistency of $\hat{\beta}$ is demonstrated after imposing appropriate conditions on the error terms, covariates, and censoring values; one important step in the proof is to show that the true parameter value β_0 is a unique solution to the population versions of the minimization problem or estimating equations,

$$\beta_0 = \arg\min_{\beta} E[\rho(y_i, c_i, x_i, \beta)] \tag{2.6}$$

or

$$0 = E[\psi(y_i, c_i, x_i, \beta)] \qquad \text{iff} \qquad \beta = \beta_0. \tag{2.7}$$

Given such an identification condition, application of a uniform law of large numbers to the sample average defining $\hat{\beta}$ ensures its consistency.

Under random censorship, it is no longer possible to define an estimator of β_0 in the same fashion as above, since the censoring variables $\{c_i\}$ are not known for all i. However, if the censoring variables $\{c_i\}$ are assumed to be independent of $\{(y_i, x_i)\}$, and if the marginal c.d.f. $G(t) \equiv \Pr\{c_i \leq t\}$ of the censoring values were known, a simple modification of the estimation approach above would replace the functions $\rho(y_i, x_i, c_i, \beta)$ or $\psi(y_i, x_i, c_i, \beta)$ by their conditional expectations given the observable variables (y_i, d_i, x_i). That is, an M-estimator of β_0 corresponding to the foregoing minimization problem would be

$$
\begin{aligned}
\hat{\beta} &= \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} E[\rho(y_i, c_i, x_i, \beta) \mid (y_i, d_i, x_i)] \qquad (2.8) \\
&= \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left\{ (1 - d_i) \cdot \rho(y_i, y_i, x_i, \beta) + d_i \cdot [S(y_i)]^{-1} \cdot \int 1(y_i < c) \rho(y_i, c, x_i, \beta) dG(c) \right\},
\end{aligned}
$$

where $S(t) \equiv 1 - G(c)$ is the survivor function for the censoring value c_i. Similarly, $\hat{\beta}$ might be defined as solutions to estimating equations of the form

$$
0 \cong \frac{1}{n} \sum_{i=1}^{n} \left\{ (1 - d_i) \cdot \psi(y_i, y_i, x_i, \hat{\beta}) + d_i \cdot [S(y_i)]^{-1} \cdot \int 1(y_i < c) \psi(y_i, c, x_i, \hat{\beta}) dG(c) \right\}. \qquad (2.9)
$$

By iterated expectations, the population analogues to the sample averages defining $\hat{\beta}$ will be the same moment functions, $E[\rho(y_i, c_i, x_i, \beta)]$ or $E[\psi(y_i, c_i, x_i, \beta)]$, as appear in the fixed censorship case, so the same identification conditions imposed for fixed censoring will apply under random censoring.

Unfortunately, when the censoring values $\{c_i\}$ have a non-degenerate distribution it is unlikely that the censoring distribution function $G(t)$ will be known *a priori*. Nevertheless, because of the assumed independence of the censoring value c_i and the latent variable y_i^*, this distribution function $G(t)$ can be consistently estimated using the Kaplan-Meier product limit estimator (Kaplan and Meier, 1958); this estimator $\hat{G}(t)$ uses only the pairs $\{(y_i, d_i)\}$ of dependent and indicator variables, and does not involve the covariates $\{x_i\}$ or parameter vector β. By substitution of the Kaplan-Meier estimator $\hat{G}(t)$ and survivor function $\hat{S}(t) = 1 - \hat{G}(t)$ into the previous minimization problem or estimating equations, feasible estimators of β_0 can be constructed, and consistency will follow from a demonstration of uniform convergence of these sample moment functions to their limiting values.

The estimation approach here is similar in spirit to that adopted by Buckley and James (1979), which adapted the "*EM* algorithm" (Dempster, Laird, and Rubin 1977) for maximization of a parametric censored-data likelihood to the semiparametric setting with unknown error distribution. However, the Buckley-James estimator treats the latent dependent variable y_i^* as "missing data" when the observed dependent variable is uncensored (using the Kaplan-Meier estimator for the error distribution, applied to residuals $\hat{\varepsilon} \equiv y - x'\hat{\beta}$ and their censoring points $u - x'\hat{\beta}$, to estimate the conditional distribution of y_i^* given $d_i = 0$); in contrast, the present approach views the censoring value c_i as "missing" when the latent dependent variable is uncensored. While the Buckley-James estimator does not require that the censoring values be independent of the regressors, it does impose that requirement for the error distribution; in contrast, the present approach assumes independence of the censoring points and regressors, but may permit dependence of, say, the scale of the errors on the covariates.

To apply this general approach to a specific estimation problem, we consider the restriction of a constant conditional π'th quantile on the distribution of the errors. That is, maintaining the assumption of independence of $\{c_i\}$ and $\{(\varepsilon_i, x_i)\}$, we impose the additional restriction that the conditional distribution of the error terms ε_i given the covariates x_i satisfies

$$\Pr\{\varepsilon_i \leq 0 \mid x_i\} = \pi \tag{2.10}$$

for some known value of π in the interior of the unit interval. Under this condition, the conditional π'th quantile of the dependent variable y_i given x_i and c_i is equal to $\min\{x'\beta_0, c_i\}$, as noted by Powell (1984, 1986a) and Newey and Powell (1990); that is,

$$\Pr\{y_i \leq \min\{x_i'\beta_0, c_i\} \mid x_i\} \geq \pi \qquad \text{and} \qquad \Pr\{y_i \leq \min\{x_i'\beta_0, c_i\} \mid x_i\} \geq 1 - \pi. \tag{2.11}$$

Under fixed censorship, a quantile estimator of β_0 under this restriction was defined by Newey and Powell (1990) as

$$\hat{\beta} = \arg\min_\beta \frac{1}{n} \sum_{i=1}^{n} \rho_\pi(y_i - \min\{x_i'\beta, c_i\}), \tag{2.12}$$

where

$$\rho_\pi(u) \equiv [\pi - 1\{u < 0\}] \cdot u; \tag{2.13}$$

this estimator is the censored-data analogue to the regression quantile estimator proposed by Koenker and Bassett (1978) for the linear model. Under regularity conditions, it was shown that the estimator $\hat{\beta}$ solves a set of estimating equations obtained as approximate first-order conditions from this minimization problem:

$$o_p(n^{-1/2}) = \frac{1}{n} \sum_{i=1}^{n} [\pi - 1\{y_i \leq x_i'\hat{\beta}\}] \cdot 1\{x_i'\hat{\beta} < c_i\} \cdot x_i. \tag{2.14}$$

For the special case $\pi = 1/2$, corresponding to a linear model for the conditional median of y_i^* given x_i, an equivalent representation would replace "ρ_π" with an absolute value function in the minimization problem, and "$[\pi - 1\{y_i \leq x_i'\beta\}]$" with "$\text{sign}\{y_i - x_i'\beta\}$" in the estimating equations.

To adapt the quantile estimator for fixed censoring to a sample subject to random censorship, then, we define the estimator as

$$\begin{aligned}\hat{\beta} &= \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \hat{E}[\rho_\pi(y_i - \min\{x_i'\beta, c_i\}) \mid (y_i, d_i, x_i)] \qquad (2.15)\\ &= \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \Big\{(1-d_i)\cdot\rho_\pi(y_i - \min\{x_i'\beta, y_i\})\\ &\quad + d_i \cdot [\hat{S}(y_i)]^{-1} \cdot \int 1(y_i < c)\rho_\pi(y_i - \min\{x_i'\beta, c\})d\hat{G}(c)\Big\},\end{aligned}$$

where "$\hat{E}[\cdot]$" denotes an expectation calculated using the product-limit estimator of $G(t)$. For this minimization problem, the estimating equations obtained from the approximate first-order condition take a particularly simple form:

$$o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n}\Big(\pi \cdot 1\{y_i > x_i'\hat{\beta}\} - (1-\pi)\cdot 1\{y_i \leq x_i'\hat{\beta}\}\cdot d_i \cdot \hat{S}(x_i'\beta)/\hat{S}(y_i)\Big)\cdot x_i. \qquad (2.16)$$

To verify that the limiting form of these estimating equations (replacing the sample average and estimated survivor functions with their population analogues) has a solution at the true value β_0, note that

$$1\{y_i > x_i'\beta_0\} \equiv 1\{\varepsilon_i > 0\}\cdot 1\{c_i > x_i'\beta_0\}$$

so that

$$E[1\{y_i > x_i'\beta_0\} \mid x_i] = \Pr\{\varepsilon_i > 0 \mid x_i\}\cdot S(x_i'\beta_0) = (1-\pi)\cdot S(x_i'\beta_0);$$

also,

$$1\{y_i \leq x_i'\beta_0\}\cdot d_i \cdot S(x_i'\beta_0)/S(y_i) \equiv 1\{\varepsilon_i \leq 0\}\cdot 1\{y_i^* < c_i\}\cdot S(x_i'\beta_0)/S(y_i^*),$$

implying

$$E\left[1\{y_i \leq x_i'\beta_0\}\cdot d_i \cdot S(x_i'\beta_0)/S(y_i) \mid x_i, \varepsilon_i\right] = 1\{\varepsilon_i \leq 0\}\cdot S(y_i^*)\cdot S(x_i'\beta_0)/S(y_i^*)$$

and thus

$$E\left[1\{y_i \leq x_i'\beta_0\} \cdot d_i \cdot S(x_i'\beta_0)/S(y_i) \mid x_i\right] = \Pr\{\varepsilon_i \leq 0 \mid x_i\} \cdot S(x_i'\beta_0) = \pi \cdot S(x_i'\beta_0).$$

Therefore, the limiting estimating equations hold when evaluated at the true value β_0 :

$$\begin{aligned} & E\left[\left(\pi \cdot 1\{y_i > x_i'\beta\} - (1-\pi) \cdot 1\{y_i \leq x_i'\beta\} \cdot d_i \cdot \hat{S}(x_i'\beta)/\hat{S}(y_i)\right) \cdot x_i\right] \\ = \; & E[(\pi \cdot (1-\pi) \cdot S(x_i'\beta_0) - (1-\pi) \cdot \pi \cdot S(x_i'\beta_0)) \cdot x_i] = 0. \end{aligned} \tag{2.17}$$

Nevertheless, as noted by Powell (1984, 1986a), it is important that the estimator be defined as the minimizer of the quantile objective function rather than the solution to these estimating equations, since multiple inconsistent roots to these equations may exist.

Two other estimators under random censorship which exploit only a quantile restriction have previously been proposed; these approaches require stronger restrictions on the support of the censoring distribution $G(t)$ than are needed for the present estimator. For example, an extension of the approach of Koul, Suslara, and Van Ryzin (1981) to quantile regression was proposed by ????, which defined the estimator of the regression coefficients as

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} d_i \cdot [\hat{S}(y_i)]^{-1} \cdot \rho_\pi(y_i - x_i'\beta), \tag{2.18}$$

which can equivalently be written as the solution to the estimating equations

$$o_p(n^{-1/2}) = \frac{1}{n} \sum_{i=1}^{n} d_i \cdot [\hat{S}(y_i)]^{-1} \cdot [\pi - 1\{y_i \leq x_i'\hat{\beta}\}] \cdot x_i; \tag{2.19}$$

this estimator exploits the fact that

$$\begin{aligned} & E\left[d_i \cdot [S(y_i)]^{-1} \cdot [\pi - 1\{y_i \leq x_i'\beta_0\}] \cdot x_i\right] \\ = \; & E\left[1\{y_i^* < c_i) \cdot [S(y_i^*)]^{-1} \cdot [\pi - 1\{\varepsilon_i \leq 0\}] \cdot x_i\right] \\ = \; & E\left[[\pi - 1\{\varepsilon_i \leq 0\}] \cdot x_i\right] \\ = \; & 0 \end{aligned} \tag{2.20}$$

provided $S(y_i^*) > 0$ with probability one. Since $\hat{S}(t) = S(t) = 1\{t < c_0\}$ when the censoring points $c_i = c_0$ with probability one, this estimation approach is not applicable for fixed (and constant) censoring except in the special cases $\Pr\{y_i \leq c_\circ\} \equiv 1$ (i.e., no censored observations). Also, as noted by Halpern and Miller (1982), this estimator may be sensitive to the particular realizations of the dependent variable y_i(and corresponding regressors x_i) which are large and uncensored, since the estimated survivor function for such observations will be close to zero and imprecisely measured; however, this robustness problem may be more pronounced for the original estimator proposed by Koul, *et al.*, which is based upon squared error loss, than for its quantile variant.

More recently, Ying, Jung, and Wei (1991) proposed a quantile estimator for β_0 under the restriction $\Pr\{\epsilon \leq 0 \mid x\} \equiv \pi \in (0,1)$ using the implied relation

$$\begin{aligned} \Pr\{y_i &> x_i'\beta_0 \mid x_i\} = \Pr\{x_i'\beta_0 < c_i \quad \text{and} \quad \varepsilon_i > 0| \, x_i\} && (2.21) \\ &= \Pr\{x_i'\beta_0 < c_i| \, x_i\} \cdot \Pr\{\varepsilon_i > 0| \, x_i\} \\ &= S(x'\beta_0) \cdot (1-\pi), && (2.22) \end{aligned}$$

which yields an estimator $\hat{\beta}$ as a solution to estimating equations of the form

$$0 \cong \frac{1}{N}\sum_{i=1}^{N}\left[[\hat{S}(x_i'\hat{\beta})]^{-1} \cdot 1\{y_i > x_i'\hat{\beta}\} - (1-\pi)\right] \cdot x_i. \tag{2.23}$$

Like the ????? estimator, this estimator will be well-defined and consistent only when $\hat{S}(x_i'\hat{\beta})$ and $S(x_i'\beta_0)$ are strictly positive with probability one, which would require $\Pr\{x\prime\beta_0 \leq c_0\} \equiv 1$ when the censoring values have a degenerate distribution. In contrast, the present approach is equally amenable to constant or random censoring; indeed, if the censoring points are degenerate, so that $S(t) = 1\{t < c_0\}$, then this estimator will be identical to the censored quantile estimator proposed by Powell (1986a) for samples consisting of at least one censored observation, since $S(t) = \hat{S}(t)$ in this case.

3 Large Sample Behavior of the Quantile Estimator

In order to demonstrate the (root$-n$) consistency and asymptotic normality of the randomly-censored quantile regression estimator proposed above, it will be necessary to augment the regularity conditions imposed for its fixed-censoring counterpart to ensure, for example, that the Kaplan-Meier estimator of the censoring survivor function is sufficiently precise. Rather than searching for the most general conditions on the errors, covariates, and censoring times, we will impose stronger conditions (like compact support of the regressors) which will be straightforward to verify and simplify the derivations of the asymptotic theory for the estimator.

We rewrite the estimator defined in (2.15) here as

$$\hat{\beta} \equiv \arg\min_{\beta \in B} R_n(\beta; \hat{S}), \tag{3.1}$$

where

$$\begin{aligned} R_n(\beta; \hat{S}) \quad \equiv \quad & \frac{1}{n}\sum_{i=1}^{n}\{(1-d_i)\cdot\rho_\pi(y_i - \min\{x_i'\beta, y_i\}) \\ & + d_i \cdot [\hat{S}(y_i)]^{-1} \cdot \int 1(y_i < c)\rho_\pi(y_i - \min\{x_i'\beta, c\})d\hat{G}(c)\Big\} \end{aligned} \tag{3.2}$$

and B is the space of possible values of the parameter vector β_0. In Newey and Powell (1990), a number of regularity conditions were imposed for the analysis of the estimator with fixed censoring, defined in (2.12) above. The following assumptions are a superset of the conditions imposed in that paper to ensure root$-n$ consistency and asymptotic normality in that case.

Assumption P: The true parameter vector β_0 is an interior point of the parameter space B, which is compact.

Assumption M: The observations $\{(y_i, d_i, x_i), i = 1, \ldots, n\}$ are a random sample for which y_i and d_i are generated according to (2.2) and (2.3), for some random variables ε_i, x_i, and c_i satisfying the remaining conditions below.

Assumption E: The error terms $\{\varepsilon_i\}$ are absolutely continuously distributed with conditional density function $f(\epsilon \mid x)$ given the regressors $x_i = x$ which has π'th quantile 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. That is,

$$\int 1\{\epsilon \leq 0\} \cdot f(\epsilon \mid x)d\epsilon = \pi,$$

and for some positive constants ϕ_0, Φ_0, and η_0,

$$f(\epsilon \mid x) \leq \phi_0, \quad |f(\epsilon_1 \mid x) - f(\epsilon_2 \mid x)| \leq \Phi_0 \cdot \mid \epsilon_1 - \epsilon_2 \mid, \quad \text{and}$$

$$f(\epsilon \mid x) \geq \eta_0 \quad \text{if} \quad \mid \epsilon \mid \leq \eta_0.$$

Assumption R: The regressors $\{x_i\}$ have compact support, i.e., $\Pr\{\|x_i\| \leq \chi_0\} = 1$ for some constant χ_0.

Assumption C: The censoring values $\{c_i\}$ are distributed independently of $\{(\varepsilon_i, x_i')\}$, with c.d.f. $G(t) \equiv \Pr\{c_i \leq t\}$ which has $G(\tau_0) = \Pr\{c_i \leq \tau_0) = 1$ and $G(\tau_0) - G(\tau_0-) = \Pr\{c_i = \tau_0\} > 0$.

Assumption RC: The regressors $\{x_i\}$ and censoring values $\{c_i\}$ satisfy

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

(ii) $E[1\{c_i - x_i'\beta > \eta_0\} \cdot x_i x_i'] = E\left[S(x_i'\beta + \eta_0\} \cdot x_i x_i'\right]$ is nonsingular for some $\eta_0 > 0$.

Many of these conditions have been discussed in Powell (1984, 1986a) and Newey and Powell (1990), so we will only briefly motivate them here. The compactness condition on the parameter space is needed because the minimand $R_n(\beta)$ is not a convex function of β, and β_0 must be an interior point to guarantee validity of the usual Taylor's series expansions. The random sampling assumption is imposed mostly for convenience, and can be relaxed for the regressors $\{x_i\}$, although random sampling of the censoring values $\{c_i\}$ is essential for consistency of the Kaplan-Meier estimator of the censoring c.d.f. The boundedness and Lipschitz continuity of the conditional error density simplify the demonstration of convergence of certain remainder terms to zero; the lower bound on the conditional density near zero ensures uniqueness of the π'th quantile of the error distribution, and can be interpreted as a "bounded heteroskedasticity" requirement (using the inverse of the conditional density at zero as the relevant scale parameter for the conditional distribution). The bounded support of the regressors ensures boundedness (and thus existence of all moments) for terms appearing in $R_n(\beta; \hat{S})$ and its subgradient; this condition can be enforced without violating the remaining assumptions by truncating any observations with x_i outside a bounded set. The upper bound on the censoring values, and the positive mass on the upper boundary of their support, guarantee that terms of the form $d_i/\hat{S}(y_i)$ will be well-behaved in large samples, since then $S(y_i)$ will be bounded away from zero for all observations with $d_i = 1$; like the boundedness condition on the regressors, this condition on the censoring distribution can be ensured by artificially censoring all observations at some point τ_0 in the observed support of the $\{y_i\}$. Assumption $RC(i)$ rules out ties between the censoring values and the regression function, just as the continuity of the error distribution rules out ties between the censoring values and the latent variable y_i^*. Finally, condition $RC(ii)$ is the key identification condition which ensures that $p\lim R_n(\beta; \hat{S}) - R_n(\beta_0; \hat{S}) > 0$ when $\beta \neq \beta_0$; it is essentially a full rank condition for the cross product of the regressors corresponding to observations in which the conditional median of the latent variable y_i^* is uncensored, i.e., $x_i'\beta_0 < c_i$.

Under these conditions, it is straightforward to establish the strong consistency of the estimator $\hat{\beta}$ for $\beta_\circ$, using a direct modification of the arguments in Powell (1984, 1986a):

Theorem 3.1 *Under conditions* P, M, E, R, C, *and* RC, *the estimator* $\hat{\beta}$ *defined in (3.1) is strongly consistent, i.e.,* $\hat{\beta} \to \beta_0$ *with probability one.*

Demonstration of the root$-n$ consistency and asymptotic normality of $\hat{\beta}$ is more delicate, since it involves the asymptotic distribution associated with the empirical process $\hat{S}(t)$, the Kaplan-Meier estimator of the censoring survivor function. If $S(t)$ were known, so that an estimator of β_0 could be defined as

$$\tilde{\beta} \equiv \arg\min_{\beta \in B} R_n(\beta; S), \tag{3.3}$$

it would be relatively simple to derive the asymptotically-normal distribution of $\tilde{\beta}$. Let

$$\psi_i(\beta, S) \equiv \big(\pi \cdot 1\{y_i > x_i'\beta\} - (1-\pi) \cdot 1\{y_i \leq x_i'\beta\} \cdot d_i \cdot S(x_i'\beta)/S(y_i)\big) \cdot x_i \tag{3.4}$$

and

$$M_0 \equiv M(\beta_0, S) \equiv E[f(0 \mid x_i) \cdot S(x_i'\beta_0) \cdot x_i x_i'] = E[f(0 \mid x_i') \cdot 1\{x_i'\beta_0 < c_i\} \cdot x_i x_i']; \tag{3.5}$$

then the same arguments used in Powell (1984) could be used to show that, under the conditions imposed above, the estimator $\tilde{\beta}$ would solve the estimating equations

$$o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n} \psi_i(\tilde{\beta}, S), \tag{3.6}$$

and would have asymptotic distribution given by

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, M_0^{-1} V_0 M_0^{-1}), \tag{3.7}$$

where

$$V_0 \equiv E[\psi_i(\beta_0, S) \cdot \psi_i(\beta_{0\circ}, S)']. \tag{3.8}$$

However, the feasible estimator $\hat{\beta}$ solves the estimating equations

$$o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\beta},\hat{S}), \tag{3.9}$$

and since $\sqrt{n}(\hat{S}(t)-S(t)) = O_p(1)$, a "correction term" for the preliminary estimation of the censoring survivor function $S(t)$ is needed for the asymptotic distribution of $\hat{\beta}$. To obtain the form of this correction term, we first define the following term:

$$\mathbb{X}_n(t) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[H(y_i)]^{-1}\cdot 1\{y_i < t\}\cdot(1-d_i) - \int_{-\infty}^{t}[H(s)]^{-1}\cdot 1\{y_i \geq s\}\cdot d\Lambda(s),$$

for $H(t) \equiv Pr\{y_i > t\}$, the survivor function for y_i, and

$$\Lambda(t) \equiv \int_{-\infty}^{t}[S(s)]^{-1}dG(s),$$

the cumulative hazard function for c_i. The correction term for the estimation of the survivor function $S(t)$ in the construction of $\hat{\beta}$ involves the integral of $\mathbb{X}(t)$, with respect to the measure

$$q(s) \equiv \lim Q_n(s) \qquad \text{a.s.}, \tag{3.10}$$

where

$$Q_n(t) \equiv \frac{1}{n}\sum_{i=1}^{n}(1\{y_i \leq \min\{t, x_i'\beta_0\}\} + 1\{\max\{y_i,t\} < x_i'\beta\}\cdot S(x_i'\beta)/S(y_i))\cdot d_i\cdot x_i. \tag{3.11}$$

Defining

$$\begin{aligned}\xi_i &\equiv \xi_i(\beta_0, S, H, \Lambda, \pi) \qquad (3.12)\\ &\equiv (1-\pi)\cdot\int_{-\infty}^{\infty}\left([H(y_i)]^{-1}\cdot 1\{y_i < t\}\cdot(1-d_i) - \int_{-\infty}^{t}[H(s)]^{-1}\cdot 1\{y_i \geq s\}\cdot d\Lambda(s)\right)dq(s)\end{aligned}$$

the asymptotic distribution of $\hat{\beta}$ depends on ξ_i, as follows:

Theorem 3.2 *Under Assumptions P, M, E, R, C, RC, the estimator $\hat{\beta}$ satisfies the asymptotic linearity condition*

$$\sqrt{n}(\hat{\beta}-\beta_0) = M_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_i(\beta_0, S) + \xi_i(\beta_0, S, H, \Lambda, \pi)] + o_p(1)$$

and is asymptotically normal,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, M_0^{-1} V_1 M_0^{-1}),$$

for

$$V_1 \equiv E[(\psi_i + \xi_i) \cdot (\psi_i + \xi_i)'].$$

In order to use the asymptotic normality result of Theorem 4.2 to form asymptotic confidence regions and hypothesis tests, a consistent estimator of the asymptotic covariance matrix of $\hat{\beta}$ is needed. Estimation of each of the matrices M_0 and V_1 poses technical problems, the former because of its dependence on the error density (and thus requiring nonparametric estimation techniques), and the latter due to the complicated form of the correction term for preliminary estimation of the survivor function $\hat{S}(t)$. For estimation of the Hessian matrix M_0, the method proposed by Powell (1984), which replaces the unknown density with a (uniform) kernel term in a sample analogue to the definition of M_0 in (3.5), can be easily adapted to the present case. Another means to consistently estimate M_0 was proposed by Pakes and Pollard (1989), who suggested that the Hessian be estimated from a numerical derivative of the function appearing in the estimating equations,

$$\Psi_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^{n} \psi_i(\beta, \hat{S}), \tag{3.13}$$

about the point $\beta = \hat{\beta}$; they note that this estimator will be consistent if the perturbations used to construct the numerical derivative are of larger order than $\sqrt{n}$, the rate of convergence of $\hat{\beta}$. Consistent estimation of an asymptotic covariance matrix analoguous to V_1(but with a different definition of $\psi(\cdot)$ and $Q_n(\cdot)$) was considered by Ying, Jung, and Wei (1991), who proposed an estimator of the form

$$\hat{V} \equiv \frac{1}{n} \sum_{i=1}^{n} (\hat{\psi}_i + \hat{\xi}_i) \cdot (\hat{\psi}_i + \hat{\xi}_i)', \tag{3.14}$$

where, in this setting, $\hat{\psi}_i$ and $\hat{\xi}_i$ would be sample analogues of ψ_i and ξ_i. That is,

$$\hat{\psi}_i \equiv \Big(\pi \cdot 1\{y_i > x_i'\hat{\beta}\} - (1 - \pi) \cdot 1\{y_i \leq x_i'\hat{\beta}\} \cdot d_i \cdot \hat{S}(x_i'\beta)/\hat{S}(y_i)\Big) \cdot x_i \tag{3.15}$$

and

$$\hat{\xi}_i \equiv n^{-1/2} \cdot (1-\pi) \cdot \int_{-\infty}^{\infty} \hat{\nu}_i(s) dQ_n(s), \tag{3.16}$$

for $Q_n(s)$ defined in (3.11),

$$\hat{\nu}_i(t) \equiv \left([\hat{H}(y_i)]^{-1} \cdot 1\{y_i < t\} \cdot (1-d_i) - \int_{-\infty}^{t} [\hat{H}(s)]^{-1} \cdot 1\{y_i \geq s\} \cdot d\hat{\Lambda}(s) \right) \tag{3.17}$$

and $\hat{H}(s)$ and $\hat{\Lambda}(s)$ are the sample analogues of the survivor function for y_i and the cumulative hazard for c_i,

$$\hat{H}(s) \equiv \frac{1}{n} \sum_{i=1}^{n} 1\{y_i > s\} \tag{3.18}$$

and

$$\hat{\Lambda}(s) \equiv [\hat{H}(s)]^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (1-d_i) \cdot 1\{y_i \leq s\}. \tag{3.19}$$

Verification of consistency of $\hat{V}$ of (3.14) would require a tedious verification that $\max_i \| \hat{\psi}_i - \psi_i + \hat{\xi}_i - \xi_i \| = o_p(1)$, and then routine application of a law of large numbers.

A simpler alternative to direct construction of a sample analogue to $M_0^{-1} V_1 M_0^{-1}$, which we adopt in the next section, is to use bootstrap methods to assess the sampling variability of $\hat{\beta}$. Specifically, a prespecified number R of random samples of size n, drawn from the empirical distribution of the data set $\{(y_i, d_i, x_i).\ i = 1, \ldots, n\}$, can be used to calculate R simulated replications of $\hat{\beta}$, and the empirical distribution of these replicated values can be used as an estimator of the sampling distribution of $\hat{\beta}$. For the fixed censoring quantile estimator, this bootstrap estimator of the asymptotic distribution was shown to be consistent by Hahn (1993), and the simulation study by Buchinsky (1991) shows that this bootstrap method works well for an empirically-based design. While the theoretical results of Hahn (1993) do not directly apply to the randomly-censored regression quantile estimator considered here, we think it likely that consistency of the bootstrap c.d.f. will hold under the conditions imposed in this section, and, further, that the bootstrap method may give a better approximation to the finite-sample distribution of test statistics involving $\hat{\beta}$ than an asymptotic normal approximation using the covariance matrix estimator described above.

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: results of a small-scale Monte Carlo study are reported, and the method is applied to a much-studied empirical example, the Stanford heart transplant data.

The Monte Carlo designs considered here are chosen to illustrate the method for simple examples, and are not meant to mimic a design that would be encountered for a particular data set. Nevertheless, some features of the designs - namely, the number of observations, percentage of observations, small number of parameters, and uniform distribution of the censoring points - are not too far from the corresponding characteristics of the empirical example. The model used in this simulation study is

$$y_i = \min\{\alpha_0 + x_i\beta_0 + \varepsilon_i, c_i\}, \qquad i = 1, \ldots, 200, \tag{4.1}$$

where the scalar regressor x_i has a standard normal distribution, the censoring variable c_i is uniformly distributed on the interval [-1.5, 1.5], and the true values α_0 and β_0 of the parameters are -1 and 1, respectively. Four homoskedastic distributions are considered for the error term σ_i: the standard normal distribution and student$-t$ distributions with 1, 2, and 3 degrees of freedom (all normalized to have the same interquartile range as the standard normal). In addition, two designs with heteroskedastic errors were considered: $\varepsilon_i = \sigma(x_i) \cdot \eta_i$, with η_i having a standard normal distribution and either $\sigma^2(x_i) = \exp(-x_i)$ or $\sigma^2(x_i) = \exp(x_i)$. For these designs, the overall censoring probabilies vary between 25% and 35%.

For each replication of the model, the following estimators were calculated:

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

b) The randomly-censored least absolute deviations estimator $\hat{\beta}$ defined in (3.1) above (with $\pi = 1/2$); and

c) A modification of the symmetrically-censored least squares estimator derived by applying (2.8) (with an estimated censoring survivor function) to the objective function for Powell's (1986b) STLS estimator (as discussed in the concluding section).

The randomly-censored least absolute deviations estimator was computed using the iterative Barrodale-Roberts algorithm described by Buchinsky (1991); in the random censoring setting, the objective function $R_n(\beta;\hat{S})$ of (3.2) can be transformed into a weighted version of the objective function for the censored quantile estimator with fixed censoring, with the quantile criterion function $\rho_\pi(\cdot)$ for each censored observation being evaluated at every support point of the product-limit estimator of the censoring distribution $G(t)$, with weights proportional to the estimated probabilities at each support point. The STLS estimator described in c) was calculated using an obvious extension of the iteration scheme described in Powell (1986b).

The results of 1000 replication of these estimators for each design are summarized in Table 1, which reports the true values of α and β, the mean bias and root-mean-squared error of the estimators, as well as robust measures of location scale, the median bias and median absolute error. Theoretically, the randomly-censored least absolute deviations and symmetrically-trimmed least squares estimators are consistent under all of the designs considered, whereas the Buckley-James estimator is inconsistent when the errors are $t(1)$(i.e., Cauchy) distributed or heteroskedastic. The results in Table 1 indicate that the estimation methods proposed here perform almost as well as the Buckley-James estimator under normality, and that the superiority of the latter disappears when the errors are nonnormal. As might be expected, the procedures proposed here, which do not impose homoskedasticity of the error terms, are superior to Buckley-James when the errors are heteroskedastic.

Turning now to the empirical example, we consider the well-known Stanford heart transplant data set published in Miller and Halpern (1982). An earlier subset of these data were analyzed using parametric methods (and the Cox 1972, 1975 proportional hazards model) in the text by Kalbfleisch and Prentice (1980), while Miller and Halpern (1982) and Ying, Jung, and Wei (1991) apply several semiparametric methods to the data available through February 1980. Summarized in this data set are the survival times of 184 patients who received heart transplants at the Stanford University Medical Center, as well as an indicator variable which equals one if the patient was dead (uncensored) at the time the data were collected, the age of the patient (in years) at the time of the transplant, and a tissue-mismatch variable. In the analyses of Miller and Halpern (1982) and Ying, Jung, and Wei (1991), 27 observations with missing values of the tissue mismatch scores were dropped, even though the main specification of the regression function in these papers was a quadratic function of age, and excluded the mismatch variable. Following these earlier studies, we consider the same data set of 157 observations (including 55 censored observations), and the same model of the survival times,

$$y_i = \min\{\alpha_0 + \beta_0 x_i + \gamma_0 (x_i)^2 + \varepsilon_i, c_i\}, \tag{4.2}$$

where the dependent variable y_i is the logarithm (base 10) of the observed survival time (in days), and x_i is the age of patient i. (For one observation, the survival time was listed as zero days; this was recoded to one for the statistical analysis here.)

Table 2 reports the randomly-censored regression quantile coefficient estimates at the three quartiles — $\pi = 0.25, 0.50$, and 0.75 — along with the Buckley-James estimator and the Ying-Jung-Wei coefficient estimator given in the aforementioned study. The standard errors for Buckley-James and the three quartile estimators were calculated as the median absolute deviation of the bootstrap c.d.f. (based upon $R = 250$ replications) divided by 0.67, which would (approximately) equal one for a standard normal distribution. Our results for Buckley-James differ from those reported by Miller and Halpern (1982), which deleted 5 observations from the sample with survival times less than 10 days.

Looking across the various coefficient estimates, the results appear fairly similar for all methods, except that the slope coefficients for the Ying, Jung, and Wei (1991) estimator are of smaller magnitude than those for the other procedures. Also, for the quartile estimators there appears to be a "flattening" of the inverted-U shape of the regression function estimates as π moves from 0.25 to 0.75. This flattening, if statistically significant, would indicate heteroskedasticity of the error distribution (or, admittedly, some other misspecification of the model), with the conditional distributions for younger and older patients being more dispersed and skewed downward. To test for significance of the difference between the estimated upper and lower quartile regression lines, a chi-squared statistic was constructed using a quadratic form in these differences about the inverse of a bootstrap estimator of the covariance matrix of the estimator, but the resulting test statistic was insignificant at all conventional levels of significance, so the hypothesis of independence of the error terms and regressors would not be rejected using this test.

5 Concluding Remarks

Although the analysis of the preceding sections has concentrated on the properties of quantile estimators of the slope coefficients, other estimation methods developed for fixed censoring are easily adapted to the present setting. For example, under the assumption of conditional symmetry of the error terms ε_i around zero given x_i, Powell (1986b) proposed an estimator which can be written as a minimizer of the form (2.4) above, with first-order condition of the form (2.5) having

$$\psi(y_i, c_i, x_i, \beta) \equiv [\max\{y_i, 2x_i'\beta - c_i\} - x_i'\beta] \cdot x_i, \tag{5.1}$$

which has expectation zero when evaluated at the true value β_0 under conditional symmetry. Modification of this estimator, developed for fixed censoring, to random right censorship is immediate using (2.8) and the Kaplan-Meier estimator, as described in section 2. The Monte Carlo results of section 4 suggests this estimator may have similar behavior to the randomly-censored quantile estimator with $\pi = 1/2$, at least for symmetric error distributions like the ones considered there. While conditional symmetry may not be an attractive assumption for an accelerated failure time model (ruling out, for example, a Weibull model for durations), it may be more reasonable for other randomly-censored regression models.

Another fixed-censoring estimation method which is easily adapted to random censoring is the method proposed by Honoré (1992) for estimation of panel data models with censoring. For the special case of $T = 2$ time periods, the model Honoré (1992) considers is

$$y_{it} = \min\{x_{it}'\beta_0 + \delta_i + \varepsilon_{it}, c_{it}\}, \qquad i = 1, \ldots, n, \qquad t = 1, 2, \tag{5.2}$$

where the term δ_i is an unobservable "fixed effect" which need not be independent of the covariate vector x_{it}. Under the assumption that $\varepsilon_{i2} - \varepsilon_{i1}$ is symmetrically distributed about zero given the regressors, Honoré proposed an estimator which solves a first-order condition of the form,

$$o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n} \zeta(e_{i,12}(\hat{\beta}) - e_{i,21}(\hat{\beta})) \cdot (x_{i2} - x_{i1}), \tag{5.3}$$

where $\zeta(\cdot)$ is a nondecreasing and odd function of its argument and

$$e_{i,12}(\beta) \equiv \min\{y_{i1} - x_{i1}'\beta, c_{i1} - x_{i2}'\beta\}, \tag{5.4}$$

with an analogous definition of $e_{i,21}(\beta)$. With an appropriate redefinition of the variables, these estimating equations are obviously of the form (2.5), so the transformation (2.9) yields estimating equations for random censoring when the censoring distribution $G(t)$ is replaced by its Kaplan-Meier estimator. When $\zeta(\cdot) = \text{sign}(\cdot)$, this estimator is similar in structure to the randomly-censored regression quantile estimator studied above, and a simple extension of the assumptions imposed in section 3 will suffice to demonstrate the root$-n$ consistency and asymptotic normality of this estimator and others based upon different choices of $\zeta(\cdot)$.

Under the assumption of independence of the error terms and regressors, Honoré and Powell (1993) propose an estimator of β_0 for model (2.3) which uses the same strategy as Honoré's censored panel data estimator, but is based upon pairwise differences across individuals rather than across time periods for each individual. That is, the estimator $\hat{\beta}$ solves estimating equations defined in terms of a second-order U-statistic,

$$o_p(n^{-1/2}) = \binom{n}{2}^{-1} \sum_{i<j} \xi(e_{ij}(\hat{\beta}) - e_{ji}(\hat{\beta})) \cdot (x_i - x_j), \tag{5.5}$$

with

$$e_{ij}(\beta) \equiv \min\{y_i - x_i'\beta, c_i - x_j'\beta\}. \tag{5.6}$$

The approach described in section 2 will also work here, but may be computationally difficult; since calculation of the empirical expectations over the unobserved values of c_i using the Kaplan-Meier c.d.f. estimator involves $O(n)$ calculation, computing a random censoring version of the estimating equations (5.5) will involve $O(n^4)$ summations, which may take some time if n is large. It may be possible to reduce the number of calculations needed, at some cost of statistical precision, by replacing the calculation of an expectation over the censoring value by a single draw from its estimated conditional distribution given the observed data. Whether such an approach would yield a root$-n$ consistent estimator is an interesting question for additional research.

Of the regularity conditions imposed in section 3 above, some may be relaxed without affecting the consistency or asymptotic normality of the estimator (for example, the assumption of randomly-sampled regressors may be relaxed to permit deterministic regressors). However, the assumption of independence of the censoring values $\{c_i\}$ and the regressors $\{x_i\}$ is crucial to the analysis above, and this assumption may be suspect in some settings. For example, in the Stanford heart transplant data set, larger censoring times correspond to earlier transplants; if transplants for younger or older patients were not typically performed in the earlier years, this would induce a dependence between censorship and the covariate, age. In general, if the regressors $\{x_i\}$ have finite support, then it should be possible to obtain consistent estimators of the conditional censoring distribution $G(t \mid x) \equiv \Pr\{c_i \leq t \mid x_i = x\}$

at each possible value of x_i, which could then be substituted into the expectations in (2.8) and (2.9). If some components of the regressors are continuously distributed, it should be possible to nonparametrically estimate the conditional censoring distribution by grouping observations with adjacent values of x_i; whether substitution of a conditional version of the product-limit estimator into (2.8) will yield a root$-n$ consistent estimator is an interesting open question for additional study.

6 References

Amemiya, T. (1985), *Advanced Econometrics.* Cambridge, Mass, Harvard University Press.

Buchinsky, M. (1991), "A Monte Carlo Study of the Asymptotic Covariance Estimators for Quantile Regression Coefficients," manuscript, Harvard University, January.

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

Cox D.R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society Series B*, 34,187-220.

Cox D.R. (1975), "Partial Likelihood" *Biometrika*, 62, 269-276

Csörgö, S. and L. Horvath (1983), "The Rate of Strong Uniform Consistency for the Product Limit Estimator," *Z. Wahrsch. verw. Gebiete*, 62, 411-426.

Duncan, G.M. (1986), "A Semiparametric Censored Regression Estimator," *Journal of Econometrics*, 32, 5-34.

Fernandez, L. (1986), "Nonparametric Maximum Likelihood Estimation of Censored Regression Models," *Journal of Econometrics*, 32, 35-57.

Gill, R. D. (1980), "Censoring and Stochastic Integrals", *Mathematical Centre Tracts*, 124, Mathematisch Centrum, Amsterdam.

Hahn, J. (1993), "Bootstrapping Quantile Regression Estimators," manuscript, Department of Economics, Harvard University.

Honoré, B.E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

Honoré, B.E. and J.L. Powell (1993), "Pairwise Difference Estimators of Linear, Censored, and Truncated Regression Models," *Journal of Econometrics*, forthcoming.

Horowitz, J.L. (1986), "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models," *Journal of Econometrics*, 32, 59-84.

Horowitz, J.L. (1988a), "Semiparametric M-Estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45-83.

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 4, 221-233.

Huber, P.J. (1981), *Robust Statistics*. New York: Wiley.

Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*. New York: Wiley.

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

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

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

Lai, T.L. and Z. Ying (1991), "Rank Regression Methods for Left Truncated and Right Censored Data," *Annals of Statistics*, 19, 531-554.

Lee, M.J. (1993a), "Windsorized Mean Estimator for Censored Regression Model," *Econometric Theory*, forthcoming.

Lee, M.J. (1993b), "Quadratic Mode Regression," *Journal of Econometrics*, forthcoming.

Leurgans, S. (1987), "Linear Models, Random Censoring, and Synthetic Data," *Biometrika*, 74, 301-309.

Lin, D.Y. and C.J. Geyer (1992), "Computational Methods for Semiparametric Linear Regression with Censored Data," *Journal of Computational and Graphical Statistics*, forthcoming.

Lin, J.S. and L.J. Wei (1992), "Linear Regression Analysis for Multivariate Failure Time Observations," *Journal of the American Statistical Association*, forthcoming.

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

Miller, R. (1976), "Least Squares Regression with Censored Data," *Biometrika*, 63, 449-464.

Miller, R. and J. Halpern (1982), "Regression with Censored Data," *Biometrika*, 69, 521-531.

Moon, C-G. (1989), "A Monte Carlo Comparison of Semiparametric Tobit Estimators," *Journal of Applied Econometrics*, 4, 361-382.

Nawata, K. (1990), "Robust Estimation Based on Grouped-Adjusted Data in Censored Regression Models," *Journal of Econometrics*, 43, 337-362.

Newey, W.K. (1985), "Semiparametric Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Annales de l'Insee*, 59/60, 219-236.

Newey, W.K. (1989a), "Efficient Estimation of Tobit Models Under Symmetry," in Barnett, W.A., J.L. Powell, and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge: Cambridge University Press.

Newey, W.K. and J.L. Powell (1990), "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6: 295-317.

Nolan, D. and D. Pollard (1987), "U-Processes: Rates of Convergence", *Annals of Statistics*, 15, 780-799

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1058.

Pollard, D. (1985), "New Ways to Prove Central Limit Theorems," *Econometric Theory,* 1, 295-314.

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

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

Powell, J.L. (1986b), "Symmetrically Trimmed Least Squares Estimation of Tobit Models," *Econometrica*, 54, 1435-1460.

Powell, J.L. (1991), "Estimation of Monotonic Regression Models Under Quantile Restrictions," in W.A. Barnett, J.L. Powell, and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press.

Prentice, R.L. (1978), "Linear Rank Tests with Right Censored Data," *Biometrika*, 65, 167-179.

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

Robins, J.M. and A.A. Tsiatis (1992), "Semiparametric Estimation of an Accelerated Failure Time Model with Time-Dependent Covariates," *Biometrika*, forthcoming.

Serfling, R.J. (1980), *Approximations of Mathematical Statistics*, Wiley, New York.

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

Shorak, G. and J. Wellner (1986), *Empirical Processes With Applications to Statistics.* New York: Wiley.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24-36.

Tsiatis, A.A. (1992), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *Annals of Statistics*, 18, 354-372.

Wei, L.J., Z. Ying, and D.Y. Lin (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests." *Biometrika*, 19, 417-442.

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

Ying, Z., S.H. Jung, and L.J. Wei (1991), "Survival Analysis with Median Regression Models," manuscript, Department of Statistics, University of Illinois.

A Proofs of Theorems in Text

A.1 Proof of Theorem 3.1

In this section, for any vector x, we let $\|x\|$ denote its Euclidean norm. Define:

$$R(\beta) \equiv E\left[\rho_\pi(y_i - \min(x_i'\beta, c_i)) - \rho_\pi(y_i - \min(x_i'\beta_0, c_i))\right] \tag{A.1}$$

then the key step in showing consistency of:

$$\hat{\beta} \equiv \text{argmin}_{\beta\in\mathcal{B}} R_n(\beta;\hat{S}) \equiv \text{argmin}_{\beta\in\mathcal{B}}\left(R_n(\beta;\hat{S}) - R_n(\beta_0;\hat{S})\right)$$

is the demonstration of:

$$\sup_{\beta\in\mathcal{B}}\left|\left(R_n(\beta;\hat{S}) - R_n(\beta_0;\hat{S})\right) - R(\beta)\right| = o(1) \quad \text{a.s.} \tag{A.2}$$

where

$$R_n(\beta;\hat{S}) \equiv \frac{1}{n}\sum_{i=1}^{n}\left((1-d_i)\rho_\pi(y_i - \min(x_i'\beta, y_i)) + d_i\hat{S}(y_i)^{-1}\int I[y_i < c]\rho_\pi(y_i - \min(x_i'\beta, c))d\hat{G}(c)\right)$$

as defined in (3.2) above.

To show result (A.2), first note that for all $\beta \in \mathcal{B}$,

$$\begin{aligned} R_n(\beta;\hat{S}) - R_n(\beta;S) &= \frac{1}{n}\sum_{i=1}^{n} d_i\left(\hat{S}(y_i)^{-1} - S(y_i)^{-1}\right)\int I[y_i < c]\rho_\pi(y_i - \min(x_i'\beta, c))d\hat{G}(c) \quad \text{(A.3)} \\ &+ \frac{1}{n}\sum_{i=1}^{n} d_i S(y_i)^{-1}\int I[y_i < c]\rho_\pi(y_i - \min(x_i'\beta, c))\left(d\hat{G}(c) - dG(c)\right) \end{aligned}$$

Since, for any c and $\beta_1, \beta_2 \in \mathcal{B}$,

$$|\rho_\pi(y_i - \min(x_i'\beta_1, c)) - \rho_\pi(y_i - \min(x_i'\beta_2, c))| \le \|x_i\|\|\beta_1 - \beta_2\| \le \chi_0\|\beta_1 - \beta_2\| \tag{A.4}$$

where χ_0 is the upper bound for $\|x_i\|$ given in assumption R, it follows that:

$$\sup_{\beta\in\mathcal{B}}\left|R_n(\beta;\hat{S}) - R_n(\beta;S) - R_n(\beta_0;\hat{S}) + R_n(\beta_0;S)\right| \tag{A.5}$$

$$\begin{aligned} &\le \sup_{y<\tau_0}\left|\hat{S}(y)^{-1} - S(y)^{-1}\right|\chi_0(2b_0) \\ &+ \sup_{\beta\in\mathcal{B}}\left|\int I[y_i < c]\left(\rho_\pi(y_i - \min(x_i'\beta, c)) - \rho_\pi(y_i - \min(x_i'\beta_0, c))\right)d\left(\hat{G}(c) - G(c)\right)\right| \end{aligned}$$

where τ_0 is the upper support point for the censoring distribution and b_0 is an upper bound for $\|\beta\|$ on the compact set $\mathcal{B}$. Now by the uniform convergence results of Shorack and Wellner(1986, Theorem 7.3.1 and equation (3) of section 7.3) (see also Proposition 1 of Wang(1987)),

$$\sup_y \left|\hat{S}(y) - S(y)\right| = \sup_y \left|\hat{G}(y) - G(y)\right| = o(1) \quad \text{a.s.} \tag{A.6}$$

along with the condition that $S(\tau_0-) > 0$ (from Assumption C), this implies that the first term on the right hand side of inequality (A.5) converges to 0 alomost surely. Moreover, the almost sure consistency of the Kaplan-Meier estimator $\hat{G}(y)$ for $G(y)$ implies that, for each value of $\beta \in \mathcal{B}$, the integral in the second term of the right hand side of (A.5) converges to 0 almost surely, and this pointwise convergence can be easily extended to uniform convergence over $\beta \in \mathcal{B}$ using the compactness of $\mathcal{B}$, the Lipschitz condition in (A.4), and a standard partitioning argument (e.g., in the proof of Theorem 4.2.1 of Amemiya(1985)).

This argument has established

$$\sup_{\beta\in\mathcal{B}} \left|R_n(\beta;\hat{S}) - R_n(\beta;S) - R_n(\beta_0;\hat{S}) + R_n(\beta_0;S)\right| = o(1) \quad \text{a.s.} \tag{A.7}$$

Also,

$$R_n(\beta;S) - R_n(\beta_0;S) = \frac{1}{n}\sum_{i=1}^{n} E\left[\rho_\pi(y_i - \min(x_i'\beta, c)) - \rho_\pi(y_i - \min(x_i'\beta_0, c))\middle|y_i, d_i, x_i\right] \tag{A.8}$$

is an empirical process satisfying the conditions for applicability of a uniform law of large numbers (e.g., Amemiya(1985), Theorem 4.2.1), so

$$\sup_{\beta\in\mathcal{B}} |(R_n(\beta;S) - R_n(\beta_0;S)) - R(\beta)| = o(1) \quad \text{a.s.} \tag{A.9}$$

which together with (A.7), establishes (A.2).

A.2 Proof of Theorem 3.2

In this section, we derive the limiting distribution of the estimator, using the consistency result established in the previous section. The argument is based on deriving a preliminary rate of convergence for the estimator which is slower than the parametric rate, and then in turn establishing root-n consistency and asymptotic normality. Throughout this section, for any matrix A, we let $\|A\|$ denote $(\sum_{i,j} A_{ij}^2)^{1/2}$ where A_{ij} denotes the components of A. Also, all asymptotically negligible remainder terms will be denoted by $\mathcal{R}_n(\cdot)$.

The first lemma establishes a linear representation for an estimator which solves an infeasible first order condition that assumes the distribution of the censoring variable is known:

Theorem A.1 *If $\hat{\beta} \xrightarrow{p} \beta_0$, $\chi_i \equiv \chi(y_i, x_i, d_i)$ is any mean 0 random vector with finite variance, and $\hat{\beta}$ solves the following relationship:*

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i(\hat{\beta}, S) + \chi_i = o_p(n^{-\delta}) \tag{A.10}$$

where $0 < \delta \leq 1/2$*, then:*

$$\hat{\beta} - \beta_0 = M_0^{-1}\frac{1}{n}\sum_{i=1}^{n}(\psi_i + \chi_i) + o_p(n^{-\delta}) \tag{A.11}$$

Proof: Noting that $E[\psi_i(\beta, S)|x_i] = S(x_i'\beta)(\pi - F_{\epsilon|X}(x_i'(\beta - \beta_0))$ we first evaluate the expansion of $E[\psi_i(\beta, S)]$ for β in a neighborhood of β_0:

Lemma A.1 *as* $\|\beta - \beta_0\| \to 0$*,*

$$E[S(x_i'\beta)(\pi - F_{\epsilon|X}(x_i'(\beta - \beta_0)))x_i] = E[S(x_i'\beta_0)f_{\epsilon|X}(0)x_i x_i'](\beta - \beta_0) + o(\|\beta - \beta_0\|) \tag{A.12}$$

Proof: We add and substract $E[S(x_i'\beta_0)(\pi - F_{\epsilon|X}(x_i'(\beta - \beta_0)))x_i]$ from the left hand side of (A.12). We first show that:

$$E[(S(x_i'\beta) - S(x_i'\beta_0))(\pi - F_{\epsilon|X}(x_i'(\beta - \beta_0)))x_i] = o(\|\beta - \beta_0\|) \tag{A.13}$$

Note that a mean value expansion of $F_{\epsilon|X}(x_i'(\beta - \beta_0))$ around 0 implies by the bound on the conditional density of ϵ_i in a neighborhood of 0 of Assumption E, the bound on $\|x_i\|$ in Assumption R, and the Cauchy Schwartz inequality that the left hand side of (A.13) is bounded above by:

$$\mathcal{C}E[|S(x_i'\beta) - S(x_i'\beta_0)|]\|\beta - \beta_0\| \tag{A.14}$$

where $\mathcal{C}$ is a constant reflecting the bounds in Assumptions E and R. By the dominated convergence theorem, $E[|S(x_i'\beta) - S(x_i'\beta_0)|]$ is $o(1)$ as $\beta \to \beta_0$ since $S(x_i'\beta_0)$ is discontinuous on a set of probability zero by Assumption RC. This establishes (A.13). We next show that

$$E[S(x_i'\beta_0)(\pi - F_{\epsilon|X}(x_i'(\beta - \beta_0)))x_i] = E[S(x_i'\beta_0)f_{\epsilon|X}(0)x_i x_i'](\beta - \beta_0) + O(\|\beta - \beta_0\|^2) \tag{A.15}$$

A mean value expansion of the left hand side of (A.15) yields:

$$E[S(x_i'\beta_0)f_{\epsilon|X}(0)x_i x_i'](\beta - \beta_0) + \mathcal{R}_n \tag{A.16}$$

where $\|\mathcal{R}_n\|$ is bounded above by:

$$E[|f_{\epsilon|X}(0) - f_{\epsilon|X}(x_i'(\tilde{\beta} - \beta_0))|\|x_i\|^2]\|\beta - \beta_0\|$$

with $\tilde{\beta}$ denoting the intermediate value in the mean value expansion. By the Lipschitz assumption on the conditional density of ϵ_i in a neighborhood of 0 (Assumption E), and the bound on $\|x_i\|$ (Assumption R), the above term is is $O(\|\beta - \beta_0\|^2)$, establishing (A.15). This shows (A.12). ■

Turning attention to the proof of the theorem, we let $E[\psi_i(\hat{\beta}, S)]$ denote $E[\psi_i(\beta, S)]\Big|_{\beta=\hat{\beta}}$. Express $\frac{1}{n}\sum_{i=1}^n \psi_i(\hat{\beta}, S)$ as

$$\frac{1}{n}\sum_{i=1}^n \psi_i(\beta_0, S) + \frac{1}{n}\sum_{i=1}^n E[\psi_i(\hat{\beta}, S)] + \frac{1}{n}\sum_{i=1}^n \psi_i(\hat{\beta}, S) - \psi_i(\beta_0, S) - E[\psi_i(\hat{\beta}, S)] \tag{A.17}$$

Turning attention to the second term in (A.17), we note that it immediately follows by Lemma A.1 and the consistency of $\hat{\beta}$ that

$$\frac{1}{n}\sum_{i=1}^n E[\psi_i(\hat{\beta}, S)] = M_0 + o_p(1)$$

We next show that the third term in (A.17) is $o_p(n^{-1/2})$. By the consistency of $\hat{\beta}$, it will suffice to show that for a sequence of numbers δ_n converging to 0 slowly enough, we have:

$$\sup_{\|\beta-\beta_0\|\leq\delta_n} \left\| \frac{1}{n}\sum_{i=1}^n \psi_i(\beta, S) - \psi_i(\beta_0, S) - E[\psi_i(\beta, S)] \right\| = o_p(n^{-1/2}) \tag{A.18}$$

To show (A.18), by applying Lemma 2.17 in Pakes and Pollard(1989), it will suffice to show the following two results:

I The class of functions $(\psi_i(\beta, S) : \beta \in \mathcal{B})$ is Euclidean with respect to the envelope F, where $E[F^2] < \infty$.

II $\lim_{\beta\to\beta_0} E[(\psi_i(\beta, S) - \psi_i(\beta_0, S)^2] = 0$.

To show **I**, we note by Lemmas 2.14(i) and 2.14(ii) of Pakes and Pollard(1989), it will suffice to show the Euclidean property for the three classes a)$(I[y_i > x_i'\beta] : \beta \in \mathcal{B})$, b) $(I[y_i \leq x_i'\beta] : \beta \in \mathcal{B})$ c) $(S(x_i'\beta) : \beta \in \mathcal{B})$. The Euclidean property for all three classes for the envelope $F \equiv 1$ follows directly from Lemma 22(ii) in Nolan and Pollard(1987) since the functions $I[\cdot]$ and $S(\cdot)$ are of bounded variation. This establishes **I**.

To establish **II**, we note that it will suffice to show that both $E[|I[y_i > x_i'\beta] - I[y_i > x_i'\beta_0]|]$ and $E[(I[\epsilon_i \leq x_i'(\beta - \beta_0)]S(x_i'\beta) - I[\epsilon_i \leq 0]S(x_i'\beta_0))^2]$ converge to 0 as $\|\beta - \beta_0\| \to 0$. To show the former, we note that $|I[y_i > x_i'\beta] - I[y_i > x_i'\beta_0]|$ is bounded above by $I[|y_i - x_i'\beta_0| \leq \|x_i\|\|\beta - \beta_0\|]$, and that:

$$P(|y_i - x_i'\beta_0| \leq \|x_i\|\|\beta - \beta_0\|) \leq P(|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|) + P(|c_i - x_i'\beta_0| \leq \|x_i\|\|\beta - \beta_0\|)$$

By Assumption E, the first term on the right hand side of the above expression converges to 0 as $\beta \to \beta_0$ since $\|x_i\|$ is bounded by Assumption R. By Assumption RC, the second term converges to 0 as $\beta \to \beta_0$, again using the assumption that $\|x_i\|$ is bounded. To show that $E[(I[\epsilon_i \leq x_i'(\beta - \beta_0)]S(x_i'\beta) - I[\epsilon_i \leq 0]S(x_i'\beta_0))^2]$ converges to 0, it will suffice to show that both $E[|I[\epsilon_i \leq x_i'(\beta - \beta_0)] - I[\epsilon_i \leq 0]|]$ and $E[(S(x_i'\beta) - S(x_i'\beta_0)^2]$ converge to 0 as $\beta \to \beta_0$. The first term is bounded above by $E[I[|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|]]$ which converges to 0 by assumption E, and the second term converges to 0 by the dominated convergence theorem, as $S(x_i'\beta_0)$ is discontinuous on a set of probability 0 by Assumption RC. This establishes **II** and hence (A.18). Thus we have shown that:

$$\frac{1}{n}\sum_{i=1}^n \psi_i(\hat{\beta}, S) = \frac{1}{n}\sum_{i=1}^n \psi_i(\beta_0, S) + (M_0 + o_p(1))(\hat{\beta} - \beta_0) + o_p(n^{-1/2}) \tag{A.19}$$

Combining this with (A.10), we have:

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\beta_0,S)+\chi_i+(M_0+o_p(1))(\hat{\beta}-\beta_0)=o_p(n^{-\delta}) \tag{A.20}$$

which by applying the Lindeberg-Levy central limit theorem and Slutsky's theorem, can be rearranged to yield the conclusion of the theorem. ■

Our next step is to establish a uniform linear representation for the Kaplan-Meier product limit estimator used in the first stage.

Lemma A.2 *Let $H(x)$ denote $P(y_i \geq x)$ and let $\Lambda(\cdot)$ denote the cumulative hazard function of c_i. Letting $\Delta\Lambda(x)$ denote $\Lambda(x)-\Lambda(x-)$ we have the following linear representation for the product limit estimator:*

$$\begin{aligned}\hat{S}(t)-S(t) &= S(t)\frac{1}{n}\sum_{i=1}^{n}H(y_i)^{-1}(1-\Delta\Lambda(y_i))^{-1}I[y_i\leq t](1-d_i) \\ &- \int_0^t H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i\geq s]d\Lambda(s)+\mathcal{R}_n(t)\end{aligned} \tag{A.21}$$

where

$$\sup_{0\leq t<\infty}|\mathcal{R}_n(t)|=o_p(n^{-1/2}) \tag{A.22}$$

Proof: Note by the assumption that τ_0, the upper support point of c_i, is mass point, we have $\hat{S}(t)\equiv 0=S(t)$ for all $t>\tau_0$. It will thus suffice to show that the linear representation holds uniformly over the interval $[0,\tau_0]$. We first define the following processes:

$$N(t)=\sum_{i=1}^{n}I[y_i\leq t,d_i=0]$$

$$Y(t)=\sum_{i=1}^{n}I[y_i\geq t]$$

$$M(t)=N(t)-\int_0^t Y(s)d\Lambda(s)$$

From the proof of Theorem 4.2.2. in Gill(1980), we have

$$\hat{S}(t)-S(t)=S(t)\frac{1}{n}\int_0^t(1-\Delta\Lambda(s))^{-1}\frac{\hat{S}(s-)}{S(s-)}\frac{nI[Y(s)>0]}{Y(s)}dM(s) \tag{A.23}$$

for all $t\in[0,\tau_0]$. We thus have:

$$\hat{S}(t)-S(t)=S(t)\frac{1}{n}\int_0^t(1-\Delta\Lambda(s))^{-1}H(s)^{-1}dM(s)+\mathcal{R}_n(t) \tag{A.24}$$

where

$$n^{1/2}\mathcal{R}_n(t) = S(t)\int_0^t (1-\Delta\Lambda(s))^{-1}\left(n^{-1/2}H(s)^{-1} - \frac{\hat{S}(s-)}{S(s-)}\frac{n^{1/2}I[Y(s)>0]}{Y(s)}\right)dM(s) \tag{A.25}$$

so note it will suffice to show that:

$$\sup_{0\le s\le \tau_0} n^{1/2}|\mathcal{R}_n(s)| = o_p(1) \tag{A.26}$$

Let

$$\mathcal{H}(s) = (1-\Delta\Lambda(s))^{-1}\left(n^{-1/2}H(s)^{-1} - \frac{\hat{S}(s-)}{S(s-)}\frac{n^{1/2}I[Y(s)>0]}{Y(s)}\right)$$

The process $(1-\Delta\Lambda(s))^{-1}\frac{\hat{S}(s-)}{S(s-)}\frac{n^{1/2}I[Y(s)>0]}{Y(s)}$ is bounded and predictable by the arguments used in the proof of Theorem 4.2.2 in Gill(1980). It immediately follows that the process $\mathcal{H}(s)$ is bounded and predictable, and note that $\mathcal{H}^2(s)Y(s)$ is

$$(1-\Delta\Lambda(s))^{-2}n^{-1}H(s)^{-2}Y(s)+ \tag{A.27}$$

$$(1-\Delta\Lambda(s))^{-2}\frac{\hat{S}^2(s-)}{S^2(s-)}\frac{nI[Y(s)>0]}{Y(s)}- \tag{A.28}$$

$$2(1-\Delta\Lambda(s))^{-2}H(s)^{-1}\frac{\hat{S}(s-)}{S(s-)} \tag{A.29}$$

By Theorem 3.1, we have:

$$\sup_{0\le s\le \tau_0} |\hat{S}(s) - S(s)| = o_p(1) \tag{A.30}$$

and note that $Y(s)/n$ converges in probability to $H(s)$, uniformly in $[0,\tau_0]$. Since $H(s)$ is bounded away from 0 on $[0,\tau_0]$, this implies that terms in (A.27)-(A.29) converge uniformly on $[0,\tau_0]$ to

$$(1-\Delta\Lambda(s))^{-2}H(s)^{-1}$$

$$(1-\Delta\Lambda(s))^{-2}H(s)^{-1}$$

and

$$2(1-\Delta\Lambda(s))^{-2}H(s)^{-1}$$

respectively. It thus follows that

$$\sup_{0\le s\le \tau_0} \mathcal{H}^2(s)Y(s) \xrightarrow{p} 0 \tag{A.31}$$

So by Theorem 4.2.1 of Gill(1980)

$$n^{1/2}\mathcal{R}_n(\cdot) \Rightarrow Z_0 \quad \text{in } D[0,\tau_0] \tag{A.32}$$

where $D[0,\tau_0]$ is the space of right continuous functions with left hand limits, and Z_0 is a process degenerate at 0. It immediately follows by Theorem 2.4.3 in Gill(1980) that

$$\sup_{0\le s\le\tau_0} n^{1/2}|\mathcal{R}_n(s)| = o_p(1) \tag{A.33}$$

This establishes (A.21). ■

Implicit in the result of the uniform linear representation is a rate of uniform convergence of the Kaplan-Meier estimator. To formally establish the uniform rate, we first show the Euclidean property of the class of functions in the summation of the linear representation:

Lemma A.3 *The class of functions*

$$\begin{aligned}(H(y_i)^{-1}(1-\Delta\Lambda(y_i))^{-1}I[y_i\le t](1-d_i) \quad &- \quad \int_0^t H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i\ge s]d\Lambda(s) \\ &: \quad t\in[0,\tau_0])\end{aligned} \tag{A.34}$$

is Euclidean for a constant envelope.

Proof: Note that the class $H(y_i)^{-1}(1-\Delta\Lambda(y_i))^{-1}(1-d_i)$ is trivially Euclidean for a constant envelope, and the class $I[y_i\le t]$ is Euclidean for the envelope $F\equiv 1$ by Example 2.11 in Pakes and Pollard(1989). It follows by Lemma 2.14(ii) of Pakes and Pollard(1989) that the class:

$$(H(y_i)^{-1}(1-\Delta\Lambda(y_i))^{-1}I[y_i\le t](1-d_i)$$

is Euclidean for a constant envelope. Next we show the Euclidean property for the class of functions of y_i and s, indexed by t:

$$H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i\ge s]I[y_i\le t] \tag{A.35}$$

The class of functions $H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i\ge s]$ is trivially Euclidean for a constant envelope, and the class $I[y_i\le t]$ is Euclidean for the envelope $F\equiv 1$ by Example 2.11 in Pakes and Pollard(1989). It follows that the class in (A.35) is Euclidean for a constant envelope by Lemma 2.14(ii) of Pakes and Pollard(1989). Therefore, by Lemma 5 in Sherman(1994), the class of functions of y_i, indexed by t:

$$\int_0^t H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i\ge s]d\Lambda(s) : t\in[0,\tau_0]$$

is Euclidean for a constant envelope. It follows by Lemma 2.14(i) in Pakes and Pollard(1989) that the class in (A.34) is Euclidean for a constant envelope. ■

We now have the following result:

Lemma A.4 *For any $\delta < 1/2$:*

$$\sup_{t \in \mathbf{R}^+} |\hat{S}(t) - S(t)| = o_p(n^{-\delta}) \tag{A.36}$$

Proof: Note that for any $t \geq \tau_0$, we have $\hat{S}(t) \equiv 0 = S(t)$, so it suffices to show that:

$$\sup_{t \in [0,\tau_0)} |\hat{S}(t) - S(t)| = o_p(n^{-\delta})$$

Working with the linear representation in (A.21), by the fact that the remainder term is $o_p(n^{-1/2})$ uniformly over $[0, \tau_0]$, it remains to show that:

$$\sup_{t \in [0,\tau_0)} \Big| \frac{1}{n} \sum_{i=1}^{n} (H(y_i)^{-1}(1 - \Delta\Lambda(y_i))^{-1} I[y_i \leq t](1 - d_i) \tag{A.37}$$
$$- \int_0^t H(s)^{-1}(1 - \Delta\Lambda(s))^{-1} I[y_i \geq s] d\Lambda(s) \Big| = o_p(n^{-\delta})$$

By the Euclidean property of the class in (A.34) this follows directly by Corollary 9 in Sherman(1994). ■

The uniform rate of convergence will suffice to establish a preliminary rate of convergence for the estimator $\hat{\beta}$.

Lemma A.5 *For any $\delta \in (0, 1/2)$,*

$$\hat{\beta} - \beta_0 = o_p(n^{-\delta}) \tag{A.38}$$

Proof: We rewrite the first order condition as:

$$\frac{1}{n} \sum_{i=1}^{n} \psi_i(\hat{\beta}, S) + \frac{1}{n} \sum_{i=1}^{n} \psi_i(\hat{\beta}, \hat{S}) - \psi_i(\hat{\beta}, S) = o_p(n^{-1/2}) \tag{A.39}$$

By linearizing the ratio $\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)}$ around $\frac{S(x_i'\hat{\beta})}{S(y_i)}$ and the assumptions that $d_i/S(y_i)$ and $\|x_i\|$ are bounded, Lemma A.4 implies that:

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \psi_i(\hat{\beta}, \hat{S}) - \psi_i(\hat{\beta}, S) \right\| = o_p(n^{-\delta}) \tag{A.40}$$

for any $\delta \in (0, 1/2)$. Thus we have:

$$\frac{1}{n} \sum_{i=1}^{n} \psi_i(\hat{\beta}, S) = o_p(n^{-\delta}) \tag{A.41}$$

to which we can apply Theorem A.1 with $\chi_i \equiv 0$ to conclude that $\hat{\beta} - \beta_0 = o_p(n^{-\delta}) + O_p(n^{-1/2}) = o_p(n^{-\delta})$. ■

We next show the following result:

Lemma A.6 *Let ξ_i be defined as in equation (3.12). Then*

$$(1-\pi)\frac{1}{n}\sum_{i=1}^{n} I[y_i \le x_i'\beta_0]d_i\left(\frac{\hat{S}(x_i'\beta_0)}{\hat{S}(y_i)}-\frac{S(x_i'\beta_0)}{S(y_i)}\right)x_i \tag{A.42}$$

has the representation:

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i + o_p(n^{-1/2}) \tag{A.43}$$

Proof: The proof is facilitated by decomposing ξ_i as the sum of two components, which we denote by ξ_{1i}, ξ_{2i}, and are defined as :

$$\begin{aligned}
\xi_{1i} &= \pi(1-\pi)\int_{\mathcal{X}} I[x_i'\beta_0 \le \tau_0]\Big(H(y_i)^{-1}I[y_i \le x_i'\beta_0](1-d_i) && \text{(A.44)}\\
&- \int_0^{x_i'\beta_0} H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i \ge s]d\Lambda(s)\Big)xdF_X(x) \\
\xi_{2i} &= -(1-\pi)\int I[\epsilon \le 0]I[x'\beta_0+\epsilon < c]\frac{S(x'\beta_0)}{S(x'\beta_0+\epsilon)} && \text{(A.45)}\\
&\times \left(H(y_i)I[y_i \le x'\beta_0+\epsilon](1-d_i) - \int_0^{x'\beta_0+\epsilon} H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i \ge s]d\Lambda(s)\right) \\
&\times xdF_{X,\epsilon}(x,\epsilon)dF_C(c)
\end{aligned}$$

Linearizing the ratio $\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)}$ around $\frac{S(x_i'\hat{\beta})}{S(y_i)}$, we have by Lemma A.4 and the assumptions that $\|x_i\|$ and $d_i/S(y_i)$ are bounded that (A.42) can be written as:

$$\frac{1}{n}\sum_{i=1}^{n}(1-\pi)I[y_i \le x_i'\beta_0]d_iS(y_i)^{-1}(\hat{S}(x_i'\beta_0)-S(x_i'\beta_0))x_i + \tag{A.46}$$

$$\frac{1}{n}\sum_{i=1}^{n}(1-\pi)I[y_i \le x_i'\beta_0]d_iS(x_i'\beta_0)S(y_i)^{-2}(\hat{S}(y_i)-S(y_i))x_i + o_p(n^{-1/2}) \tag{A.47}$$

We first establish a representation for (A.46). Here, we "plug in" the linear representation for $\hat{S}(\cdot)-S(\cdot)$ established in Lemma A.2. Noting that the "own observation" terms are asymptotically negligible, this yields a U-statistic plus and asymptotically negligible term:

$$\frac{1}{n(n-1)}\sum_{i\neq j}(1-\pi)I[y_i \le x_i'\beta_0]d_iS(x_i'\beta_0)S(y_i)^{-1}\times \tag{A.48}$$

$$\Big(H(y_j)^{-1}(1-\Delta\Lambda(y_j))^{-1}I[y_j \le x_i'\beta_0](1-d_j)$$

$$-\int_0^{x_i'\beta_0} H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_j \ge s]d\Lambda(s)\Big)x_i + o_p(n^{-1/2})$$

The left hand side of the above expression is a second order U-statistic, and we denote its kernel function by $\mathcal{F}(\zeta_i, \zeta_j)$ where $\zeta_i = (y_i, x_i', d_i)'$. It is straightforward to show that $E[\|\mathcal{F}(\zeta_i, \zeta_j)\|^2] < \infty$ by the assumptions that $d_i/S(y_i)$, $H(\cdot)^{-1}$ and $\|x_i\|$ are bounded. We note that $E[\mathcal{F}(\zeta_i, \zeta_j)] = E[\mathcal{F}(\zeta_i, \zeta_j)|\zeta_i] = 0$ and $\frac{1}{n}\sum_{j=1}^n E[\mathcal{F}(\zeta_i, \zeta_j)|\zeta_j]$ can be written as $\frac{1}{n}\sum_{i=1}^n \xi_{1i}$. Thus by a standard projection theorem for U-statistics (see for example Serfling(1980)), (A.46) can be expressed as:

$$\frac{1}{n}\sum_{i=1}^n \xi_{1i} + o_p(n^{-1/2}) \tag{A.49}$$

The same arguments can be used to represent (A.47) as:

$$\frac{1}{n}\sum_{i=1}^n \xi_{2i} + o_p(n^{-1/2}) \tag{A.50}$$

This establishes the conclusion of the lemma. ■

We next establish the following lemma:

Lemma A.7

$$\frac{1}{n}\sum_{i=1}^n (1-\pi)d_i(I[y_i \leq x_i'\hat{\beta}] - I[y_i \leq x_i'\beta_0])d_i \left(\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)} - \frac{S(x_i'\hat{\beta})}{S(y_i)}\right) x_i = o_p(n^{-1/2}) \tag{A.51}$$

Proof: Fix $\delta \in (1/4, 1/2)$. By linearizing the ratio $\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)}$ around $\frac{S(x_i'\hat{\beta})}{S(y_i)}$, we have by Lemma A.4, and the assumption that $\|x_i\|$ and $d_i/S(y_i)$ are bounded, that it suffices to show:

$$\frac{1}{n}\sum_{i=1}^n |I[y_i \leq x_i'\hat{\beta}] - I[y_i \leq x_i'\beta_0]|d_i = o_p(n^{-1/2+\delta}) \tag{A.52}$$

We note that the left hand side of the above expression is bounded above by:

$$\frac{1}{n}\sum_{i=1}^n I[|\epsilon_i| \leq \|x_i\|\|\hat{\beta} - \beta_0\|] \tag{A.53}$$

we can multiply this expression by $I[\|\hat{\beta} - \beta_0\| \leq n^{-\delta}]$ and the resulting remainder term is $o_p(n^{-1/2})$ by Lemma A.5. Note that for any β, by Assumption E we have $E[I[|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|]]$ is $O(\|\beta - \beta_0\|)$. It will thus suffice to show that:

$$\sup_{\|\beta-\beta_0\| \leq n^{-\delta}} \left| \frac{1}{n}\sum_{i=1}^n I[|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|] - E[I[|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|]] \right| = o_p(n^{-1/2}) \tag{A.54}$$

This follows by Lemma 2.17 in Pakes and Pollard(1989), since the class of functions indexed by β at hand is Euclidean for the envelope $F \equiv 1$ by example 2.11 in Pakes and Pollard(1989), and $P\left(|\epsilon_i| \leq \|x_i\|\|\beta - \beta_0\|\right) \to 0$ as $\beta \to \beta_0$ by Assumption E. ■

The final result which needs to be established before proving the main theorem is an equicontinuity condition for the Kaplan-Meier estimator:

Lemma A.8

$$\frac{1}{n}\sum_{i=1}^{n} I[y_i \le x_i'\hat{\beta}]d_i \left(\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)} - \frac{S(x_i'\beta_0)}{S(y_i)} - \frac{\hat{S}(x_i'\beta_0)}{\hat{S}(y_i)} + \frac{S(x_i'\beta_0)}{S(y_i)}\right) x_i = o_p(n^{-1/2}) \tag{A.55}$$

Proof: Again, we linearize the ratios $\frac{\hat{S}(x_i'\hat{\beta})}{\hat{S}(y_i)}$, and $\frac{\hat{S}(x_i'\beta_0)}{\hat{S}(y_i)}$, which by Lemma A.4, and the bounds on $\|x_i\|, d_i/S(y_i)$ makes it suffice to show that:

$$\frac{1}{n}\sum_{i=1}^{n} I[y_i \le x_i'\hat{\beta}]d_i S(y_i)^{-1}\left(\hat{S}(x_i'\hat{\beta}) - S(x_i'\hat{\beta}) - \hat{S}(x_i'\beta_0) + S(x_i'\beta_0)\right) x_i = o_p(n^{-1/2}) \tag{A.56}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I[y_i \le x_i'\hat{\beta}]d_i S(y_i)^{-2}\left(S(x_i'\hat{\beta}) - S(x_i'\beta_0)\right)\left(\hat{S}(y_i) - S(y_i)\right) x_i = o_p(n^{-1/2}) \tag{A.57}$$

We first show (A.57). We note by Lemma A.4 and the bound on $I[y_i \le x_i'\hat{\beta}]d_i S(y_i)^{-2}x_i$ that it will suffice to show:

$$\frac{1}{n}\sum_{i=1}^{n} |S(x_i'\hat{\beta}) - S(x_i'\beta_0)| = o_p(n^{-\delta}) \tag{A.58}$$

for $\delta \in (1/4, 1/2)$. Assumption RC(i) implies that $E[|S(x_i'\hat{\beta}) - S(x_i'\beta_0)|] = O_p(\|\hat{\beta} - \beta_0\|)$, so by the consistency of $\hat{\beta}$, it will suffice to show that for a sequence of numbers δ_n converging to 0 slowly enough that:

$$\sup_{\|\beta-\beta_0\|\le\delta_n} \frac{1}{n}\sum_{i=1}^{n} |S(x_i'\beta) - S(x_i'\beta_0)| - E\left[|S(x_i'\beta) - S(x_i'\beta_0)|\right] = o_p(n^{-1/2}) \tag{A.59}$$

Note that the class of functions $(S(x_i'\beta) : \beta \in \mathcal{B})$ is Euclidean for the envelope $F \equiv 1$ by Lemma 22(ii) in Nolan and Pollard(1987). It immediately follows that the class $(|S(x_i'\beta) - S(x_i'\beta_0)| : \beta \in \mathcal{B})$ is Euclidean for a constant envelope. Also, by Assumption RC(i) it follows that $E\left[(S(x_i'\beta) - S(x_i'\beta_0))^2\right] \to 0$ as $\beta \to \beta_0$. (A.59) follows from Lemma 2.17 in Pakes and Pollard(1989), showing (A.57)

We next show (A.56). Note that it can be shown as before that:

$$\frac{1}{n}\sum_{i=1}^{n} d_i(I[y_i \le x_i'\beta] - I[y_i \le x_i'\beta_0]) = O_p(\|\beta - \beta_0\|)$$

so $I[y_i \le x_i'\hat{\beta}]$ can be replaced with $I[y_i \le x_i'\beta_0]$ in (A.56) and the resulting remainder term is $o_p(n^{-1/2})$. By Lemma A.4 and the fact that $\hat{S}(t) - S(t) = 0$ for $t > \tau_0$, it will suffice to show:

$$\sup_{\|\beta-\beta_0\|\le n^{-\delta}} \quad \frac{1}{n}\sum_{i=1}^{n} I[y_i \le x_i'\hat{\beta}]d_i S(y_i)^{-1} I[x_i'\beta \le \tau_0](\hat{S}(x_i'\beta) - S(x_i'\beta)) \tag{A.60}$$
$$- \quad I[x_i'\beta \le \tau_0](\hat{S}(x_i'\beta_0) - S(x_i'\beta_0))x_i = o_p(n^{-1/2})$$

We next plug in the linear representation of Lemma A.2. Again, by noting that the own observation terms are asymptotically negligible, the summation of the left hand side in (A.60) can be written as a U-statistic:

$$\frac{1}{n(n-1)}\sum_{i\neq j} I[y_i \leq x_i'\hat{\beta}]d_i S(y_i)^{-1}(\mathcal{Q}_j(x_i'\beta) - \mathcal{Q}_j(x_i'\beta_0))x_i \tag{A.61}$$

where here we let $\mathcal{Q}_i(t)$ denote the mean 0 process:

$$S(t)(H(y_i)^{-1}(1-\Delta\Lambda(y_i))^{-1}I[y_i \leq t](1-d_i) - \int_0^t H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i \geq s]d\Lambda(s))$$

Again, we let $\zeta_i \equiv (y_i, x_i, d_i)$, and let $\mathcal{F}(\zeta_i, \zeta_j, \beta)$ denote the kernel of the U-process. Note to show (A.60), it will suffice to show that:

$$\sup_{\|\beta-\beta_0\|\leq n^{-\delta}} \left\| \frac{1}{n(n-1)}\sum_{i\neq j} \mathcal{F}(\zeta_i,\zeta_j,\beta) - E[\mathcal{F}(\zeta_i,\zeta_j,\beta)|\zeta_j] \right\| = o_p(n^{-1/2}) \tag{A.62}$$

$$\sup_{\|\beta-\beta_0\|\leq n^{-\delta}} \left\| \frac{1}{n}\sum_{j=1}^{n} E[\mathcal{F}(\zeta_i,\zeta_j,\beta)|\zeta_j] \right\| = o_p(n^{-1/2}) \tag{A.63}$$

We first show (A.63). Note that $\frac{1}{n}\sum_{j=1}^n E[\mathcal{F}(\zeta_i,\zeta_j,\beta)|\zeta_j]$ can be written as:

$$\frac{1}{n}\sum_{i=1}^{n} \pi \int_{\mathcal{X}} (\mathcal{Q}_i(x'\beta) - \mathcal{Q}_i(x'\beta_0))x dF_X(x)$$

To which we can apply Lemma 2.17 in Pakes and Pollard(1989). We first show the class of functions of y_i, d_i, indexed by β :

$$\left(\int_{\mathcal{X}} \mathcal{Q}_i(x'\beta)x dF_X(x) : \beta \in \mathcal{B} \right) \tag{A.64}$$

is Euclidean for a constant envelope. To do so, we first note the Euclidean property (for a constant envelope) of the class of functions of y_i, d_i, x_i, indexed by β,

$$(\mathcal{Q}_i(x_i'\beta)x_i : \beta \in \mathcal{B})$$

follows from the same arguments used in showing the Euclidean property for the class in (A.34). Thus the class in (A.64) is Euclidean for a constant envelope by Lemma 5 in Sherman(1994). We next show that:

$$E\left[\left\| \int_{\mathcal{X}} (\mathcal{Q}_i(x'\beta) - \mathcal{Q}_i(x'\beta_0))^2 x dF_X(x) \right\| \right] \to 0 \tag{A.65}$$

as $\beta \to \beta_0$. For this it will suffice to show that as $\beta \to \beta_0$:

$$E[|I[x_i'\beta \leq \tau_0] - I[x_i'\beta_0 \leq \tau_0]|] \quad \to \quad 0 \tag{A.66}$$

$$E[|I[c_i \leq x_i'\beta] - I[c_i \leq x_i'\beta_0]|] \quad \to \quad 0 \tag{A.67}$$

$$E\left[\left(\int_{x_i'\beta}^{x_i'\beta_0} H(s)^{-1}(1-\Delta\Lambda(s))^{-1}I[y_i \geq s]d\Lambda(s) \right)^2 \right] \quad \to \quad 0 \tag{A.68}$$

All three of these conditions follow from Assumption RC. This shows (A.64) and hence (A.63). To show (A.62) we note the U-process with kernel $\mathcal{F}(\zeta_i, \zeta_j, \beta) - E[\mathcal{F}(\zeta_i, \zeta_j, \beta)|\zeta_j]$ is degenerate. Similar arguments as above can be used to establish the Euclidean property of this class of functions indexed by $\beta \in \mathcal{B}$, as well as an analogous L^2-continuity condition. Thus (A.63) follows directly from Corollary 8 in Sherman(1994). This shows (A.55). ■

We can now proceed to the main theorem:

Theorem A.2 *(Theorem 3.2 in text) The estimator $\hat{\beta}$ has the following asymptotic linear representation:*

$$\hat{\beta} - \beta_0 = \frac{1}{n}\sum_{i=1}^{n} M_0^{-1}(\psi_i(\beta_0, S) + \xi_i) + o_p(n^{-1/2}) \tag{A.69}$$

Proof: Write $\psi_i(\beta, S)$ as

$$\psi_{1i}(\beta) - \psi_{2i}(\beta)\psi_{3i}(\beta, S)$$

where

$$\psi_{1i}(\beta) = \pi I[y_i > x_i'\beta]x_i \tag{A.70}$$

$$\psi_{2i}(\beta) = (1-\pi)I[y_i \leq x_i'\beta]d_i x_i \tag{A.71}$$

$$\psi_{3i}(\beta, S) = S(x_i'\beta)/S(y_i) \tag{A.72}$$

Rearrange the first order condition:

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i(\hat{\beta}, \hat{S}) = o_p(n^{-1/2}) \tag{A.73}$$

as:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \psi_i(\hat{\beta}, S) + \\
&\frac{1}{n}\sum_{i=1}^{n} \psi_{2i}(\beta_0)(\psi_{3i}(\beta_0, \hat{S}) - \psi_{3i}(\beta_0, S) + \\
&\frac{1}{n}\sum_{i=1}^{n}(\psi_{2i}(\hat{\beta}) - \psi_{2i}(\beta_0))(\psi_{3i}(\beta_0, \hat{S}) - \psi_{3i}(\beta_0, S)) + \\
&\frac{1}{n}\sum_{i=1}^{n} \psi_{2i}(\hat{\beta})(\psi_{3i}(\hat{\beta}, S) - \psi_{3i}(\hat{\beta}, \hat{S}) - \psi_{3i}(\beta_0, \hat{S}) + \psi_{3i}(\beta_0, S)) = o_p(n^{-1/2})
\end{aligned}$$

which by Lemmas 2-7 yields:

$$\frac{1}{n}\sum_{i=1}^{n} \psi_i(\hat{\beta}, S) + \xi_i = o_p(n^{-1/2}) \tag{A.74}$$

so the desired result follows from Theorem A.1 with $\delta = 1/2$ and $\chi_i = \xi_i$. ■

The limiting distribution in Theorem 3.2 follows by applying the Lindeberg-Levy central limit theorem to the linear representation in Theorem A.2.

Table 1: Monte Carlo Results

Student t(1)

	Buckley-James		CLAD ($\pi = 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	1.000	-1.000	1.000
Mean Bias	-1.6000	0.695	-0.005	-0.004	0.009	0.015
Median Bias	-0.648	0.004	-0.011	-0.004	-0.000	0.007
RMSE	9.635	23.761	0.089	0.093	0.116	0.164
MAE	1.611	2.188	0.071	0.074	0.090	0.122

Student t(2)

	Buckley-James		CLAD ($\pi - 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	1.000	-1.000	1.000
Mean Bias	-0.109	-0.000	-0.010	-0.002	0.001	0.008
Median Bias	-0.095	-0.001	-0.017	-0.001	-0.000	0.002
RMSE	0.205	0.191	0.095	0.105	0.097	0.123
MAE	0.146	0.125	0.076	0.083	0.077	0.098

Student t(3)

	Buckley-James		CLAD ($\pi - 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	1.000	-1.000	1.000
Mean Bias	-0.032	0.008	-0.001	0.000	0.010	0.010
Median Bias	-0.035	0.004	-0.002	-0.006	0.007	0.0008
RMSE	0.115	0.005	0.096	0.109	0.095	0.115
MAE	0.091	0.089	0.076	0.087	0.075	0.091

Table 1: Monte Carlo Results (continued)

Standard Normal

	Buckley-James		CLAD ($\pi = 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	0.000	-1.000	1.000
Mean Bias	0.000	0.000	-0.002	-0.007	0.007	0.005
Median Bias	0.001	-0.003	-0.000	-0.009	0.008	0.005
RMSE	0.079	0.083	0.099	0.106	0.085	0.095
MAE	0.062	0.066	0.079	0.085	0.067	0.075

Heteroskedastic Normal, $\sigma^2(x) = \exp(-x)$

	Buckley-James		CLAD ($\pi = 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	1.000	-1.000	1.000
Mean Bias	0.161	0.178	0.004	0.000	0.026	0.020
Median Bias	0.156	0.173	0.003	-0.016	0.021	0.002
RMSE	0.196	0.237	0.113	0.149	0.118	0.221
MAE	0.168	0.196	0.090	0.118	0.083	0.165

Heteroskedastic Normal, $\sigma^2(x) = \exp(x)$

	Buckley-James		CLAD ($\pi - 0.50$)		STLS	
	α	β	α	β	α	β
True Values	-1.000	1.000	-1.000	1.000	-1.000	1.000
Mean Bias	-0.150	-0.224	-0.005	-0.007	0.005	0.004
Median Bias	-0.149	-0.224	-0.001	-0.006	0.007	0.005
RMSE	0.173	0.247	0.113	0.091	0.098	0.083
MAE	0.152	0.225	0.090	0.071	0.078	0.067

Table 2: Estimation Results for Stanford Heart Transplant Data.
157 Observations with Complete Records (Unless Otherwise Noted)

	Constant	Age	Age^2
Ying-Jung Wei[a]	2.731	0.034	-0.0007
	(0.684)	(0.011)	(0.0110)
Buckley-James,[b]	1.35	0.107	-0.0017
152 Observation	(0.71)	(0.037)	(0.0005)
Buckley-James[c]	1.046	0.113	-0.007
	(1.035)	(0.057)	(0.0007)
SCLS[c]	1.132	0.129	-0.0023
	(1.129)	(0.060)	(0.0008)
CRQ, $\pi = 0.50^c$ (CLAD)	1.460	0.123	-0.0021
	(1.446)	(0.078)	(0.0011)
CRQ, $\pi = 0.25^c$	-0.696	0.165	-0.0023
	(1.894)	(0.113)	(0.0015)
CRE, $\pi = 0.75Hc$	1.880	0.090	-0.0013
	(1.028)	(0.060)	(0.0008)

[a] Reported by Ying, Jung, and Wei (1991); standard errors calculated as width of reported 95% confidence intervals, divided by 2 $\times 1.9$

[b] Reported by Miller and Halpern (1982); sample excludes 5 observations with survival times less than 10 days.

[c] Standard errors calculated from bootstrap distribution with $R = 1000$ replications, using median absolute deviation divided by 0.67.