The Annals of Statistics
1997, Vol. 25, No. 3, 1088–1108

A GENERALIZATION OF THE PRODUCT-LIMIT ESTIMATOR WITH AN APPLICATION TO CENSORED REGRESSION[1]

BY SONG YANG

Texas Tech University

The product-limit estimator (PLE) and weighted empirical processes are two important ingredients of almost any censored regression analysis. A link between them is provided by the generalized PLEs introduced in this paper. These generalized PLEs are the product-limit integrals of the empirical cumulative hazard function estimators in which weighted empirical processes are used to replace the standard empirical processes. The weak convergence and some large sample approximations of the generalized PLEs are established. As an application these generalized PLEs are used to define some minimum distance estimators which are shown to be asymptotically normal. These estimators are qualitatively robust. In some submodels an optimal choice of the weight matrix is the covariate matrix and some of these estimators are quite efficient at a few common survival distributions. To implement these estimators some computational aspects are discussed and an algorithm is given. From a real data example and some preliminary simulation results, these estimators seem to be very competitive to and more robust than some more traditional estimators such as the Buckley–James estimator.

1. Introduction. The censored regression model is an important model in survival analysis and has received much attention in the past as well as in the recent literature. In almost all the works in censored regression, a fundamental ingredient is the product-limit estimator (PLE). The PLE was developed by Kaplan and Meier (1958) and is in fact the nonparametric maximum likelihood estimator of the underlying survival function. By now the PLE is almost a standard part of any regression analysis.

In this paper we propose a generalization of the PLE. This generalized PLE is defined via empirical cumulative hazard function estimators using weighted empirical processes. By varying the weights we actually have a class of generalized PLEs. Weighted empirical processes arise naturally in regression analysis and have been exploited successfully in the complete data case [cf. Koul (1992)]. The connection between the PLE and the weighted empirical processes so far has not been demonstrated, despite their being two seemingly natural parts of censored regression analysis. The generalized PLEs proposed in this paper link the two important tools. Potentially the generalized PLEs could be as useful to the censored regression as the weighted empirical processes to the

---

Received November 1995; revised June 1996.

[1]Revision partially supported by NSF Grant DMS-95-04918 and Advance Research Program of Texas Grant 0206-44-3812.

AMS 1991 *subject classifications*. Primary 62G05, 60F05; secondary 62J05.

Key words and phrases. Product-limit estimator, weighted empirical process, censored regression, martingales, asymptotic normality, robustness.

ordinary regression. As an application of these generalized PLEs, we will use them to define and study some Koul (1985) type minimum L_2 distance estimators of the regression parameter. These estimators are qualitatively robust. Under certain submodels, to be specified later, using the covariates matrix as the weights minimizes the asymptotic variance within the class of all suitable weights and some of the estimators are quite efficient at a few common survival distributions. Some goodness-of-fit tests of the model also naturally arise from these estimators. For implementing these estimators, some computational aspects are discussed and an algorithm given. From an application to the Stanford heart transplant data and preliminary simulation results, the minimum distance estimators seem very competitive to and more robust than some more traditional estimators such as the Buckley–James estimator.

We will organize the material as follows. In Section 2 we first discuss the model and related literature. Then the generalized PLEs are introduced. The convergence of these estimators is obtained via the martingale and counting process theory [cf. Gill (1980), Shorack and Wellner (1986)]. Some large sample approximations are also established. In Section 3 the generalized PLEs are used to define the regression estimators. These estimators are shown to be asymptotically normal. Various properties such as the asymptotic efficiency and the choice of the weight matrix are discussed. In Section 4 robustness is briefly discussed. Finally Section 5 discusses some computational issues and an algorithm is applied to the Stanford heart transplant data.

2. The model and the generalized PLEs.

Let

$$Y_i = z_i^* \beta + \varepsilon_i, \qquad i = 1, 2, \dots, n, \tag{1}$$

where the covariates $z_1, \dots, z_n$ are k-dimensional nonrandom vectors, * denotes the transpose, and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. with an unknown cumulative distribution function (c.d.f.) F. In the censored regression model, one only observes $X_i = \min(Y_i, C_i)$ and $\delta_i = [Y_i \le C_i]$, $i = 1, \dots, n$, where $C_1, \dots, C_n$ are independent and $[A]$ denotes the indicator function of an event A. The c.d.f. G_i of C_i is unknown and Y_i, C_i are independent. For some discussion on related models, see Tsiatis (1990) and Kalbfleisch and Prentice (1980).

In the censored regression model, earlier works include some extensions of the well-known least squares method. See Buckley and James (1979), Koul, Susarla and Van Ryzin (1981) and Miller (1976). Except for the Koul–Susarla–Van Ryzin estimator, these extensions of the least squares method require iteration. Applying these three estimators to the Stanford heart transplant data, Miller and Halpern (1982) concluded that the Buckley–James estimator was the most reliable of the three and thus was preferred together with the partial likelihood estimator of Cox (1972) for the proportional hazard model. Later Leurgans (1987) proposed an estimator that is similar in spirit to the Koul–Susarla–Van Ryzin estimator and appears to be competitive with the Buckley–James estimator. Notice that the Buckley and James estimator inherits the nonrobustness of the least squares method and loses the edge of computational simplicity of the original least squares estimator. Recently there also have been some extensions of other familiar procedures to the censored regression model. These include the M-estimation [Ritov (1990), Zhou (1992)] and the rank method [Tsiatis (1990), Lai and Ying (1991), Ying (1993)]. We will introduce some minimum distance estimators in Section 4. The minimum distance method often has certain robustness advantages in the complete data case. See Section 4 for some more discussion.

Let $D = (d_{ij})_{n\times k}$ be a matrix whose elements are, for technical reasons, nonnegative. For $-\infty < t < \infty$ and any fixed $b \in R^k$, define the weighted empirical processes

$$\mathbf{H}^1_{nj}(t;b) = \sum_{i=1}^n d_{ij}\delta_i[X_i - z_i^* b \le t], \qquad \mathbf{K}_{nj}(t;b) = \sum_{i=1}^n d_{ij}[X_i - z_i^* b \ge t].$$

Note that for notational convenience we have suppressed the dependence of $\mathbf{H}^1_{nj}(t;b)$, $\mathbf{K}_{nj}(t;b)$ on D. These weighted empirical processes are the building blocks in our analysis here.

Define the empirical cumulative hazard estimator

$$\hat{\Lambda}_{nj}(t;b) = \int_{-\infty}^t \frac{d\mathbf{H}^1_{nj}(u;b)}{\mathbf{K}_{nj}(u;b)}, \qquad -\infty < t < \infty, \tag{2}$$

where and throughout this paper, $\int_{-\infty}^t = \int_{(-\infty,t]}$ for $t < \infty$. Note that (2) is scale invariant with respect to D.

Now we can define, for each fixed b, the generalized product-limit estimator

$$\hat{F}_{nj}(t;b) = 1 - \prod_{s\le t}^{n}\big(1 - \Delta\hat{\Lambda}_{nj}(s;b)\big), \qquad -\infty < t < \infty, \tag{3}$$

where $\Delta h(s) = h(s) - h_-(s)$ for any function h and the subscript $-$ denotes the left limit. Also $\Delta\hat{\Lambda}_{nj}(s;b)$ is taken to be zero if $\Lambda_{nj}(s;b)$ is not defined. Note that when $d_{1j} = \cdots = d_{nj} = 1$, (2) and (3) reduce to the usual Aalen–Nelson estimator (of the cumulative hazard function) and the Kaplan and Meier estimator (of the lifetime distribution), respectively. For convenience we will write $\hat{F}_{nj}(t) \equiv \hat{F}_{nj}(t;\beta)$.

Let $H^1_{nj} = E\mathbf{H}^1_{nj}$ and $K_{nj} = E\mathbf{K}_{nj}$. For any c.d.f. H, denote $1-H$ by $\bar{H}$. Then

$$\begin{aligned} H^1_{nj}(t;b) &= \sum_{i=1}^n d_{ij}\int_{-\infty}^t \bar{G}_i(u + z_i^* b)\,dF\big(u + z_i^*(b-\beta)\big),\\ K_{nj}(t;b) &= \sum_{i=1}^n d_{ij}\bar{G}_i(t + z_i^* b)\bar{F}\big(t + z_i^*(b-\beta)\big). \end{aligned} \tag{4}$$

Again for convenience we will use $H^1_{nj}(t;\beta) \equiv H^1_{nj}(t)$, and so on. Let $\Lambda(t) \equiv \int_{-\infty}^t \bar{F}_-^{-1}\,dF$ be the cumulative hazard function corresponding to F. A crucial observation is that, for $t < \sup\{u\colon K_{nj}(u) > 0\}$,

$$\int_{-\infty}^t \frac{dH^1_{nj}(u)}{K_{nj}(u)} = \int_{-\infty}^t \frac{(\sum_{i=1}^n d_{ij}\bar{G}_{i-}(u + z_i^*\beta))dF(u)}{\sum_{i=1}^n d_{ij}\bar{G}_{i-}(u + z_i^*\beta)\bar{F}_-(u)} = \Lambda(t),$$

regardless of what the weights d_{ij}'s may be.

By what has now become a standard result in censored data analysis,

$$M_{nj}(t) = \mathbf{H}^1_{nj}(t;\beta) - \int_{-\infty}^{t} \mathbf{K}_{nj}(u;\beta)\, d\Lambda(u), \qquad -\infty < t < \infty,\ j = 1,\dots,k$$

are mean zero square integrable martingales with respect to σ-algebras,

$$\mathscr{F}_t = \sigma\{\delta_i[X_i - z_i^*\beta \le t], [X_i - z_i^*\beta \le t]: i = 1,\dots,n\}, \qquad -\infty < t < \infty.$$

Their predictable covariance processes are

$$\langle M_{nj}, M_{nl}\rangle(t) = \int_{-\infty}^{t} \left(\sum_{i=1}^{n} d_{ij} d_{il} [X_i - z_i^*\beta \ge u] \right) (1 - \Delta\Lambda(u))\, d\Lambda(u),$$

for $-\infty < t < \infty$ and $j, l = 1,\dots,k$. See Corollary 3.1.1 of Gill (1980) or (7.8.10) of Shorack and Wellner (1986) for related assertions.

For $j = 1,\dots,k$ let

$$U_{nj} = \frac{\hat{F}_{nj} - F}{1 - F}, \qquad R_{nj} = \frac{1 - \hat{F}_{nj}}{1 - F}.$$

Define $T_{nj} = \sup\{u: \mathbf{K}_{nj}(u) > 0\}$, $J_{nj}(\cdot) = [\cdot \le T_{nj}]$ and the stopped process $U_{nj}^{T_{nj}}(\cdot) = U_{nj}^{T_{nj}}(T_{nj} \wedge \cdot)$, where $\wedge$ denotes the minimum. Since $\mathbf{K}_{nj}$ is left continuous, J_{nj} is predictable. Thus similarly to (3.2.12) of Gill (1980) or (7.8.17) of Shorack and Wellner (1986), for $j = 1,\dots,k$,

$$U_{nj}^{T_{nj}} = \int_{-\infty}^{t} R_{nj\_} J_{nj} d(\hat{\Lambda}_{nj} - \Lambda) = \int_{-\infty}^{t} R_{nj\_} \frac{J_{nj}}{\mathbf{K}_{nj}}\, dM_{nj},$$

which are mean zero locally square integrable martingales with predictable covariation processes:

$$\langle U_{nj}^{T_{nj}}, U_{nl}^{T_{nl}}\rangle(t) = \int_{-\infty}^{t} \frac{J_{nj} J_{nl} R_{nj-}(u) R_{nl-}(u)}{\mathbf{K}_{nj}(u)\mathbf{K}_{nl}(u)} \left(\sum_{i=1}^{n} d_{ij} d_{il} [X_i - z_i^*\beta \ge u] \right) \times (1 - \Delta\Lambda(u))\, d\Lambda(u), \qquad t < \infty. \tag{5}$$

Let

$$\tau_j = \sqrt{\sum_{i=1}^{n} d_{ij}^2}, \qquad s_j = \sum_{i=1}^{n} d_{ij}, \qquad W_{nj} = \frac{s_j}{\tau_j} U_{nj}, \qquad j = 1,\dots,k. \tag{6}$$

The asymptotic behavior of our regression estimators depends on the asymptotic behavior of the W_{nj}'s, which is given by Theorem 1. Let

$$\Gamma_{nj}(t;b) = \frac{1}{s_j} \sum_{i=1}^{n} d_{ij} \bar{G}_{i-}(t + z_i^* b),$$
$$\Gamma_{ndjl}(t;b) = \frac{1}{\tau_j \tau_l} \sum_{i=1}^{n} d_{ij} d_{il} \bar{G}_{i-}(t + z_i^* b). \tag{7}$$

We will abbreviate $\Gamma_{nj}(t;\beta) \equiv \Gamma_{nj}(t)$, and so on. For later use we will also consider $\hat{F}_{n0}$, W_{n0} and Γ_{n0}, and so on, which are defined using the weights $d_{i0} \equiv 1$.

THEOREM 1. *Suppose for some* $T < \infty$, $\bar{F}(T) > 0$ *and* $\Gamma_{nj}, \Gamma_{ndjl}$ *converge pointwise to some continuous limits* Γ_j, Γ_{djl} *on* $(-\infty, T]$ *with* $\inf_j \Gamma_j(T) > 0$. *Then*

$$(W_{n0}, \dots, W_{nk}) \to (W_0, \dots, W_k) \quad \text{in } D^{k+1}(-\infty, T],$$

where $W_0, \dots, W_k$ *are jointly Gaussian with mean zero and predictable covariance processes*:

$$\langle W_j, W_l \rangle(t) = \int_{-\infty}^{t} \frac{\Gamma_{djl}\bar{F}}{\Gamma_j \Gamma_l \bar{F}_-^3}\, dF, \qquad t < T. \tag{8}$$

PROOF. By the continuity and monotonicity, the convergences of $1 - \Gamma_{nj}$, $1 - \Gamma_{ndjl}$ to Γ_j, Γ_{djl} are uniform. Thus by the law of large numbers, with probability 1,

$$\sup_{t \in (-\infty, T]} \left| \frac{1}{s_j} \mathbf{K}_{nj}(t) - \Gamma_j \bar{F}(t) \right| \to 0$$

and

$$\sup_{t \in (-\infty, T]} \left| \frac{1}{\tau_j \tau_l} \sum_{i=1}^{n} d_{ij} d_{il} [X_i - z_i^* \beta \geq u] - \Gamma_{djl} \bar{F}(t) \right| \to 0.$$

From these and similarly to Theorem 7.3.1 in Shorack and Wellner (1986),

$$\sup_{t \in (-\infty, T]} |\hat{F}_{nj}(t) - F(t)| \to 0.$$

Now Theorem 1 follows from combining these results and (5) in an argument similar to Theorem 4.2.1 of Gill (1980) or Theorem 7.8.1 of Shorack and Wellner (1986). □

To apply the generalized PLEs to censored regression, we establish some large sample approximations next. For that purpose we make the following assumptions. Let $Z = (z_1, \dots, z_n)^*$. We will use $\|\cdot\|$ to denote the Euclidean norm in R^k.

A1. $\max_i \|z_i\| \leq K$ and $n^a = O(\tau_j)$ for some $a > 0$, $K < \infty$ and $j = 1, \dots, k$.

A2. F and G_i have Lebesgue densities f, g_i, respectively. In addition, $\sup_t |f'(t)| \leq C$, $\sup_t |g_i(t)| \leq C$, $i = 1, \dots, n$, for some $C > 0$ and $\sup_i E|Y_i \wedge C_i|^r < \infty$ for some $r > 0$; f' is uniformly continuous and integrable. As $a_n \to 0$, $\int_{-\infty}^{\infty} w(t, a_n)\, dt = o(1)$ where

$$w(t, a_n) = \sup_{|s| < a_n} (|f'(t+s) - f'(t)| + |f(t+s) - f(t)|).$$

A3. For some $T, B < \infty$, $\|\beta\| \leq B$, $\bar{F}(T + 2KB) > 0$ and $\liminf_n \Gamma_{nj}(T; b) > 0$, $j = 0, \dots, k$ for $\Gamma_{nj}(t; b)$ in (7) and $\|b\| \leq B$.

The assumption A2 and the rate assumption in A1 are needed in the approximations in Theorem 2 and Theorem 3. The quantity T in A3, like the T in Theorem 1, corresponds to a usual practice in survival analysis. The purpose is to avoid possible upper tail instability. Also under A3 we do not have to stop U_{nj} at T_{nj} in the asymptotic analysis.

Theorem 2 gives uniform approximations of empirical processes by their expectations. Lai and Ying (1988) used Alexander's (1984) inequality to obtain these results for the case of equal weights. Under the rate assumption in A1, the first part of Theorem 2 can be adapted from their Theorem 1 and the second from their Theorem 3. Thus the proofs will be omitted. Note that we use s_j, τ_j in scaling, rather than n as in Lai and Ying (1988). To avoid technical complications we consider the range $(-\infty, T]$; thus the modification in Lai and Ying (1988) on $\hat{F}_{nj}$ is not needed. Let

$$\Lambda_{nj}(t;b) = \int_{-\infty}^{t} \frac{h_{nj}(u;b)\,du}{K_{nj}(u;b)}, \qquad F_{nj}(t;b) = 1 - \exp\{-\Lambda_{nj}(t;b)\},$$

where $h_{nj}(u;b) = \partial H^1_{nj}(u;b)/\partial u$. We will abbreviate "with probability 1" to w.p.1.

THEOREM 2. *Suppose* A1 *and* A2 *hold. Then for any* $c, \varepsilon > 0$, $0 < r < 1$, *w.p.*1,

$$\sup_{\|b\|<B,\, t\le T} |Q_{nj}(t;b) - EQ_{nj}(t;b)| = o(\tau_j^{1+\varepsilon}),$$

$$\sup_{t\le T,\, \|b-b'\|\le cn^{-r}} |Q_{nj}(t;b) - EQ_{nj}(t;b) - Q_{nj}(t;b') + EQ_{nj}(t;b')| = o(\tau_j^{1-r/2+\varepsilon}),$$

where Q_{nj} *is either* $\mathbf{H}^1_{nj}$ *or* $\mathbf{K}_{nj}$. *Thus when* A3 *also holds*,

$$\sup_{t\le T,\, \|b\|<B} |\hat{F}_{nj}(t;b) - F_{nj}(t;b)| = o\left(\frac{\tau_j^{1+\varepsilon}}{s_j}\right), \tag{9}$$

$$\sup_{t\le T,\, \|b-b'\|\le \tau_j^{-r}} |\hat{F}_{nj}(t;b) - F_{nj}(t;b) - \hat{F}_{nj}(t;b') + F_{nj}(t;b')| = o\left(\frac{\tau_j^{1-r/2+\varepsilon}}{s_j}\right). \tag{10}$$

In Theorem 3 we approximate $F_{nj}(t;\beta+\Delta)$ by a linear function of Δ. For $j = 0, \ldots, k$ and $l = 1, \ldots, k$ let

$$\Gamma_{nzjl}(t;b) = \frac{1}{s_j}\sum_{i=1}^{n} d_{ij} z_{il} \bar{G}_{i-}(t + z_i^* b),$$

$$\Gamma_{nzj\bullet}(t;b) = (\Gamma_{nzj1}(t;b), \ldots, \Gamma_{nzjk}(t;b))^*.$$

Let $\lambda(t) = f(t)/\bar{F}(t)$ be the hazard rate function of F. Then $d\lambda(t) = (f'\bar{F} + f^2)/\bar{F}^2\,dt$. Define the k-dimensional vector functions

$$\xi_j(t) = \bar{F}(t)\int_{-\infty}^{t} \frac{\Gamma_{nzj\bullet}(u;\beta)}{\Gamma_{nj}(u;\beta)}\,d\lambda(u), \qquad t \le T,\ \ j = 0,\dots,k.$$

Also note that $F_{nj}(\cdot;\beta) = F(\cdot)$.

THEOREM 3. *Under* A1 *to* A3, *for* $b_n \to 0$,

$$F_{nj}(t;\beta + b_n) = F(t) + \xi_j^*(t)b_n + o(\|b_n\|),$$

uniformly for $t \le T$.

PROOF. Since f' is uniformly continuous and $\max_i \|z_i\| \le K$, by the Taylor expansion we have

$$\begin{aligned} f(u + z_i^* b_n) &= f(u) + f'(u)z_i^* b_n + O(w(u, K\|b_n\|))z_i^* b_n \\ &= f(u) + f'(u)z_i^* b_n + o(\|b_n\|), \end{aligned}$$

uniformly in u and i. Similarly, $\bar{F}(u + z_i^* b_n) = \bar{F}(u) - f(u)z_i^* b_n + o(\|b_n\|)$, uniformly in u and i. Let

$$R = \frac{f(u)\Gamma_{nj}(u;\beta + b_n) + f'(u)\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n}{\bar{F}(u)\Gamma_{nj}(u;\beta + b_n) - f(u)\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n}.$$

Using the expansions of f and $\bar{F}$ in R, we have

$$\begin{aligned} &\frac{h_{nj}(u;\beta + b_n)}{K_{nj}(u;\beta + b_n)} - R \\ &\quad= \frac{(\bar{F}(u) - f(u))\Gamma_{nj}(u;\beta + b_n) - (f(u) + f'(u))\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n}{\bar{F}(u)\Gamma_{nj}(u;\beta + b_n) - f(u)\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n + \Gamma_{nj}(u;\beta + b_n)o(\|b_n\|)} \\ &\qquad \times \frac{\Gamma_{nj}(u;\beta + b_n)o(\|b_n\|)}{\bar{F}(u)\Gamma_{nj}(u;\beta + b_n) - f(u)\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n} \\ &\quad= r_n(u), \end{aligned}$$

say. Under A3, $\Gamma_{nj}(u;\beta + b_n) > 0$. Dividing both the numerator and the denominator of r_n by $\Gamma_{nj}(u;\beta + b_n)$ and using $\Gamma_{nzjl} \le K\Gamma_{nj}$, and $\bar{F}(T) > 0$, we see that, for $b_n \to 0$, $r_n(u) = o(\|b_n\|)$, uniformly in i and $u < T$. Looking back at the beginning of the proof, we see that the $o(\|b_n\|)$ is actually $O(w(u, K\|b_n\|)) \cdot O(\|b_n\|)$. Thus $\int_{-\infty}^{t} r_n(u)\,du = o(\|b_n\|)$.

Now some algebra gives

$$\begin{aligned} R - \frac{f(u)}{\bar{F}(u)} &= \frac{(f'(u)\bar{F}(u) + f^2(u))\Gamma^*_{nzj\bullet}(u;\beta + b_n)b_n}{\bar{F}^2\Gamma_{nj}(u;\beta + b_n)} \\ &\qquad + (f'(u)\bar{F}(u) + f^2(u))o(\|b_n\|), \end{aligned}$$

uniformly in i and $u < T$. Note that to get the last equality we have again used $\Gamma_{nzjl} \le K\Gamma_j$ and $\bar{F}(T) > 0$. Combining these results, we have

$$\begin{aligned}\Lambda_{nj}(t;\beta+b_n) &= \int_{-\infty}^{t} \frac{f(u)\,du}{\bar{F}(u)} + \frac{f'(u)\bar{F}(u)+f^2(u)}{\bar{F}^2(u)}\frac{\Gamma^*_{nzj\bullet}(u;\beta+b_n)b_n}{\Gamma_{nj}(u;\beta+b_n)}\,du \\ &\quad + o(\|b_n\|) \\ &= \Lambda(t) + \int_{-\infty}^{t} \frac{\Gamma^*_{nzj\bullet}(u;\beta)}{\Gamma_j(u;\beta)}\,d\lambda(u)b_n + o(\|b_n\|),\end{aligned}$$

uniformly for $t \le T$. Note that in obtaining the last equality we have used

$$\int_{-\infty}^{T} \lambda'(u)|\Gamma_{nj}(u;\beta+b_n) - \Gamma_{nj}(u;\beta)|\,du = o(1),$$

$$\int_{-\infty}^{T} \lambda'(u)|\Gamma_{njl}(u;\beta+b_n) - \Gamma_{njl}(u;\beta)|\,du = o(1),$$

$j = 0,\dots,k$, $l = 1,\dots,k$, which follow from A1 to A3. Since F_{nj} is obtained from Λ_{nj} through a simple exponential function, the assertion of Theorem 3 follows. $\square$

3. Estimating the regression parameter. In the complete data case, the minimum L_2 distance estimation has been an important and interesting alternative in parametric estimation, especially in the context of robust inferences. When the underlying distribution has some known parametric form, the minimum L_2 distance estimators are usually asymptotically minimax robust [Beran (1977), (1982), Millar (1983)], have some optimal properties of quantitative robustness [Donoho and Liu (1988a)] and are void of certain pathologies of non-L_2 distance estimators [Donoho and Liu (1988b)]. Also, many minimum L_2 distance estimators are highly efficient at some familiar parametric models for proper choices of the integrating measure [Koul and DeWet (1983), Parr and Schucany (1980)]. Using the generalized PLEs, we now introduce some minimum L_2 distance estimators in the censored regression.

For easier applications, let us modify the generalized PLEs $\hat{F}_{nj}(t;b)$, $j = 0,1,\dots,k$, to be 1 on and after their largest jump points, respectively. Let H be a σ-finite measure that assigns finite measure to any bounded interval. Define the estimation function

$$M_{nd}(b) = \sum_{j=1}^{k} \frac{s_j^2}{\tau_j^2} \int_{-\infty}^{\infty} \left(\hat{F}_{nj}(t;b) - \hat{F}_{n0}(t;b)\right)^2 dH(t). \tag{11}$$

We define the estimator $\hat{\beta}_{nd}$ of β as any minimizer in b, or within $o(1)$ of achieving the minimum, of $M_{nd}(b)$:

$$M_{nd}(\hat{\beta}_{nd}) \le \inf_b M_{nd}(b) + o(1). \tag{12}$$

A large class of estimators are obtained by varying the weights D and the integrating measure H. The choices of D and H will be discussed in the remarks after the proof of Theorem 4.

Note that, when there is no censoring, $\hat{F}_{nj}(t;b)$'s reduce to some weighted empirical processes and $M_{nd}(b)$ reduces to

$$\sum_{j=1}^{k}\int_{-\infty}^{\infty}\biggl(\sum_{i=1}^{n}\frac{1}{\tau_j}(d_{ij}-\bar{d}_{\cdot j})[Y_i-z_i^*b\le t]\biggr)^2\,dH(t).$$

Thus when there is no censoring and for $b=\beta$, each sum $\sum_{i=1}^{n}$ in $M_{nd}(b)$ has expectation zero. The estimators $\hat{\beta}_{nd}$ are extensions to the multiple censored regression of the minimum L_2 distance estimators of Koul (1985) for regression with complete data. Let $Z=(z_1,\dots,z_n)^*$ and $\bar{Z}$ be the $n\times k$ matrix whose row vectors are identically $\bar{z}=\sum_{i=1}^{n}z_i$. Koul (1985) used the weights $((Z-\bar{Z})^*(Z-\bar{Z}))^{-1/2}z_i$'s, which may be negative even if z_i's are nonnegative. Thus these weights cannot be used here in the censored regression.

To study the asymptotic properties of $\hat{\beta}_{nd}$ we first give some uniform approximations of $M_{nd}(b)$ in local neighborhoods of β. We make the following additional assumption.

A4. $(Z-\bar{Z})^*(Z-\bar{Z})$ is invertible. For $A_n=((Z-\bar{Z})^*(Z-\bar{Z}))^{-1/2}$ and $j=1,\dots,k$, $(\tau_j/s_j)A_n^{-1}$ converges to a positive definite matrix. H is a finite measure with support in $(-\infty,T]$.

The assumption on A_n gives some control over the norm of $(\tau_j/s_j)A_n^{-1}$. If $(z_1,\dots,z_n)$ is a realization of i.i.d. copies of a random vector z, then this convergence assumption is satisfied with probability 1 if $\mathrm{Cov}(z,z^*)$ is positive definite. The assumption on H makes the integral small when the integrand is small.

Now recall W_{nj} in (6) and ξ_j in Theorem 3. For $j=1,\dots,k$ let

$$\eta_j(t)=\frac{s_j}{\tau_j}(\xi_j(t)-\xi_0(t)),\qquad V_{nj}(t)=\frac{s_j}{\tau_j}(\hat{F}_{nj}(t)-\hat{F}_{n0}(t))$$

and

$$\tilde{M}(b)=\sum_{j=1}^{k}\int_{-\infty}^{\infty}\{V_{nj}(t)+\eta_j^*A_nb\}^2\,dH(t).$$

Then $\tilde{M}(b)$ is quadratic in b and has the minimizer

$$\biggl(\sum_{j=1}^{k}\int_{-\infty}^{\infty}A_n\eta_j\eta_j^*A_n\,dH(t)\biggr)^{-1}\sum_{j=1}^{k}\int_{-\infty}^{\infty}A_n\eta_j(t)V_{nj}(t)\,dH(t).$$

Some approximations of M_{nd} are given in the following result.

LEMMA 1. *Suppose* A1 *to* A3 *hold. Then for any* $b_n\to 0$, *w.p.*1,

$$M_{nd}(\beta+b_n)=\sum_{j=1}^{k}\int_{-\infty}^{\infty}\biggl\{V_{nj}(t)+\eta_j^*b_n+o\biggl(1+\frac{s_j}{\tau_j}\|b_n\|\biggr)\biggr\}^2\,dH(t),\tag{13}$$

where $o(\cdot)$ is uniform in t. Hence if A4 *also holds, then for any $L > 0$, w.p.1,*

$$\sup_{\|\Delta\|<L} |M_{nd}(\beta + A_n\Delta) - \tilde{M}(\Delta)| = o(1). \tag{14}$$

PROOF. First we will derive a more convenient form of approximation from Theorem 2. Note that $\tau_j^2 \le Ks_j$. For a fixed $0 < r < 1$, if $\|b_n\| > \tau_j^r$, then

$$\frac{\tau_j^{1+(1-r)}}{s_j} = \frac{\tau_j^2}{s_j\tau_j^r} \le B\tau_j^{-r} < B\|b_n\|.$$

If $\|b_n\| \le \tau_j^{-r}$, then for $\varepsilon < r/2$, $-r/2+\varepsilon < 0$ and so $\tau_j^{-r/2+\varepsilon} = o(1)$ under A1. Thus from (9) and (10) in Theorem 2 we always have, for $b_n \to 0$, w.p.1,

$$\begin{aligned}&\sup_{t\le T}\left|\hat{F}_{nj}(t;\beta+b_n) - F_{nj}(t;\beta+b_n) - \hat{F}_{nj}(t;\beta) + F_{nj}(t;\beta)\right|\\ &\qquad = o\left(\frac{\tau_j}{s_j} + \|b_n\|\right).\end{aligned} \tag{15}$$

Now from (15) and Theorem 3 we have, w.p.1,

$$\begin{aligned}&M_{nd}(\beta+b_n)\\ &\quad = \sum_{j=1}^k \int_{-\infty}^{\infty}\left\{\frac{s_j}{\tau_j}\big(\hat{F}_{nj}(t;\beta) - F(t) + \hat{F}_{nj}(-t;\beta) - F(-t)\big)\right.\\ &\qquad\qquad \left. + o\left(1+\frac{s_j}{\tau_j}\|b_n\|\right) + \frac{s_j}{\tau_j}(\xi_j(t)-\xi_0(t))^* b_n + o\left(\frac{s_j}{\tau_j}\|b_n\|\right)\right\}^2 dH(t)\\ &\quad = \sum_{j=1}^k \int_{-\infty}^{\infty}\left\{V_{nj}(t) + o\left(1+\frac{s_j}{\tau_j}\|b_n\|\right) + \eta_j^* b_n\right\}^2 dH(t),\end{aligned}$$

where $o(\cdot)$ is uniform in t. This proves the first assertion of Lemma 1.

By A4, the smallest and largest eigenvalues of $(\tau_j/s_j)A_n^{-1}$ converge to two positive numbers, respectively. Thus for some $c_1, c_2 > 0$ and large n,

$$c_1\|\Delta\| \le \frac{\tau_j}{s_j}\|A_n^{-1}\Delta\| \le c_2\|\Delta\| \quad \text{or} \quad c_1\|A_n\Delta\| \le \frac{\tau_j}{s_j}\|\Delta\| \le c_2\|A_n\Delta\|, \tag{16}$$

for any Δ. Note that since $\tau_j \le Ks_j$, we have $(\tau_j/s_j) \le (K/\tau_j) = o(1)$ under A1. Thus for fixed $\|\Delta\| < L$, $\|A_n\Delta\| \to 0$. Now the second assertion of Lemma 1 follows from taking $b_n = A_n\Delta$ in (13). □

Using the convergence in Theorem 1 and the approximations in Theorems 2 and 3 and Lemma 1, we can obtain the asymptotic normality of the estimator $\hat{\beta}_{nd}$.

THEOREM 4. (a) *Suppose* A1 *to* A3 *hold and, for some* $\varepsilon > 0$, $\alpha > 0$,

$$\liminf_n \inf_{\|b\|>\alpha} \sum_{j=1}^k \int_{-\infty}^{\infty} \left\{ \frac{s_j}{\tau_j^{1+\varepsilon}} (F_{nj}(t;\beta+b) - F_{n0}(t;\beta+b)) \right\}^2 dH(t) > 0. \tag{17}$$

Then for any $\hat{\beta}_{nd}$ *in* (12), $\|\hat{\beta}_{nd} - \beta\| = o_p(1)$.

(b) *Suppose* A1 *to* A4 *hold. In addition, suppose* $\|\hat{\beta}_{nd} - \beta\| = o_p(1)$,

$$\sum_{j=1}^k \int_{-\infty}^{\infty} \frac{\tau_j^2}{s_j^2} \eta_j \eta_j^*(t)\, dH(t) \text{ converges to a positive definite matrix} \tag{18}$$

and

$$\Sigma_d = \lim_n \Sigma_{d1}^{-1} \Sigma_{d2} \Sigma_{d1}^{-1} \text{ exists,} \tag{19}$$

where

$$\Sigma_{d1} = \sum_{j=1}^k \int A_n \eta_j \eta_j^* A_n \, dH(t),$$

$$\Sigma_{d2} = 2 \sum_{j,l=1}^k \int_{-\infty}^{\infty} \int_{-\infty}^{t} A_n \eta_j(s) \eta_l^*(t) A_n \cdot C(s) \bar{F}(s) \bar{F}(t)\, dH(s)\, dH(t),$$

$$\begin{aligned} C(t) &= \left\langle W_j - \frac{s_j}{\tau_j \sqrt{n}} W_0, W_l - \frac{s_j}{\tau_j \sqrt{n}} W_0 \right\rangle(t) \\ &= \int_{-\infty}^{T} \left(\frac{\Gamma_{jl}}{\Gamma_j \Gamma_l} - \frac{s_j}{\tau_j \sqrt{n}} \frac{\Gamma_{0l}}{\Gamma_0 \Gamma_l} - \frac{s_l}{\tau_l \sqrt{n}} \frac{\Gamma_{0j}}{\Gamma_0 \Gamma_j} + \frac{s_j s_l}{n \tau_j \tau_l} \frac{\Gamma_{00}}{\Gamma_0 \Gamma_0} \right) \frac{dF}{\bar{F}^2}. \end{aligned}$$

Then

$$A_n^{-1}(\hat{\beta}_{nd} - \beta) \to_D N(0, \Sigma_d).$$

PROOF. (a) By Theorem 2, w.p.1,

$$\begin{aligned} &\frac{1}{\min_j \tau_j^{2\varepsilon}} M_{nd}(\beta+b) \\ &\quad \geq \sum_{j=1}^k \int_{-\infty}^{\infty} \left\{ \frac{s_j}{\tau_j^{1+\varepsilon}} \big(\hat{F}_{nj}(t;\beta+b) - \hat{F}_{n0}(t;\beta+b)\big) \right\}^2 dH(t) \\ &\quad = o(1) + \sum_{j=1}^k \int_{-\infty}^{\infty} \left\{ \frac{s_j}{\tau_j^{1+\varepsilon}} (F_{nj}(t;\beta+b) - F_{n0}(t;\beta+b)) \right\}^2 dH(t), \end{aligned}$$

uniformly in b. But $M_{nd}(\beta) = \sum_{j=1}^k \int_{-\infty}^{\infty} V_{nj}^2(t) dH(t) = O_p(1)$ by Theorem 1. Thus

$$\frac{1}{\min_j \tau_j^{2\varepsilon}} M_{nd}(\beta) = \frac{1}{\min_j \tau_j^{2\varepsilon}} O_p(1) = o_p(1).$$

Therefore the assertion follows.

(b) From (a) and (13) we have

$$M_{nd}(\hat{\beta}_{nd}) = \sum_{j=1}^{k} \int_{-\infty}^{\infty} \left(V_{nj}(t) + \eta_j^*(\hat{\beta}_{nd} - \beta) + o_p\left(1 + \frac{s_j}{\tau_j}\|\hat{\beta}_{nd} - \beta\|\right)\right)^2 dH(t)$$

$$= \sum_{j=1}^{k} \int_{-\infty}^{\infty} \left(V_{nj}(t) + o_p(1 + \|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|) + \eta_j^*(\hat{\beta}_{nd} - \beta)\right)^2 dH(t),$$

where $o_p(1)$ and $o_p(\|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|)$ are uniform in t. Now using $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$, we have, for large n,

$$\begin{aligned} M_{nd}(\hat{\beta}_{nd}) & \\ &\geq \sum_{j=1}^{k} \frac{1}{2}(\hat{\beta}_{nd} - \beta)^* \int_{-\infty}^{\infty} \eta_j \eta_j^*(t)\, dH(t)(\hat{\beta}_{nd} - \beta) \\ &\quad - \sum_{j=1}^{k} \int_{-\infty}^{\infty} \left(V_{nj}(t) + o_p(1) + o_p(\|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|)\right)^2 dH(t) \\ &\geq c\left\|\frac{s_j}{\tau_j}(\hat{\beta}_{nd} - \beta)\right\|^2 - \sum_{j=1}^{k} \int_{-\infty}^{\infty} \{V_{nj}(t) + o_p(1 + \|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|)\}^2\, dH(t) \\ &\geq \frac{c}{c_2^2}\|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|^2 - \sum_{j=1}^{k} \int_{-\infty}^{\infty} \{V_{nj}(t) + o_p(1 + \|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|)\}^2\, dH(t), \end{aligned}$$

where to obtain the second last and the last inequalities we have used (18) and (16), respectively. From this and the Cauchy–Schwarz inequality,

$$\begin{aligned} &\frac{c}{c_2^2}\|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|^2 \\ &\qquad \leq M_{nd}(\hat{\beta}_{nd}) + 2\sum_{j=1}^{k} \int_{-\infty}^{\infty} V_{dj}^2(t)\, dH(t) + o_p(1 + \|A_n^{-1}(\hat{\beta}_{nd} - \beta)\|^2). \end{aligned}$$

Since $M_{nd}(\hat{\beta}) \leq M_{nd}(\beta) + o(1) = O_p(1)$, it follows that $\|A_n^{-1}(\hat{\beta}_{nd} - \beta)\| = O_p(1)$. Thus by (14) we get

$$\begin{aligned} A_n^{-1}(\hat{\beta}_{nd} - \beta) &= \left(\sum_{j=1}^{k} \int_{-\infty}^{\infty} A_n \eta_j \eta_j^* A_n\, dH(t)\right)^{-1} \\ &\quad \times \sum_{j=1}^{k} \int_{-\infty}^{\infty} A_n \eta_j(t) V_{nj}(t)\, dH(t) + o_p(1). \end{aligned}$$

Note that

$$V_{nj}(t) = \bar{F}(t)(W_{nj}(t) - \frac{s_j}{\tau_j\sqrt{n}} W_{n0}(t)).$$

Thus by Theorem 1 we have

$$A_n^{-1}(\hat{\beta}_{nd} - \beta) \to_D N(0, \Sigma_d),$$

for Σ_d in (19).

REMARKS. (i) *Random covariates*. Sometimes the covariates $z_1, \dots, z_n$ are assumed to be i.i.d. random vectors. In this case the theory for the generalized PL estimators and the Koul type estimators is still valid with proper modification.

(ii) *The weights D*. To better understand the choice of D, let us look at the special case when

$$\bar{G}_i(t + z_i^* \beta) \equiv \bar{F}^{\alpha}(t), \qquad i = 1, \dots, n, \tag{20}$$

for some $\alpha \geq 0$. These conditions can arise from certain reliability models and are often used to simplify the conditions in censored regression analysis [Chen, Hollander and Langberg (1982)]. The uncensored complete data case corresponds to $\alpha = 0$. Under (20), η_j reduces to the jth column of $f(Z - \bar{Z})^* D_1$, where $D_1 = D \operatorname{diag}\{1/\tau_1, \dots, 1/\tau_k\}$ and $C(t)$ reduces to the (j, l) element of $\int_{-\infty}^{T} \bar{F}^{-(2+\alpha)}\, dF (D_1 - \bar{D}_1)^*(D_1 - \bar{D}_1)$. Thus

$$\sum_{j=1}^{k} A_n \eta_j \eta_j^* A_n = f^2 A_n (Z - \bar{Z})^* D_1 D_1^* (Z - \bar{Z}) A_n$$

and $\sum_{j=1}^{k} A_n \eta_j(s) \eta_j^*(t) A_n C(s, t)$ reduces to

$$\begin{aligned} &\frac{1}{1+\alpha} f(s) f(t) \bar{F}(s) \bar{F}(t) (\bar{F}^{-(1+\alpha)} - 1) A_n (Z - \bar{Z})^* \\ &\qquad \times (D_1 - \bar{D}_1)^*(D_1 - \bar{D}_1) D_1^* (Z - \bar{Z}) A_n. \end{aligned}$$

If $(D^*(Z - \bar{Z}))^{-1}$ exists, then

$$\Sigma_d = \lim_n \sigma^2 A_n^{-1} (D_1^*(Z - \bar{Z}))^{-1} (D_1^* D_1) (Z^* D_1)^{-1} A_n^{-1},$$

where

$$\begin{aligned} \sigma^2 = {} & \left(\int_{-\infty}^{\infty} f^2\, dH \right)^{-2} \\ & \times \int_{-\infty}^{\infty} \int_{-\infty}^{t} \frac{1}{1+\alpha} f(s) f(t) \bar{F}(s) \bar{F}(t) (\bar{F}^{-(1+\alpha)} - 1)\, dH(s)\, dH(t). \end{aligned} \tag{21}$$

For nonnegative definite matrices $A,\ B$, write $A \geq B$ if $A - B$ is nonnegative definite. As in Koul (1985), from the multivariate Cauchy–Schwarz inequality we get [note that $\bar{D}^*(Z - \bar{Z}) = (0)_{k \times k}$],

$$(D_1 - \bar{D}_1)^*(D_1 - \bar{D}_1) \geq D_1^*(Z - \bar{Z}) A_n^2 (Z - \bar{Z})^* D_1 \quad \text{or} \quad \Sigma_d \geq \sigma^2 I_{k \times k},$$

with equality if $D_1 = ZM$ for some invertible M. Hence if the covariates z_i's are nonnegative, the choice $D = Z$ will minimize the variance of $b^* A_n^{-1}(\hat{\beta}_{nd} -$

β), for any $b \in R^k$. In survival analysis, often the covariates, such as the age, blood pressure of the patients, gender or treatment indicator, are nonnegative. In general, the nonnegativity can be achieved by transforming the parameter space.

(iii) *Asymptotic efficiency*. To better understand the asymptotic efficiency of $\hat{\beta}_{nd}$, again suppose (20) holds. Then the proper version of the asymptotic normality of the rank or M-estimator $\hat{\beta}$ [cf. Ritov (1990), Tsiatis (1990), Lai and Ying (1991)] with efficient score and integration range $(-\infty, T]$ is

$$A_n^{-1}(\hat{\beta} - \beta) \to_D N(0, \sigma_e^2 I_{k\times k}),$$

where

$$\sigma_e^2 = \left(\int_{-\infty}^{T} \left(\frac{\lambda'^2}{\lambda} \bar{F}^{1+\alpha}\right)(t)\, dt\right)^{-1}.$$

With the use of some "tampering" functions, Lai and Ying (1991) allow $T = \infty$. To adaptively estimate the efficient score function, the density f and its derivative f' often have to be estimated by splitting the sample into two parts.

Let σ^2 be as in (21), with H the Lebesgue measure restricted to $(-\infty, T]$. For the purpose of comparison we take $T = \infty$ in σ^2 and σ_e^2. For some of the common error distributions in survival analysis, Table 1 gives the numerical values of σ_e^2 and σ^2 for $\alpha = 0$, 1/3, 1/2 and 1, corresponding to 0%, 25%, 33% and 50% censoring, respectively. Where available, the exact theoretical values are given in the parentheses. The densities of Weibull(c) and gamma(c) are $cx^{c-1}\exp(-x^c)$ and $\Gamma^{-1}(c)x^{c-1}\exp(-x)$, respectively, and log(Weibull(c)) denotes the natural logarithm transform of the Weibull(c) distribution, and so on.

Table 1 shows that, under (20), the asymptotic efficiency of the minimum distance estimators with $dH(t) = dt$ is quite high at normal and logistic survival distributions. For the other distributions, the efficiency often increases as the censoring is increased. In some preliminary simulation studies the minimum distance estimators also appeared highly efficient. Note that these estimators require neither estimating the densities nor splitting the sample. Thus they are easier to compute and may have more stable finite sample behavior. The small sacrifice in the asymptotic efficiency may well be compensated for.

(iv) *Inference and goodness-of-fit tests*. For inference on β, a Wald-type statistic requires a consistent estimator of the asymptotic covariance matrix Σ_d. The matrix Σ_d contains the density function f and is difficult to estimate reliably in the presence of censoring. There are some possible resampling alternatives to avoid this difficulty. One is the one sample bootstrap, with the individual data points being (X_i, z_i). Another is to use the limiting behavior of $M_{nd}(\beta)$, which involves integrals of squared transformed Brownian bridges. Orthogonal decomposition of the covariance processes [cf. Loéve (1963), page 478] shows that $M_{nd}(\beta)$ is asymptotically an infinite quadratic form of standard normal random variables. Numerically simulating the estimated quadratic form provides a basis for inference on β. A third possibility is to directly simulate the involved transformed Brownian bridges.

TABLE 1
Comparison of the factors σ^2 and σ_e^2 at various error distributions and under the submodel (20)

	$\boldsymbol{\sigma^2}$	$\boldsymbol{\sigma_e^2}$		$\boldsymbol{\sigma^2}$	$\boldsymbol{\sigma_e^2}$
Normal			log(Weibull(1))		
α=0	1.0953	1.0007(1)	$\alpha = 0$	1.3140	1.0051(1)
α=1/3	1.2259	1.1281	$\alpha = 1/3$	1.5006	1.3423
α=1/2	1.3016	1.1873	$\alpha = 1/2$	1.6111	1.5113
α=1	1.5851	1.3521	$\alpha = 1$	2.0391	2.0201
Logistic			log(Weibull(2))		
α=0	3.0548	3.0000(3)	$\alpha = 0$	0.3285	0.2500(1/4)
α=1/3	3.4201	3.3333	$\alpha = 1/3$	0.3752	0.3333
α=1/2	3.6315	3.5000	$\alpha = 1/2$	0.4028	0.3750
α=1	4.4180	4.0000	$\alpha = 1$	0.5098	0.5000
Weibull(1)			log(Gamma(2))		
$\alpha = 0$	0.2061	∞	$\alpha = 0$	0.6013	0.5002(1/2)
Weibull(2)			$\alpha = 1/3$	0.6823	0.6311
$\alpha = 0$	0.2194	∞	$\alpha = 1/2$	0.7300	0.6944
Weibull(3)			$\alpha = 1$	0.9128	0.8782
$\alpha = 0$	0.1214	0.0955			
$\alpha = 1/3$	0.1355	0.1053			
$\alpha = 1/2$	0.1438	0.1096			
$\alpha = 1$	0.1745	0.1210			

To test the goodness-of-fit of the model, we can use $M_{nd}(\hat\beta_{nd})$, whose limiting behavior again involves integrals of squared transformed Brownian bridges. Based on their asymptotic distributions, some goodness-of-fit tests can be obtained. In these tests, $\hat\beta_{nd}$ can also be replaced by the Buckley–James estimator, rank estimator or M-estimator, and so on, though certain simplifications due to using $\hat\beta_{nd}$ will be lost. The asymptotic properties of the resampling inference procedures and the goodness-of-fit tests are worthy of further investigation.

(v) *About the conditions* (17) *and* (18). Let us verify (17) and (18) in some special cases. Suppose the range of z_i's does not depend on n and contains k independent vectors: $z_i = e_m$, $i = n_{m-1}+1, \dots, n_m$, $0 = n_0 < n1 < \cdots < n_l = n$, where $l \geq k$ and $e_1, \dots, e_k$ are independent. This covers the k-sample case. For the sake of clarity, we illustrate the two sample cases in detail. Let $z_i = e_m$, $i = n_{m-1} + 1, \dots, n_m$, $m = 1, 2$, $0 = n_0 < n_1 < n_2 = n$, Suppose $d_{ij} = 0$, $i \notin \{n_{j-1} + 1, \dots, n_j\}$. Then we have, for any t, b,

$$F_{n1}(t; b) = 1 - \exp\left\{-\int_{-\infty}^{T} \frac{f(u + e_2^* b)}{\bar F(u + e_2^* b)}\, du\right\},$$

$$F_{n1}(t; b) - F_{n0}(t; b) = \bar F_{n1}(t; b) \exp\left\{\int_{-\infty}^{T} \frac{(\lambda_1 - \lambda_0) S_0 \bar F_0}{S_0 \bar F_0 + S_1 \bar F_1}\, du\right\},$$

where

$$f_0 = f(u + e_1^*(b - \beta)), \qquad S_0 = \frac{1}{n} \sum_{i=1}^{n_1} \bar G_i(u + e_1^* b),$$

and so on. Thus, if $\lim_n S_0$, and so on, exist, uniformly in u and b so that the limits can be taken inside the inf and the integral in (17), and the hazard rate λ is not periodic, then (17) is satisfied.

For (18), using $\Gamma_{nzj\bullet}(u;\beta)/\Gamma_{nj}(u;\beta) = e_j$, $u < T$, we have

$$\xi_j(t) = f(t)e_j, \qquad \eta_j(t) = 2f(t)e_j.$$

Hence (18) is satisfied if $\int_{-\infty}^{\infty} f^2\,dH > 0$. Also, under (20) the matrix in (18) reduces to

$$\left(\int_{-\infty}^{\infty} f^2\,dH\right)(Z-\bar{Z})^* D_2 D_2^*(Z-\bar{Z}), \tag{22}$$

where $D_2 = D\,\mathrm{diag}\{1/s_1,\dots,1/s_k\}$. Thus (18) is satisfied if the matrix (22) converges to a positive definite matrix.

(vi) *A data dependent H*. There are many other choices of H in addition to that in part (iii) above. For many parametric models, the choice $H = F$ has high asymptotic efficiency [cf. Remark 3.4 of Koul and DeWet (1983)]. Thus we are motivated to look at the estimation function

$$\tilde{M}_{nd}(b) = \sum_{j=1}^{k} \frac{s_j^2}{\tau_j^2} \int_{-\infty}^{T} \left(\hat{F}_{nj}(t;b) - \hat{F}_{n0}(t;b)\right)^2 d\hat{F}_{n0}(t;b) \tag{23}$$

and define $\tilde{\beta}_{nd}$ analogously to (12). The asymptotic analysis is similar. Under conditions similar to those in Theorem 4, we have

$$A_{nd}^{-1}(\tilde{\beta}_{nd} - \beta) \to_D N(0, \tilde{\Sigma}_d),$$

where $\tilde{\Sigma}_d$ is Σ_d with H being the restriction of F on $(-\infty, T]$.

4. Robustness. As mentioned in the introduction, minimum L_2 distance estimation has various robustness properties in the i.i.d. complete data case. It is expected that some of those properties are still present in the censored regression. Intuitively, small contamination in the model or data does not severely change the estimation function (11). However, quantitative or even qualitative robustness analysis is difficult. We have the following limited result.

THEOREM 5. *Suppose the following statements hold*:

(i) *z_i's are as in Theorem* 5(a) *and the c.d.f. of Y_i is F_{mn}, $i = n_{m-1} + 1, \dots, n_m$, $m = 1, \dots, l$;*

(ii) *with F replaced by F_{mn}'s,* A1 *to* A3 *are satisfied uniformly in n.*

(iii) $\sup_{t,m} |F_{mn}(t) - F(t - e_m^*\beta)| = o(1)$.

If $d_{ij} = 0$, $i \notin \{n_{j-1}+1, \dots, n_j\}$, then $\|\hat{\beta}_{nd} - \beta\| = o_p(1)$.

PROOF. First note that Theorem 2 is still valid. As in (v) after the proof of Theorem 4, we also have

$$F_{nj}(t;\beta+b)=F_{jn}(t+e_j^*(\beta+b)).$$

Thus the discussion in (v) and assumption (ii) of Theorem 5 give the result. □

5. Computational aspects and an application to the Stanford data. As discussed in Section 3, when the covariates are nonnegative, the weight matrix $D = Z$ is desirable in certain situations. Using a linear transform if necessary, we may assume the covariates to be nonnegative for numerical computations.

From the definitions of $\hat{\beta}_{nd}$ and $\tilde{\beta}_{nd}$, it seems that the search range is the entire R^k space. Practically we can reduce this to a compact neighborhood of β. However, it may not be easy to determine the size and shape of this neighborhood, two of the factors that affect the computational efficiency. We first discuss a proper transform that helps resolve this issue.

Note that, in the random design case of model (1), if Y is standardized and z standardized and orthogonalized, then β is inside the unit sphere. This is because

$$1=\mathrm{Var}(Y)>\mathrm{Var}(z^*\beta)=\beta^*\beta.$$

Thus for practical purposes we can restrict the search range to the unit sphere if Y_i's and z_i's are standardized and orthogonalized using the sample covariances. This standardization and orthogonalization is equivalent to using a proper transform $\tilde{Z}$ of the covariates: $Z=\tilde{Z}M$, where M is a $k\times k$ matrix. Such $\tilde{Z}$ will no longer be nonnegative; thus we still use Z. But now we know that we can reduce the search range to $\beta = M\gamma$, where γ is in the unit sphere. If we have an initial guess or estimator $\beta_0 = M\gamma_0$ from a grid search or other methods, then it seems reasonable to further reduce the range to within $S(\gamma_0)$, the sphere centered at γ_0 and inscribed to the unit sphere.

The estimating function M_{nd} in (12) may be nondifferentiable or even discontinuous. The estimating function $\tilde{M}_{nd}$ in (23) also has many local minimums. Thus algorithms such as Newton's method, the steepest descent or the conjugate gradient method [cf. Press, Flannery, Teukolsky and Vetterling (1986)] are not very reliable for finding $\hat{\beta}_{nd}$ and $\tilde{\beta}_{nd}$. For general optimization problems, Monte Carlo methods have often been used, such as the random search, multistart random search, adaptive random search and simulated annealing algorithms. See Rubinstein (1986), among others, for more discussions. These algorithms are for general minimization purposes and do not use particular features of the function to be minimized. Exploiting the limiting quadratic nature of M_{nd} and $\tilde{M}_{nd}$ and flexible definitions of $\hat{\beta}_{nd}$ and $\tilde{\beta}_{nd}$, we describe an algorithm that works reasonably well and then briefly report its application to the Stanford heart transplant data. The basic idea of the following algorithm is the multistage random search with the added twist of quadratic fitting, as follows:

1. Choose a starting point γ_0. This could be (0,0) or an initial estimator in the unit sphere.
2. Sample $\gamma_i, i = 1, \dots, p$ from $N(\gamma_0, s^2 I_{k\times k})$. Evaluate M_{nd} at γ_0 and those γ_i's within the unit sphere. Then use the least squares method and the $p_1 \le p+1$ points with smallest function values to fit
$$Q(\gamma) = \gamma^* U\gamma + v^*\gamma + w,$$
where U is a $k \times k$ symmetric matrix, $v \in R^k$ and $w \in R^1$. Obtain the estimated minimizer $\gamma' = -U^{-1}v/2$ of $Q(\gamma)$.
3. Let
$$\gamma_0 = \arg\min_{\gamma_0,\, \gamma',\, \gamma_i} M_{nd}$$
and go to 2 with smaller s and p. Repeat q times after adequate precision in γ_0 and/or the relative stable value of M_{nd} is achieved.

As discussed at the beginning of this section, if there is a reasonable initial estimator, in Step 2 we can reduce the random search to those γ_i's in $S(\gamma_0)$. The normal sampling variance s^2 should be such that many sample points will be within the unit sphere but not too clustered around γ_0 (unless γ_0 is a very good initial estimator).The choice $s = 1/6$ seemed to work well. Other distributions may also be used. The first time the sample size p should be large so as to obtain enough points around $\hat{\beta}$, since the approximate quadraticity is only true near it. Similarly, p_1 should be small so as not to include points too far away from $\hat{\beta}$. For successive iterations, $s_{\text{new}} = s/3$ worked well. The number q of iterations does not have to be very large, for often γ_0 and the function value are much stabilized in about five iterations. Often the quadratic fit did improve the search.

Now we report the result of the above algorithm applied to the Stanford heart transplant data. For the data set and the background, see Miller and Halpern (1982). As in that paper, we will use the model $\beta_1(age) + \beta_2(age)^2$ for $\log_{10}$ of the survival time (in days). There were 157 patients with complete records. Among these, 55 were alive, that is, were censored, as of February 1980. The other 102 survival times were uncensored. In various analyses, the five patients with survival times less than ten days were usually deleted [cf. Miller and Halpern (1982)].

Let $\bar{a}$ be the mean age of the 152 patients, s_x^2 the variance of X_i's and $\Sigma = (s_{ij})_{2\times 2}$ the covariance matrix of $age - \bar{a}$ and $(age - \bar{a})^2$. Let $d = \det(\Sigma)$. Then a proper transform matrix is

$$M = \frac{s_x}{\sqrt{s_{11}d}} \begin{pmatrix} \sqrt{d} & -(s_{12} + 2\bar{a}s_{11}) \\ 0 & s_{11} \end{pmatrix}.$$

Consider $\hat{\beta}$ that corresponds to $H =$ Lebesgue, the weight matrix $D = Z$ and $T = \infty$ (the results for finite T were similar). The above algorithm was run in Matlab. In running the program it turned out that the computing time, measured in number of floating point operations, was dominated by that of evaluating the function. Thus we tried not to evaluate the function too many times. In each run we started from $(0, 0)$. The successive p's in step 2 above were 100, 40, 20, 20, 20 (five iterations) and $p_1 = 8$ each time. In various runs, within three to five iterations, γ had the first three significant digits stable and M_{nd} the first four. Thus, in view of the size $n = 152$ of the data set, satisfactory answers were obtained after five iterations, which evaluate the function about 200 times. Evaluating the function $200 \sim 250$ times consumes about the same computing time as that of using the *fmins* procedure in Matlab once, in either computing $\hat{\beta}$ or the partial likelihood estimator in the proportional hazard model. Figure 1 gives the contours of M_{nd} based on grids −1: 0.1: 1 with range 0.13: 0.01: 0.3. The small circle inscribed is centered at the naive least squares estimator, pretending all observations were uncensored. The figure shows that there are no other minimizers in the unit circle and also that the search could be restricted to the smaller circle.

For comparison, the Buckley–James estimator and Cox's partial likelihood estimator (for the proportional hazard model) were computed. Also computed were the estimator $\tilde{\beta}$ in (23), with $T = \infty$ and $D = Z$. Moreover, to investigate robustness, these estimators were computed for the 152 and 157 patients, that is, excluding or including the five short survival times. The results from one

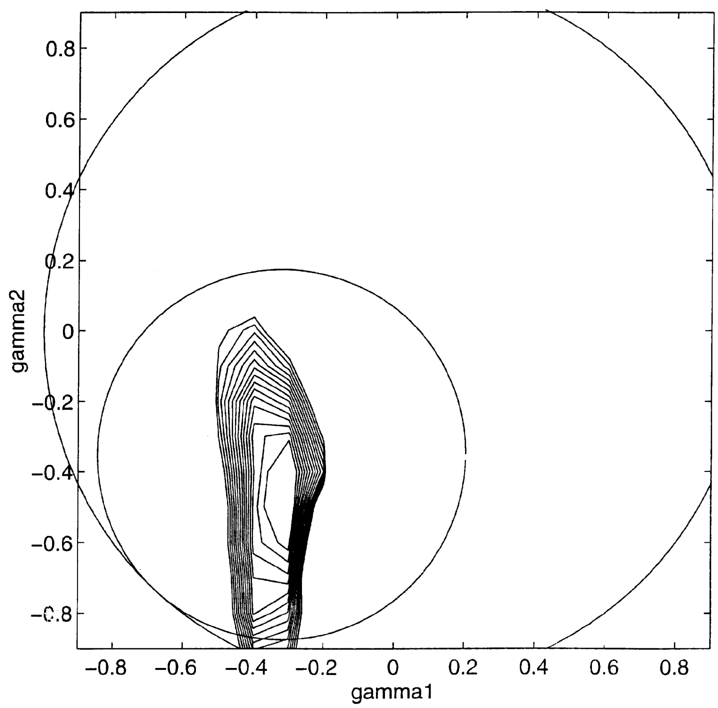

FIG. 1. *Contours of* M_{nd}.

TABLE 2
Regression estimates for the Stanford heart transplant data, with and without the five extreme survival times

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\Delta\,\hat{\beta}_1$	$\Delta\,\hat{\beta}_2$
BJ152	0.1070	−0.0017		
BJ157	0.0794	−0.0012	−25.8%	+29.4%
$\hat{\beta}$152	0.1434	−0.0022		
$\hat{\beta}$157	0.1325	−0.0020	−7.6%	8.6%
$\tilde{\beta}$152	0.1426	−0.0022		
$\tilde{\beta}$157	0.1391	−0.0021	−6.7%	3.5%
Cox152	0.1462	−0.0024		
Cox157	0.1399	−0.0022	−4.3%	8.3%

run are summarized in Table 2 for the original parameter β. Changes with the inclusion, denoted by Δ, are also given.

From Table 2 the minimum distance estimators and the partial likelihood estimator seem to be more robust than the Buckley–James estimator. In some preliminary simulation results, the minimum distance estimators also seemed very competitive and more robust than the Buckley–James estimator.

Acknowledgments. I thank Professors Hira Koul, Benjamin Duran and the referees for valuable comments and James Dunyak and Carroll Nunn for helpful discussions on numerical computations. All their efforts led to the improved presentation of this paper.

REFERENCES

ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067. [Correction: *Ann. Probab.* **15** 428–430.]

BERAN, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, R. J. (1982). Robust estimation in models for independent non-identically distributed data. *Ann. Statist.* **10** 415–428.

BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

CHEN, Y. Y., HOLLANDER, M. and LANGBERG, N. A. (1982). Small-sample results for the Kaplan–Meier estimator. *J. Amer. Statist. Assoc.* **77** 141–144.

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

DONOHO, D. L. and LIU, R. C. (1988a). The "automatic" robustness of minimum distance functionals. *Ann Statist.* **16** 552–586.

DONOHO, D. L. and LIU, R. C. (1988b). Pathologies of some minimum distance estimators. *Ann Statist.* **16** 587–608.

GILL, R. (1980). *Censoring and Stochastic Integrals.* Math. Centrum, Amsterdam.

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

KAPLAN, E. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KOUL, H. (1985). Minimum distance estimation in linear regression with unknown error distributions. *Statist. Probab. Lett.* **3** 1–8.

KOUL, H. (1992). *Weighted Empiricals and Linear Models*. IMS, Hayward, CA.
KOUL, H. and DEWIT, T. (1983). Minimum distance estimation in a linear regression model. *Ann. Statist.* **11** 921–932.
KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9** 1276–1288.
LAI, T. L. and YING, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27** 334–358.
LAI, T. L. and YING, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19** 531–554.
LEURGANS, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74** 301–309.
LOÉVE, M. (1963). *Probability Theory.* Van Nostrand, Princeton.
MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. *Lecture Notes in Math.* **976** 75–265. Springer, New York.
MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.
MILLER, R. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.
PARR, W. C. and SCHUCANY, W. R. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75** 616–624.
PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. and VETTERLING, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge Univ. Press.
RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.
RUBINSTEIN, R. Y. (1986). *Monte Carlo Optimization, Simulation and Sensitivity of Queuing Networks*. Wiley, New York.
SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.
TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372.
YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21** 76–99.
ZHOU, M. (1992). M-estimation in censored linear models. *Biometrika* **79** 837–841.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
P.O. BOX 41042
LUBBOCK, TEXAS 79409
E-MAIL: yesoy@ttacs1.ttu.edu