Statistica Sinica 2(1992), 17-46

ASYMPTOTICALLY EFFICIENT ESTIMATION IN CENSORED AND TRUNCATED REGRESSION MODELS

Tze Leung Lai and Zhiliang Ying

Stanford University and University of Illinois

Abstract: Information bounds are developed for estimation of regression parameters in the presence of left truncation and right censoring on the observed responses, assuming that the vectors of covariates and censoring/truncation variables are independent (but possibly non-identically distributed). Under certain regularity conditions, asymptotically efficient estimators that attain these information bounds are also given.

Key words and phrases: Censoring and truncation, regression, Fisher information, regular estimators, asymptotic minimax bounds, semiparametric models, adaptive rank estimators, martingales and stochastic integrals, asymptotic normality.

1. Introduction

Consider the linear regression model

$$y_i^* = \beta^T x_i^* + \epsilon_i^* \quad (i = 1, 2, \ldots), \tag{1.1}$$

where the ϵ_i^* are i.i.d. random variables representing unobservable disturbances and having a common continuously differentiable distribution function F with density f, β is a $d \times 1$ vector of unknown parameters and the x_i^* are either nonrandom or are independent random $d \times 1$ vectors independent of $\{\epsilon_i^*\}$. Suppose that the responses y_i^* are not completely observable due to left truncation and right censoring by random variables t_i^* and c_i^* such that $-\infty \le t_i^* < \infty$ and $-\infty < c_i^* \le \infty$. Let $\tilde{y}_i^* = y_i^* \wedge c_i^*$ and $\delta_i^* = I_{\{y_i^* \le c_i^*\}}$, where we use $\wedge$ and $\vee$ to denote minimum and maximum, respectively. In addition to right censorship of the responses y_i^* by c_i^*, suppose that there is also left truncation in the sense that $(\tilde{y}_i^*, \delta_i^*, x_i^*)$ can be observed only when $\tilde{y}_i^* \ge t_i^*$. The data, therefore, consist of n observations $(\tilde{y}_i, t_i, \delta_i, x_i)$ with $\tilde{y}_i \ge t_i, i = 1, \ldots, n$.

The vectors (t_i^*, c_i^*, x_i^{*T}) are usually assumed to be independent random vectors that are independent of the sequence $\{\epsilon_n^*\}$. The special case $t_i^* \equiv -\infty$ corresponds to the "censored regression model" which is of basic importance in statistical modelling and analysis of failure time data (cf. Kalbfleisch and

Prentice (1980), Lawless (1982)). The special case $c_i^* \equiv \infty$ corresponds to the "truncated regression model" in econometrics (cf. Tobin (1958), Goldberger (1981), Amemiya (1985)) and in astronomy (cf. Segal (1975), Nicoll and Segal (1980)), which assumes the presence of truncation variables T_i^* so that (x_i^*, y_i^*) can be observed only when $y_i^* \le T_i^*$ (or equivalently, when $-y_i^* \ge -T_i^* = t_i^*$).

Instead of assuming the (t_i^*, c_i^*, x_i^{*T}) to be independent so that the sample $\{(t_i, c_i, x_i, y_i) : 1 \le i \le n\}$ can be regarded as having been generated by a larger, randomly stopped sample of independent random vectors $(t_i^*, c_i^*, x_i^{*T}, y_i^*)$, $1 \le i \le m(n) = \inf\{m : \sum_{i=1}^m I_{\{t_i^* \le y_i^* \wedge c_i^*\}} = n\}$, an alternative setting proposed by Turnbull (1976) is to assume that (t_i, c_i, x_i^T) are independent random vectors that are independent of $\{\epsilon_n^*\}$ and such that $c_i \ge t_i$ and

$$\begin{aligned} &\tau_0 = 0,\ \tau_j = \inf\{i > \tau_{j-1} : y_i^* \ge t_j\},\ \tilde{y}_j = y_{\tau_j}^* \wedge c_j, \\ &(t_i^*, c_i^*, x_i^*) = (t_j, c_j, x_j) \quad \text{for } \tau_{j-1} < i \le \tau_j. \end{aligned} \tag{1.2}$$

In this formulation, (t_i, c_i, x_i^T, y_i) are independent random vectors such that the conditional distribution of y_i given (t_i, c_i, x_i) is

$$\begin{aligned} &P\{y_i \le y | t_i, c_i, x_i\} \\ &\quad = \{F(y - \beta^T x_i) - F(t_i - \beta^T x_i)\}/\{1 - F(t_i - \beta^T x_i)\},\ y \ge t_i. \end{aligned} \tag{1.3}$$

Suppose that (t_i, c_i, x_i^T) are i.i.d. random vectors whose distributions do not depend on β and that the conditional distribution of y_i given (t_i, c_i, x_i) is determined by (1.3). Under the assumption that the density function f of F is known, the maximum likelihood estimator of β will be shown in Section 2 to be asymptotically normal with mean 0 and covariance matrix $n^{-1}V_f$, where V_f^{-1} is the Fisher information matrix. Without assuming f to be known, it will be shown in Section 3 that adaptive estimators can nevertheless be constructed so that they are asymptotically normal with mean 0 and covariance matrix $n^{-1}V_f$ when x_i has mean 0 and is independent of $(t_i - \beta^T x_i, c_i - \beta^T x_i)$. In general, these estimators may have larger asymptotic covariance matrices (in the sense of nonnegative definite differences) than $n^{-1}V_f$, but can still be shown to attain the asymptotically minimal covariance matrix for the asymptotic distributions of regular estimators. Such optimality results follow from the generalization of the Hájek convolution theorem and asymptotic minimax bounds to semiparametric models by Begun, Hall, Huang and Wellner (1983), since the (t_i, c_i, x_i^T) are assumed to be i.i.d. random vectors. In Section 2 we further remove the restrictive assumption that the (t_i, c_i, x_i^T) be identically distributed, which excludes the important case of nonrandom t_i, c_i and x_i, and we also develop asymptotic lower bounds for minimax risks in the general setting where (t_i, c_i, x_i^T) are only assumed to be independent. Moreover, we consider the setting of independent (t_i^*, c_i^*, x_i^*)

instead of independent (t_i, c_i, x_i) and develop asymptotic lower bounds in this alternative setting. Section 3 shows how to construct estimators that are asymptotically efficient in either setting, in the sense that the covariance matrix of the asymptotic normal distribution coincides with that given by the asymptotic lower bound.

In classical regression theory, it is usually assumed that F has finite mean α, and (1.1) is usually rewritten as

$$y_i^* = \alpha + \beta^T x_i^* + \epsilon_i \quad (i = 1, 2, \dots), \tag{1.4}$$

where the ϵ_i have mean 0. When the y_i^* are not completely observable because of censorship and truncation, it is often not possible to give consistent estimates of α although β can still be estimated consistently and efficiently. Therefore it is more natural to combine the unidentifiable α with ϵ_i, leading to the model (1.1). For completely observable y_i^*, which corresponds to the special case $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$, our results in Sections 2 and 3 still yield asymptotically efficient estimators of the slope β. In Section 4 we develop asymptotic lower bounds for the minimax risks in estimating both α and β for the case of independent (but possibly non-identically distributed) covariates x_i^* but without censorship or truncation (i.e., $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$), and show how the ideas in Section 3 can be extended to give asymptotically efficient estimators of the intercept α and the slope β.

2. Information Bounds for Estimating β

Suppose that (t_i^*, c_i^*, x_i^{*T}) are i.i.d. random vectors. Then by (1.1), $(t_i^*, c_i^*, x_i^{*T}, y_i^*)$ are i.i.d., from which it follows that the observed sample $\{(t_i, x_i^T, \delta_i, \tilde{y}_i) : i = 1, \dots, n\}$ consists of n i.i.d. random vectors. Likewise, if the (t_i, c_i, x_i^T) are i.i.d. and the conditional distribution of y_i given (t_i, c_i, x_i^T) is determined by (1.3), then $(t_i, x_i^T, \delta_i, \tilde{y}_i), i = 1, \dots, n$, are i.i.d. random vectors. Parameter estimation in either setting, therefore, fits into the general setting of estimation in semiparametric models based on i.i.d. observations as discussed by Stein (1956), Koshevnik and Levit (1976), and Begun, Hall, Huang and Wellner (1983). In particular, Begun et al. (1983) develop information bounds in the form of convolution-type representations of regular estimators and asymptotic minimax bounds on risk functions, by making use of (i) the notion of a "Hellinger differentiable density" for partial differentiation with respect to the nonparametric part (infinite-dimensional nuisance parameter) of the model, and (ii) the characterization of the "effective score" for the finite-dimensional parameter of interest as that component of the score function orthogonal (in some Hilbert space) to all nuisance parameter scores. For the censored regression model (i.e.,

$t_i^* \equiv -\infty$), Ritov and Wellner (1988) have evaluated explicitly the orthogonal projections to calculate these information bounds under the assumption that the (c_i^*, x_i^{*T}) are i.i.d. with a common distribution not depending on β.

When the sample observations $(t_i, x_i^T, \delta_i, \tilde{y}_i)$, $i = 1, \ldots, n$, are not i.i.d., the framework of Begun et al. (1983) is no longer applicable. Instead of using a functional-analytic approach involving orthogonal projections to calculate information bounds, we use a more direct approach that involves replacing the nonparametric component of the model by suitably chosen parametric subfamilies that are asymptotically least favorable in some sense. The difficult step in our approach, therefore, is to "guess" such parametric subfamilies. We have actually first guessed what the information bounds should be on the basis of the asymptotic distributions of certain estimators (presented in Section 3) which we believe to be asymptotically efficient. With those information bounds in mind, it is not hard to choose the appropriate parametric subfamilies.

We now outline more specifically the arguments that will be used to develop information bounds for the estimation of β under independent but possibly non-identically distributed (t_i, c_i, x_i^T). Similar arguments will be used for the case of independent (t_i^*, c_i^*, x_i^{*T}). First, from the joint density function of $(t_i, x_i^T, \delta_i, \tilde{y}_i)$, $i = 1, \ldots, n$, we obtain the likelihood function of β and f (common density of the ϵ_i^*), assuming that the density function of (t_i, c_i, x_i^T) does not depend on β and f for every i. When f is known, the maximum likelihood estimator $\widehat{\beta}_n$ of β can then be determined and the Fisher information matrix can be computed to give the variance of the asymptotic normal distribution of $\widehat{\beta}_n$ under certain regularity conditions. Under these regularity conditions, Hájek's convolution theorem for regular estimators and asymptotic minimax bounds are applicable and we establish in Subsection 2.1 the asymptotic efficiency of the maximum likelihood estimator (cf. Hájek (1970, 1972), Ibragimov and Has'minskii (1981)). In Subsection 2.2 we drop the assumption of known f and assume only that f belongs to a family $\mathcal{F}_n$ of densities satisfying certain regularity conditions. We then define certain parametric subfamilies of $\mathcal{F}_n$ and consider asymptotic minimax bounds and regular estimators for each such parametric subfamily. Using martingale theory and ideas similar to those of Chapter 4 of Ibragimov and Has'minskii (1981), we develop asymptotic lower bounds on minimax risks, which will be shown in Section 3 to be attained by certain rank estimators. Subsection 2.3 considers the case of independent (t_i^*, c_i^*, x_i^{*T}) instead of independent (t_i, c_i, x_i^T). Here we have an additional complication involving the conditional densities of t_i^* and c_i^* given x_i^*, which can be handled by constructing suitable parametric subfamilies of the class of conditional densities.

2.1. Maximum likelihood estimators in the case of known f and inde-

pendent (t_i, c_i, x_i)

Let (t_i, c_i, x_i^T), $i = 1, \ldots, n$, be a sequence of independent random vectors. Suppose that for each i, conditional on (t_i, c_i, x_i), y_i has the density $f(y - \beta^T x_i)/(1 - F(t_i - \beta^T x_i))$, $y \geq t_i$. Let $q_i(\cdot|t_i, x_i)$ denote the conditional density of c_i given (t_i, x_i) and let p_i denote the joint density of (t_i, x_i), where both p_i and q_i are with respect to some σ-finite measures and do not depend on the regression parameter β. The likelihood function given the data $\tilde{y}_i = y_i \wedge c_i$, $\delta_i = I_{\{y_i \leq c_i\}}, t_i, x_i (i = 1, \ldots, n)$ is therefore proportional to

$$L_n(\beta) = \prod_{i=1}^{n} \left\{ \left[\frac{f(\tilde{y}_i - \beta^T x_i)}{1 - F(t_i - \beta^T x_i)} \right]^{\delta_i} \left[\frac{1 - F(\tilde{y}_i - \beta^T x_i)}{1 - F(t_i - \beta^T x_i)} \right]^{1-\delta_i} \right\}. \tag{2.1}$$

Assuming f to be continuously differentiable, the maximum likelihood estimator can be determined as a zero of the function

$$\begin{aligned} \frac{\partial}{\partial \beta} \log L_n(\beta) &= \sum_{i=1}^{n} x_i \left\{ \delta_i \frac{-f'(\tilde{y}_i - \beta^T x_i)}{f(\tilde{y}_i - \beta^T x_i)} + (1 - \delta_i) \frac{f(\tilde{y}_i - \beta^T x_i)}{1 - F(\tilde{y}_i - \beta^T x_i)} \right. \\ &\qquad \left. - \frac{f(t_i - \beta^T x_i)}{1 - F(t_i - \beta^T x_i)} \right\} \\ &= \sum_{i=1}^{n} x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)} I_{\{f(t)>0\}} [dN_i(\beta, t) - \lambda(t) J_i(\beta, t) dt], \end{aligned} \tag{2.2}$$

where $\partial/\partial\beta$ denotes the gradient vector, $\lambda = f/(1 - F)$ is the hazard function and

$$N_i(\beta, t) = I_{\{\tilde{y}_i - \beta^T x_i \leq t,\ \delta_i = 1\}}, \quad J_i(\beta, t) = I_{\{\tilde{y}_i - \beta^T x_i \geq t \geq t_i - \beta^T x_i\}}. \tag{2.3}$$

The second equality in (2.2) follows from the formulas $\lambda'/\lambda = f'/f + f/(1 - F)$ and $\int_{-\infty}^{\infty} \lambda'(t) I_{\{f(t)>0\}} J_i(\beta, t) dt = -\int_{-\infty}^{\infty} \lambda(t) I_{\{f(t)>0\}} dJ_i(\beta, t)$.

For left truncated and right censored data, it simplifies the analysis considerably by using hazard functions instead of density functions and by using stochastic integral representations as in (2.2). The following lemma enables us to apply martingale theory to analyze these stochastic integrals.

Lemma 1. *Suppose that $(t_i, c_i, x_i^T), i = 1, \ldots, n$, are independent random vectors with $c_i \geq t_i$ and that the conditional distribution of y_i given (t_i, c_i, x_i) is determined by (1.3) for every i. For $-\infty < s < \infty$, let $\mathcal{B}_s$ be the complete σ-field generated by t_i, c_i, x_i, $(y_i - \beta^T x_i) I_{\{y_i - \beta^T x_i \leq s\}}, i = 1, \ldots, n$. Let $\tilde{y}_i = y_i \wedge c_i$, $\delta_i = I_{\{y_i \leq c_i\}}$, and define $N_i(\beta, t), J_i(\beta, t)$ by (2.3). Let*

$$M_i(t) = N_i(\beta, t) - \int_{-\infty}^{t} J_i(\beta, s) \lambda(s) ds. \tag{2.4}$$

Then $\{M_i(t), \mathcal{B}_t, -\infty < t < \infty\}$ *is a martingale with predictable variation process*

$$\langle M_i\rangle(t) = \int_{-\infty}^{t} J_i(\beta, s)\lambda(s)ds. \tag{2.5}$$

Proof. Let $\mathcal{B}_{-\infty} = \cap_{s=-\infty}^{\infty}\mathcal{B}_s$. Note that (t_i, c_i, x_i) is $\mathcal{B}_{-\infty}$-measurable. Moreover, by (1.3), $y_i \geq t_i$ and the conditional hazard function of y_i given (t_i, c_i, x_i) is $\lambda(\cdot - \beta^T x_i)$. Hence the same argument as in the proof of Lemma 5 of Lai and Ying (1991a) can be used to show that

$$M_i(s) = I_{\{y_i - \beta^T x_i \leq s\wedge(c_i - \beta^T x_i)\}} - \int_{y_i - \beta^T x_i}^{s\wedge(c_i - \beta^T x_i)} I_{\{y_i - \beta^T x_i \geq u\}}\lambda(u)du$$

is a martingale with respect to $\mathcal{B}_s$, and that its predictable variation process is given by (2.5).

Throughout the sequel we shall use for any vector x the notation $x^0 = 1$ (scalar), $x^1 = x$ (vector), and $x^2 = xx^T$ (matrix). We shall denote the actual value of the unknown parameter β by β_0. Suppose that for some $C > 0$ and $0 < \delta < 1$,

$$\|x_i\| \leq C \text{ and } F(t_i - \beta_0^T x_i) \leq 1 - \delta \text{ a.s. for every } i, \tag{2.6}$$

F has a continuously differentiable density f such that

$$\int_{-\infty}^{\infty} (f'/f)^2 dF < \infty, \tag{2.7}$$

and that for $k = 0, 1, 2$ and for every s with $F(s) < 1$,

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\{x_i^k I_{\{t_i - \beta_0^T x_i \leq s \leq c_i - \beta_0^T x_i\}}/(1 - F(t_i - \beta_0^T x_i))\} = \Gamma_k(s) \text{ exists.} \tag{2.8}$$

The assumption (2.7) implies that $\int_{-\infty}^{\infty}(\lambda'/\lambda)^2 dF < \infty$, cf. Lemma 2 of Lai and Ying (1991c). By Lemma 1, $\{n^{-1/2}\sum_{i=1}^{n} x_i \int_{-\infty}^{t}\{-\lambda'(s)/\lambda(s)\}I_{\{f(s)>0\}}dM_i(s), \mathcal{B}_t, -\infty < t < \infty\}$ is a martingale with predictable variation process

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n} x_i^2 \int_{-\infty}^{t}\{\lambda'(s)/\lambda(s)\}^2 I_{\{\tilde{y}_i - \beta_0^T x_i \geq s \geq t_i - \beta_0^T x_i\}}\lambda(s) I_{\{f(s)>0\}}ds \\ &= \int_{-\infty}^{t}\{\lambda'(s)/\lambda(s)\}^2 n^{-1}\sum_{i=1}^{n} x_i^2 I_{\{c_i - \beta_0^T x_i \geq s \geq t_i - \beta_0^T x_i\}} I_{\{y_i - \beta_0^T x_i \geq s\}} \\ &\cdot (1 - F(s))^{-1}dF(s) \xrightarrow{P} \int_{-\infty}^{t}(\lambda'(s)/\lambda(s))^2\Gamma_2(s)dF(s), \end{aligned} \tag{2.9}$$

by the law of large numbers and by (2.8) and (1.3), since (t_i, c_i, x_i^T, y_i) are independent random vectors. Since $\int_{-\infty}^{\infty}(\lambda'/\lambda)^2 dF < \infty$, it follows that $\lim_{A\to\infty}\int_{|\lambda'/\lambda|>A}(\lambda'/\lambda)^2 dF = 0$, and therefore by Lenglart's (1977) inequality,

$$\sup_t \|n^{-1/2}\sum_{i=1}^n x_i \int_{\infty}^t (-\lambda'(s)/\lambda(s)) I_{\{|\lambda'(s)/\lambda(s)|>n^{1/3}\}} dM_i(s)\| \xrightarrow{P} 0. \qquad (2.10)$$

Note that the jumps of the process $\{\sum_{i=1}^n x_i \int_{-\infty}^t (\lambda'/\lambda) I_{\{|\lambda'/\lambda|\le n^{1/3}\}} dM_i, t \ge 0\}$ are $\le Cn^{1/3}$ a.s. Hence we can apply Rebolledo's (1980) martingale central limit theorem to conclude from (2.2) and (2.9) that

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\beta}\log L_n(\beta)|_{\beta=\beta_0} \xrightarrow{\mathcal{L}} N(0, I(\beta_0)), \text{ where } I(\beta_0) = \int_{-\infty}^{\infty}(\lambda'/\lambda)^2\Gamma_2 dF. \quad (2.11)$$

Here and in the sequel we use the symbol $\xrightarrow{\mathcal{L}}$ to denote convergence in distribution. Note that the quantity $I(\beta_0)$ in (2.11) represents the limiting Fisher information matrix. Under a quadratic-mean differentiability condition on λ, the following lemma gives a refinement of (2.11) so that Hájek's convolution theorem and asymptotic minimax bounds are applicable to the present estimation problem.

Lemma 2. *Under the assumptions of Lemma* 1 *and* (2.6)−(2.8), *suppose that* $I(\beta_0)$ *is nonsingular and that*

$$\int_{-\infty}^{\infty}\left\{\frac{\lambda(t+\epsilon)-\lambda(t)-\epsilon\lambda'(t)}{\epsilon}\right\}^2 \frac{1}{\lambda^2(t)} I_{\{\lambda(t)>0\}} dF(t) \to 0 \quad as\ \epsilon \to 0. \qquad (2.12)$$

Then for any $d \times 1$ *vector* u,

$$\log L_n(\beta_0 + u/\sqrt{n}) - \log L_n(\beta_0) \xrightarrow{\mathcal{L}} N(-u^T I(\beta_0)u/2, u^T I(\beta_0)u). \qquad (2.13)$$

Consequently, for any bounded continuous function $w : \mathbf{R}^d \to [0,\infty)$ *such that*

$$w(0) = 0, w(x) = w(-x), \{x : w(x) < a\} \text{ is convex for all } a > 0, \qquad (2.14)$$

we have for any $0 < \epsilon < 1/2$ *and any estimator* T_n *of* β,

$$\liminf_{n\to\infty} \sup_{\|\beta-\beta_0\|\le n^{-\epsilon}} E_\beta w(\sqrt{n}(T_n - \beta))$$
$$\ge (2\pi)^{-d/2}\int_{\mathbf{R}^d} w((I(\beta_0))^{-1/2}x)e^{-\|x\|^2/2}dx, \qquad (2.15)$$

and the lower bound in (2.15) *is attained by the solution* $\widehat{\beta}_n$ *of the estimating equation* $(\partial/\partial\beta)\log L_n(\beta) = 0$ *in a ball centered at* β_0 *with radius* $n^{-\epsilon}$. *Moreover,* *if T_n is a regular estimator in the sense that for any $d \times 1$ vector x and for $\beta = \beta_0 + n^{-1/2}x$, $\sqrt{n}(T_n - \beta)$ has a limiting distribution Ψ (not depending on x), then*

$$\Psi = N(0, (I(\beta_0))^{-1}) * H \tag{2.16}$$

where $*$ denotes convolution and H is some other distribution. In particular, $\widehat{\beta}_n$ is regular with $\Psi = N(0, (I(\beta_0))^{-1})$, i.e., with H concentrating all its mass at 0.

Proof. To prove (2.13), first note that $f(y_i - \beta^T x_i) > 0$ and $F(\tilde{y}_i - \beta^T x_i) < 1$ a.s. Let

$$\zeta_{i,n} = [\lambda(\tilde{y}_i - \beta_0^T x_i - u^T x_i/\sqrt{n})/\lambda(\tilde{y}_i - \beta_0^T x_i)]^{\delta_i} - 1.$$

Then by (1.3), (2.6) and (2.12), with probability 1,

$$\begin{aligned}
&\max_{i\le n} E[\{\zeta_{i,n} - \delta_i n^{-1/2} u^T x_i \lambda'(\tilde{y}_i - \beta_0^T x_i)/\lambda(\tilde{y}_i - \beta_0^T x_i)\}^2 | t_i, c_i, x_i] \\
&= \max_{i\le n} \int_{-\infty}^{c_i - \beta_0^T x_i} \left[\frac{\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s) - n^{-1/2} u^T x_i \lambda'(s)}{\lambda(s)}\right]^2 \\
&\quad \cdot \frac{f(s)}{1 - F(t_i - \beta^T x_i)} ds = o(n^{-1}).
\end{aligned}$$

Hence for every $\epsilon > 0$, with probability 1,

$$\begin{aligned}
&\max_{i\le n} P\{|\zeta_{i,n}| \ge \epsilon | t_i, c_i, x_i\} \\
&\le \max_{i\le n} P\{|n^{-1/2} u^T x_i| |\lambda'(\tilde{y}_i - \beta_0^T x_i)/\lambda(\tilde{y}_i - \beta_0^T x_i)| \ge \epsilon/2 | t_i, c_i, x_i\} + (2/\epsilon)^2 o(n^{-1}) \\
&\le (2/\epsilon)^2 \int_{C\|u\| |\lambda'(s)/\lambda(s)| \ge \epsilon\sqrt{n}/2} \left(\frac{u^T x_i}{\sqrt{n}}\right)^2 \left(\frac{\lambda'(s)}{\lambda(s)}\right) \frac{f(s)}{1-\delta} ds + o(n^{-1}) = o(n^{-1}),
\end{aligned}$$

again making use of (2.6). Therefore as $n \to \infty$,

$$P\Big\{\max_{i\le n} |\zeta_{i,n}| \ge \epsilon\Big\} \le nE\Big[\max_{i\le n} P\{|\zeta_{i,n}| \ge \epsilon | t_i, c_i, x_i\}\Big] \to 0,$$

by the dominated convergence theorem. Hence $\max_{i\le n} |\zeta_{i,n}| \xrightarrow{P} 0$ and for every $0 < \epsilon < 1/2$,

$$P(A_{n,\epsilon}) \to 1, \text{ where } A_{n,\epsilon} = \Big\{\max_{i\le n} |\zeta_{i,n}| < \epsilon\Big\}. \tag{2.17}$$

Since $\lambda(y_i - \beta^T x_i) > 0$ a.s., we also have $\lambda(\tilde{y}_i - \beta_0^T x_i - u^T x_i/\sqrt{n}) > 0$ a.s.

on $\{\delta_i = 1\} \cap \{|\zeta_{i,n}| < 1/2\}$. From (2.2), it follows that on $A_{n,\epsilon}$,

$$\begin{aligned}
&\log L_n(\beta_0 + u/\sqrt{n}) - \log L_n(\beta_0) \\
&= \sum_{i=1}^{n}\{\delta_i[\log\lambda(\tilde{y}_i - \beta_0^T x_i - u^T x_i/\sqrt{n}) - \log\lambda(\tilde{y}_i - \beta_0^T x_i)] \\
&\quad - [\Lambda(\tilde{y}_i - \beta_0^T x_i - u^T x_i/\sqrt{n}) - \Lambda(\tilde{y}_i - \beta_0^T x_i)] \\
&\quad + [\Lambda(t_i - \beta_0^T x_i - u^T x_i/\sqrt{n}) - \Lambda(t_i - \beta_0^T x_i)]\} \\
&= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{\log\lambda(s - u^T x_i/\sqrt{n}) - \log\lambda(s)\}dN_i(\beta_0, s) \\
&\quad + \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{\Lambda(s - u^T x_i/\sqrt{n}) - \Lambda(s)\}dJ_i(\beta_0, s),
\end{aligned}$$

where $\Lambda(t) = \int_{-\infty}^{t}\lambda(s)ds = -\log(1 - F(t))$. Therefore on $A_{n,\epsilon}$,

$$\begin{aligned}
&\log L_n(\beta_0 + u/\sqrt{n}) - \log L_n(\beta_0) \\
&= \sum_{i=1}^{n}\int_{-\infty}^{\infty} I_{\{\lambda(s)>0, |\lambda(s-u^T x_i)/\lambda(s)-1|<\epsilon\}}\Big\{[\log\lambda(s - u^T x_i/\sqrt{n}) - \log\lambda(s)]dM_i(s) \\
&+ \Big[\log\Big(1 + \frac{\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s)}{\lambda(s)}\Big)\Big]\lambda(s)J_i(\beta_0, s)ds \\
&- J_i(\beta_0, s)[\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s)]ds\Big\}. \qquad (2.18)
\end{aligned}$$

Using the law of large numbers together with (2.6)–(2.8) and (2.12), it can be shown by an argument similar to that in (2.9) that

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Big(\frac{\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s)}{\lambda(s)}\Big)^2 \lambda(s)J_i(\beta_0, s)ds \\
&\xrightarrow{P} u^T\Big[\int_{-\infty}^{\infty}\Gamma_2(s)\frac{\lambda'(s)}{\lambda(s)}ds\Big]u. \qquad (2.19)
\end{aligned}$$

Applying (2.17), (2.19) and the Taylor expansion $\log(1+r) = r - \frac{1}{2}r^2(1+r^*)^{-2}$, with r^* lying between 0 and r, we obtain by letting $n \to \infty$ and then $\epsilon \to 0$ that

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{-\infty}^{\infty}\Big\{\Big[\log(1 + \frac{\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s)}{\lambda(s)})\Big]\lambda(s) - \Big[\lambda(s - u^T x_i/\sqrt{n}) - \lambda(s)\Big]\Big\} \\
&\times J_i(\beta_0, s)I_{\{\lambda(s)>0, |\lambda(s-u^T x_i/\sqrt{n})/\lambda(s)-1|<\epsilon\}}ds \xrightarrow{P} -\frac{1}{2}u^T I(\beta_0)u. \qquad (2.20)
\end{aligned}$$

Again using the Taylor expansion for $\log(1 + r)$, an argument similar to (2.9) and (2.10) can be used to show that Rebolledo's (1980) martingale central limit theorem is applicable to yield

$$\sum_{i=1}^{n}\int_{-\infty}^{\infty} I_{\{\lambda(s)>0,|\lambda(s-u^T x_i)/\lambda(s)-1|<\epsilon\}} \log\Big(1+\frac{\lambda(s-u^T x_i/\sqrt{n})-\lambda(s)}{\lambda(s)}\Big)dM_i(s)$$
$$\xrightarrow{\mathcal{L}} N(0,u^T I(\beta_0)u),\ \text{letting } n\to\infty \text{ and then } \epsilon\to 0. \tag{2.21}$$

From (2.17), (2.18), (2.20) and (2.21), the desired conclusion (2.13) follows.

From (2.13), it follows that for any $x \in \mathbf{R}^d$ the sequence of probability measures $\{P_{\beta_0+(nI(\beta_0))^{-1/2}x}\}$ is locally asymptotically normal (LAN) with normalizing matrices $(nI(\beta_0))^{-1/2}$, cf. page 120 of IH (which we shall use to denote Ibragimov and Has'minskii (1981)). Hence the asymptotic minimax bound (2.15) follows from Theorem 12.1 on page 162 of IH, while (2.16) follows from the Hájek convolution theorem, cf. Theorem 9.1 on page 154 of IH. Finally, the desired asymptotic normality and asymptotic efficiency properties of $\widehat{\beta}_n$ follow from well known results on maximum likelihood estimators in LAN families (cf. Sections 1 and 4 of Chapter 3 of IH).

2.2. Asymptotic lower bounds on minimax risks when f is unknown in the case of independent (t_i, c_i, x_i)

To begin with, suppose that f is embedded in a parametric family $\{f_\theta\}$ such that the hazard function of f_θ is $\lambda_\theta(t) = \lambda(t) + \theta^T\eta(t)$, where θ is a $d\times 1$ vector and $\eta : \mathbf{R}\to\mathbf{R}^d$ is some function to be specified later. Replacing f in the likelihood function (2.1) by f_θ yields

$$\log L_n(\theta,\beta) = \sum_{i=1}^{n}\{\delta_i \log\lambda_\theta(\tilde{y}_i-\beta^T x_i) - \Lambda_\theta(\tilde{y}_i-\beta^T x_i) + \Lambda_\theta(t_i-\beta^T x_i)\}, \tag{2.22}$$

where $\Lambda_\theta(t) = \int_{-\infty}^{t}\lambda_\theta(s)ds = -\log(1-F_\theta(t))$. Hence if $\eta(s)=0$ whenever $\lambda(s)=0$, then

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L_n(\theta,\beta)|_{\theta=0} &= \sum_{i=1}^{n}\Big\{\frac{\delta_i\eta(\tilde{y}_i-\beta^T x_i)}{\lambda(\tilde{y}_i-\beta^T x_i)} - \int_{t_i-\beta^T x_i}^{\tilde{y}_i-\beta^T x_i}\eta(s)ds\Big\}\\ &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{\eta(s)}{\lambda(s)}[dN_i(\beta,s) - J_i(\beta,s)\lambda(s)ds],\end{aligned}$$

setting $0/0=0$ and noting that $\int_{-\infty}^{\infty}(\int_{-\infty}^{t}\eta(s)ds)dJ_i(\beta,t) = -\int_{-\infty}^{\infty}J_i(\beta,t)\eta(t)dt$; moreover,

$$\frac{\partial}{\partial\beta}\log L_n(\theta,\beta)|_{\theta=0} = \sum_{i=1}^{n}x_i\int_{-\infty}^{\infty}\frac{-\lambda'(t)}{\lambda(t)}[dN_i(\beta,t)-J_i(\beta,t)\lambda(t)dt],$$

as in (2.2). Let $\ell_n = \log L_n$. Therefore, under (2.6)–(2.8) and suitable regularity conditions on η, an argument similar to that in the proof of (2.11) yields

$$\frac{1}{\sqrt{n}}\begin{pmatrix}(\partial/\partial\theta)\ell_n(0,\beta_0)\\(\partial/\partial\beta)\ell_n(0,\beta_0)\end{pmatrix} \xrightarrow{\mathcal{L}} N(0,\mathbf{I}_\eta), \text{ where } \mathbf{I}_\eta = \begin{pmatrix} I_{\theta\theta} & I_{\theta\beta}\\ I_{\theta\beta}^T & I_{\beta\beta}\end{pmatrix}, \tag{2.23}$$

$$\begin{aligned}
I_{\theta\theta} &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{n}\sum_{i=1}^{n} E\frac{I_{\{t_i-\beta_0^T x_i\le t\le c_i-\beta_0^T x_i\}}}{1-F(t_i-\beta_0^T x_i)}\Big[\frac{\eta(t)}{\lambda(t)}\Big]^2 dF(t)\\
&= \int_{-\infty}^{\infty}\Gamma_0(t)\frac{\eta(t)\eta^T(t)}{\lambda^2(t)}dF(t),\\
I_{\theta\beta}^T &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{n}\sum_{i=1}^{n} E\Big[x_i\frac{I_{\{t_i-\beta_0^T x_i\le t\le c_i-\beta_0^T x_i\}}}{1-F(t_i-\beta_0^T x_i)}\Big]\frac{-\lambda'(t)\eta^T(t)}{\lambda^2(t)}dF(t)\\
&= \int_{-\infty}^{\infty}\frac{-\lambda'(t)}{\lambda^2(t)}\Gamma_1(t)\eta^T(t)dF(t),\\
I_{\beta\beta} &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{n}\sum_{i=1}^{n} E\Big[x_i^2\frac{I_{\{t_i-\beta_0^T x_i\le t\le c_i-\beta_0^T x_i\}}}{1-F(t_i-\beta_0^T x_i)}\Big]\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2 dF(t)\\
&= \int_{-\infty}^{\infty}\Gamma_2(t)\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2 dF(t).
\end{aligned}$$

Assume $\mathbf{I}_\eta$ to be nonsingular. The lower right $d\times d$ submatrix of the $2d\times 2d$ matrix $\mathbf{I}_\eta^{-1}$ is

$$\begin{aligned}
(I_{\beta\beta}-I_{\theta\beta}^T I_{\theta\theta}^{-1}I_{\theta\beta})^{-1} = \Big\{\int_{-\infty}^{\infty}\Gamma_2(t)\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2 dF(t) - \Big(\int_{-\infty}^{\infty}\frac{-\lambda'(t)}{\lambda^2(t)}\Gamma_1(t)\eta^T(t)dF(t)\Big)\\
\times\Big(\int_{-\infty}^{\infty}\frac{\Gamma_0(t)}{\lambda^2(t)}\eta(t)\eta^T(t)dF(t)\Big)^{-1}\Big(\int_{-\infty}^{\infty}\frac{-\lambda'(t)}{\lambda^2(t)}\Gamma_1(t)\eta^T(t)dF(t)\Big)\Big\}^{-1}.
\end{aligned} \tag{2.24}$$

For the univariate case $d = 1$, an application of the Schwarz inequality shows that the right hand side of the above equation is

$$\le\Big\{\int_{-\infty}^{\infty}\Gamma_2(t)\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2 dF(t) - \int_{-\infty}^{\infty}\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2\frac{\Gamma_1^2(t)}{\Gamma_0(t)}dF(t)\Big\}^{-1},$$

with equality when $\eta = -\lambda'\Gamma_1/\Gamma_0$. For general d, if we still choose $\eta = -\lambda'\Gamma_1/\Gamma_0$, we again have $(I_{\beta\beta} - I_{\theta\beta}^T I_{\theta\theta}^{-1} I_{\theta\beta})^{-1} = (I_f(\beta_0))^{-1}$, where

$$I_f(\beta_0) = \int_{-\infty}^{\infty}\Big[\Gamma_2(t) - \frac{\Gamma_1(t)\Gamma_1^T(t)}{\Gamma_0(t)}\Big]\Big[\frac{\lambda'(t)}{\lambda(t)}\Big]^2 dF(t). \tag{2.25}$$

Under the assumptions (2.6)–(2.8) and assuming $I_f(\beta_0)$ to be positive definite, we can choose, for every $A \ge A_0$ (sufficiently large), a continuous function

$\eta_A : \mathbf{R} \to \mathbf{R}^d$ such that

$$\eta_A(s) = 0 \text{ if } f(s) \leq 1/A, \ \sup_s \|\eta_A(s)\| \leq A, \tag{2.26a}$$

$$\mathbf{I}_{\eta_A} \text{ is positive definite, where } \mathbf{I}_{\eta_A} \text{ is defined in (2.23)}, \tag{2.26b}$$

$$\lim_{A\to\infty} \int_{-\infty}^{\infty} \left\| \frac{\eta_A(t)}{\lambda(t)} - \left[\frac{-\lambda'(t)}{\lambda(t)}\right] \frac{\Gamma_1(t)}{\Gamma_0(t)} \right\|^2 I_{\{\lambda(t)>0, \Gamma_0(t)>0\}} dF(t) = 0. \tag{2.26c}$$

Take any ϵ satisfying $0 < \epsilon < 1/2$. For $\|\theta\| \leq n^{-\epsilon}$ and all sufficiently large n, $\lambda_{\theta,A}(t) = \lambda(t) + \theta^T \eta_A(t)$ is a hazard function. Let $\phi_{\theta,A}(t) = \lambda_{\theta,A}(t) \exp\{-\int_{-\infty}^t \lambda_{\theta,A}(s)ds\}$ denote the corresponding density function. Let $\mathcal{F}_n$ be a family of density functions such that for every fixed $A \geq A_0$,

$$\phi_{\theta,A} \in \mathcal{F}_n \text{ for } \|\theta\| \leq n^{-\epsilon} \text{ and all large } n. \tag{2.27}$$

In Section 3, without assuming f to be known, we shall construct estimators $\widehat{\beta}_n$ of β such that as $n \to \infty$, the limiting distribution of $\sqrt{n}(\widehat{\beta}_n - \beta)$ is normal with mean 0 and covariance matrix $(I_f(\beta_0))^{-1}$ under $P_{\beta,\phi}$, and the convergence to normality is uniform in $\|\beta - \beta_0\| \leq n^{-\epsilon}$ and $\phi \in \mathcal{F}_n$, for some family $\mathcal{F}_n$ of densities such that (2.27) holds for every given $A \geq A_0$. The following theorem shows that the limiting covariance matrix $(I_f(\beta_0))^{-1}$ is asymptotically minimal and gives an extension of Lemma 2 in which $I(\beta_0)$ is to be replaced by $I_f(\beta_0)$ for the present setting of unknown f. Note that the "information loss" in not knowing f manifests itself in the replacement of the integrand $\Gamma_2(t)$ in $I(\beta_0)$ by $\Gamma_2(t) - \Gamma_1(t)\Gamma_1^T(t)/\Gamma_0(t)$ in $I_f(\beta_0)$.

Theorem 1. *Let* $0 < \epsilon < 1/2$. *Under* (2.6)–(2.8), (2.12) *and the assumptions of Lemma* 1, *define* $I_f(\beta_0)$ *by* (2.25) *and assume that it is positive definite. Let* $\mathcal{F}_n$ *be a family of densities satisfying* (2.27) *for every given* $A \geq A_0$.

(i) *Let* $w : \mathbf{R}^d \to [0,\infty)$ *be a bounded continuous function satisfying* (2.14). *For any estimator* T_n *of* β,

$$\begin{aligned}&\liminf_{n\to\infty}\{\sup_{\|\beta-\beta_0\|\leq n^{-\epsilon}, \phi\in\mathcal{F}_n} E_{\beta,\phi} w(\sqrt{n}(T_n - \beta))\}\\ \geq &(2\pi)^{-d/2} \int_{\mathbf{R}^d} w((I_f(\beta_0))^{-1/2}x) e^{-\|x\|^2/2} dx.\end{aligned}$$

(ii) *Suppose that an estimator* T_n *is regular in the following sense: For any sequence* $(\beta_n, \phi_n)_{n\geq 1}$ *with* $\sqrt{n}(\beta_n - \beta)$ *converging to some vector and* $\phi_n \in \mathcal{F}_n$, $\sqrt{n}(T_n - \beta_n)$ *has a limiting distribution* Ψ *under* P_{β_n,ϕ_n} *where* Ψ *does not depend on the particular sequence* $(\beta_n, \phi_n)_{n\geq 1}$. *Then* $\Psi = N(0, (I_f(\beta_0))^{-1}) * H$ *for some distribution* H.

Proof. Take any $A \geq A_0$ and consider the parametric subfamily $\{\phi_{\theta,A} : \|\theta\| \leq n^{-\epsilon}\}$ of $\mathcal{F}_n$. For this parametric subfamily, defining $L_n(\theta, \beta)$ by (2.22) with λ_θ replaced by $\lambda_{\theta,A}$, it can be shown, by making use of (2.6)–(2.8), (2.12), (2.26) and an argument similar to the proof of (2.13), that for any $d \times 1$ vectors u and v,

$$\log L_n\Big(\frac{u}{\sqrt{n}}, \beta_0 + \frac{v}{\sqrt{n}}\Big) - \log L_n(0, \beta_0)$$
$$\xrightarrow{\mathcal{L}} N\Big(-\frac{1}{2}(u^T, v^T)\mathbf{I}_{\eta_A}\binom{u}{v}, (u^T, v^T)\mathbf{I}_{\eta_A}\binom{u}{v}\Big), \tag{2.28}$$

where $\mathbf{I}_{\eta_A}$ is defined in (2.23). In view of this local asymptotic normality (LAN) property, for every bounded nonnegative function w^* on $\mathbf{R}^d \times \mathbf{R}^d$ such that $w^*(0) = 0$, $w^*(x) = w^*(-x)$ and $\{x : w^*(x) < a\}$ is convex for all $a > 0$, we have for any estimators T_n of β and $\widehat{\theta}_n$ of θ the asymptotic minimax bound

$$\liminf_{n\to\infty}\{\sup_{\|\beta-\beta_0\|\leq n^{-\epsilon}, \|\theta\|\leq n^{-\epsilon}} E_{\beta,\phi_{\theta,A}} w^*(\sqrt{n}(\widehat{\theta}_n - \theta), \sqrt{n}(T_n - \beta))\} \geq Ew^*(N(0, \mathbf{I}_{\eta_A}^{-1}))$$

(cf. page 162 of IH). In particular, for the special case $w^*(y, z) = w(z)$, this reduces to

$$\liminf_{n\to\infty}\{\sup_{\|\beta-\beta_0\|\leq n^{-\epsilon}, \|\theta\|\leq n^{-\epsilon}} E_{\beta,\phi_{\theta,A}} w(\sqrt{n}(T_n - \beta))\}$$
$$\geq Ew(Z_A) \quad \text{with } Z_A \sim N\Big(0, (I_{\beta\beta} - I_{\theta\beta}^T I_{\theta\theta}^{-1} I_{\theta\beta})_A^{-1}\Big), \tag{2.29}$$

where $(I_{\beta\beta} - I_{\theta\beta}^T I_{\theta\theta}^{-1} I_{\theta\beta})_A^{-1}$ is the lower right $d \times d$ submatrix of $\mathbf{I}_{\eta_A}^{-1}$ and is given by (2.24) with η replaced by η_A. From (2.27) and (2.29), it follows that for every $A \geq A_0$,

$$\liminf_{n\to\infty}\{\sup_{\|\beta-\beta_0\|\leq n^{-\epsilon}, \phi\in\mathcal{F}_n} E_{\beta,\phi} w(\sqrt{n}(T_n - \beta))\} \geq Ew(Z_A). \tag{2.30}$$

Since (2.30) holds for every $A \geq A_0$ and since $(I_{\beta\beta} - I_{\theta\beta}^T I_{\theta\theta}^{-1} I_{\theta\beta})_A^{-1} \to (I_f(\beta_0))^{-1}$ as $A \to \infty$ by (2.26c), (i) follows.

To prove (ii), take any $u, v \in \mathbf{R}^d$ and $A \geq A_0$. Let $\beta_{n,u} = \beta_0 + u/\sqrt{n}$, $\psi_{n,v} = \phi_{v/\sqrt{n},A}$. By (2.27), $\psi_{n,v} \in \mathcal{F}_n$ for all large n. Since $\sqrt{n}(T_n - \beta_{n,u})$ has a limiting distribution Ψ under $P_{\beta_{n,u},\psi_{n,v}}$ for any $u, v \in \mathbf{R}^d$, it then follows from the LAN property and Hajek's convolution theorem (cf. IH, page 154) that Ψ is the distribution of the sum $Z_A + R_A$ of independent random variables Z_A in (2.29) and R_A. Since this holds for every $A \geq A_0$ and since Z_A has the limiting $N(0, (I_f(\beta_0))^{-1})$ distribution as $A \to \infty$, it then follows that R_A has a limiting distribution H as $A \to \infty$, and therefore $\Psi = N(0, (I_f(\beta_0))^{-1}) * H$.

2.3. Partial likelihood and information bounds in the case of independent (t_i^*, c_i^*, x_i^*)

Instead of assuming the observed $(t_i, x_i, \delta_i, \tilde{y}_i)$ to be independent as in (1.3), where the unobservable (t_i^*, c_i^*, x_i^*) at times $i \in \{\tau_{j-1}+1, \ldots, \tau_j - 1\}$ are all set equal to (t_j, c_j, x_j) and τ_j is the first time i after τ_{j-1} to yield an observation (i.e., $\tilde{y}_i \geq t_j$), we now consider the case in which (t_i^*, c_i^*, x_i^{*T}) are independent random vectors that are independent of the sequence $\{\epsilon_n^*\}$. We shall still assume (2.7) and (2.12) on the common density f of the ϵ_i^* but replace the assumptions (2.6) and (2.8) by

$$\|x_i^*\| \leq C \text{ a.s. for all } i \text{ and some nonrandom constant } C, \tag{2.31}$$

$$\lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} E\{x_i^{*k} I_{\{t_i^*-\beta_0^T x_i^* \leq s \leq c_i^* - \beta_0^T x_i^*\}}\} = \Gamma_k^*(s) \text{ exists for } k = 0,1,2$$
$$\text{and } s < F^{-1}(1), \tag{2.32}$$

$$\lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} P\{t_i^* - \beta_0^T x_i^* \leq c_i^* - \beta_0^T x_i^* < s\} = \Gamma^*(s)$$
$$\text{exists for every } s < F^{-1}(1), \tag{2.33}$$

$$\sigma_0 \triangleq \inf\{s : \Gamma_0^*(s) > 0\} < \sigma_1 \triangleq \inf\{s > \sigma_0 : (1 - F(s))\Gamma_0^*(s) = 0\}. \tag{2.34}$$

To begin with, suppose that the common density f of the ϵ_i^* in (1.1) is known. Let $g_i(\cdot|x_i^*)$ be the conditional density of t_i^* given x_i^* with respect to Lebesgue measure, and let $C_i(x_i^*, t_i^*) = P\{c_i^* \geq t_i^* | x_i^*, t_i^*\}$. Let ν_i be the distribution function of x_i^*. Let $\tau_0 = 0$, $\tau_j = \inf\{i > \tau_{j-1} : \tilde{y}_i^* \geq t_i^*\}$. Then $x_i = x_{\tau_i}^*$, $y_i = y_{\tau_i}^*$, etc., and

$$\begin{aligned} &P\{\tau_j > \tau_{j-1} + n | (\tau_r, t_r, x_r, \delta_r, \tilde{y}_r), r \leq j-1\} \\ =&P\{\tilde{y}_i^* < t_i^* \text{ for } \tau_{j-1}+1 \leq i \leq \tau_{j-1}+n | (\tau_r, t_r, x_r, \delta_r, \tilde{y}_r), r \leq j-1\} \\ =&\prod_{\tau_{j-1}+1}^{\tau_{j-1}+n} \iint (1 - F(t - \beta^T x)) C_i(t,x) g_i(t|x)\, dt\, d\nu_i(x). \end{aligned}$$

Suppose that in addition to $(t_i, x_i, \delta_i, \tilde{y}_i)_{i\leq n}$, one also observes $\tau_1, \ldots, \tau_n$. Assuming f, C_i and ν_i to be known and g_i to belong to a smooth parametric family $\{g_{i,\gamma}\}$ with $\gamma \in \mathbf{R}^d$ and $g_i = g_{i,0}$, the likelihood function of (γ, β) based on $(\tau_i, t_i, x_i, \delta_i, \tilde{y}_i)_{i\leq n}$ is proportional to

$$\begin{aligned} &L_n(\gamma, \beta) \\ =&\prod_{j=1}^{n}\left\{\frac{f^{\delta_j}(\tilde{y}_j - \beta^T x_j)[1 - F(\tilde{y}_j - \beta^T x_j)]^{1-\delta_j}}{1 - F(t_j - \beta^T x_j)}[(1 - F(t_j - \beta^T x_j))g_{\tau_j,\gamma}(t_j|x_j)]\right. \end{aligned}$$

$$\times \prod_{i=\tau_{j-1}+1}^{\tau_j-1} \iint (1-F(t-\beta^T x)) g_{i,\gamma}(t|x) C_i(t,x)\,dt\,d\nu_i(x) \Big\}. \tag{2.35}$$

Let $G_i(\cdot|x_i^*)$ be the conditional distribution function of t_i^* given x_i^*. Letting $h_{i,\gamma}$ denote the hazard function of $g_{i,\gamma}$, suppose that $h_{i,\gamma}$ is of the form

$$h_{i,\gamma}(t|x) = h_i(t|x)\left\{1 + \gamma^T \psi(t,x) - \frac{\gamma^T \int_{-\infty}^t \psi(s,x)\,dG_i(s|x)}{1 - G_i(t|x)}\right\}, \tag{2.36}$$

where $h_i = h_{i,0}$ and $\psi(s,x)$ is a $d \times 1$ vector to be specified later. Since $g_{i,\gamma}(t|x) = h_{i,\gamma}(t|x)\exp\{-\int_{-\infty}^t h_{i,\gamma}(s|x)ds\}$, it follows from (2.36) that

$$\frac{\partial}{\partial\gamma} g_{i,\gamma}(t|x)|_{\gamma=0} = \psi(t,x) g_i(t|x), \tag{2.37}$$

noting that $[h_i(t|x)/(1-G_i(t|x))]dt = (1-G_i(t|x))^{-2}dG_i(t|x) = d[1/(1-G_i(t|x))]$ and using integration by parts to evaluate $\int_{-\infty}^u \{\int_{-\infty}^t \psi(s,x)g_i(s|x)ds\} d[1/(1 - G_i(t|x))]$. From (2.35) and (2.37), it follows that

$$\begin{aligned}
&\frac{\partial}{\partial\gamma}\log L_n(\gamma,\beta)|_{\gamma=0} \\
&= \sum_{j=1}^n \left\{\psi(t_j,x_j) + \sum_{i=\tau_{j-1}+1}^{\tau_j-1} \frac{\iint (1-F(t-\beta^T x)) C_i(t,x)\psi(t,x) g_i(t|x)\,dt\,d\nu_i(x)}{\iint (1-F(t-\beta^T x)) C_i(t,x) g_i(t|x)\,dt\,d\nu_i(x)}\right\}, \\
&\frac{\partial}{\partial\beta}\log L_n(\gamma,\beta)|_{\gamma=0} \\
&= \sum_{i=1}^n x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)}[dN_i(\beta,t) - J_i(\beta,t)\lambda(t)dt] + \sum_{i=1}^n x_i \frac{f(t_i - \beta^T x_i)}{1 - F(t_i - \beta^T x_i)} \\
&+ \sum_{j=1}^n \sum_{i=\tau_{j-1}+1}^{\tau_j-1} \frac{\iint x f(t-\beta^T x) C_i(t,x) g_i(t|x)\,dt\,d\nu_i(x)}{\iint (1-F(t-\beta^T x)) C_i(t,x) g_i(t|x)\,dt\,d\nu_i(x)},
\end{aligned}$$

where $N_i(\beta,t)$ and $J_i(\beta,t)$ are defined in (2.3). Hence, for the particular choice

$$\psi(t,x) = x f(t-\beta^T x)/(1 - F(t-\beta^T x)), \tag{2.38}$$

and letting $\ell_n = \log L_n$, we have

$$\frac{\partial}{\partial\beta}\ell_n(0,\beta) = \sum_{i=1}^n x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)}[dN_i(\beta,t) - J_i(\beta,t)\lambda(t)dt] + \frac{\partial}{\partial\theta}\ell_n(0,\beta). \tag{2.39}$$

Let $t_i^*(\beta) = t_i^* - \beta^T x_i^*$, $\tilde{y}_i^*(\beta) = \tilde{y}_i^* - \beta^T x_i^*$. Let $\mathcal{B}_s^*$ be the complete δ-field generated by

$$t_i^*, x_i^*, I_{\{t_i^* \leq \tilde{y}_i^*\}}, \delta_i^* I_{\{t_i^*(\beta_0) \leq \tilde{y}_i^*(\beta_0) \leq s\}}, I_{\{t_i^*(\beta_0) \leq u \leq \tilde{y}_i^*(\beta_0)\}}, I_{\{t_i^*(\beta_0) \leq \tilde{y}_i^*(\beta_0) \leq u\}}$$
$$(u \leq s, i = 1, 2, \ldots). \tag{2.40}$$

Let $\mathcal{B}_{-\infty}^* = \cap_{s=-\infty}^{\infty} \mathcal{B}_s^*$ and $S_n = (\partial/\partial\gamma)\ell_n(0, \beta_0)$. Note that $S_n, \tau_1, \tau_2, \ldots$ are $\mathcal{B}_{-\infty}^*$-measurable. Define $M_i(t)$ as in (2.4). Then $\{M_i(s), \mathcal{B}_s^*, -\infty < s < \infty\}$ is a martingale (cf. Lemma 5 of Lai and Ying (1991a)). Letting

$$W_n = \sum_{i=1}^n x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)} [dN_i(\beta_0, t) - J_i(\beta_0, t)\lambda(t)dt] = \sum_{i=1}^n x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)} dM_i(t),$$

it then follows, under suitable integrability assumptions, that $E\{(W_n + S_n)S_n^T\} = ES_n^2$, $E(W_n + S_n)^2 = EW_n^2 + ES_n^2$ (where $W_n^2 = W_n W_n^T$, etc.). Hence

$$\operatorname{Cov}\begin{pmatrix}(\partial/\partial\gamma)\ell_n(0,\beta_0)\\(\partial/\partial\beta)\ell_n(0,\beta_0)\end{pmatrix} = \begin{pmatrix} I_{n,\gamma\gamma} & I_{n,\gamma\beta} \\ I_{n,\gamma\beta}^T & I_{n,\beta\beta}\end{pmatrix} = \begin{pmatrix} ES_n^2 & ES_n^2 \\ ES_n^2 & EW_n^2 + ES_n^2\end{pmatrix},$$

and therefore

$$I_{n,\beta\beta} - I_{n,\gamma\beta}^T I_{n,\gamma\gamma}^{-1} I_{n,\gamma\beta} = EW_n^2, \tag{2.41}$$

where the inverse refers to the generalized Moore-Penrose inverse without assuming ES_n^2 to be nonsingular. Thus, the Fisher information for estimating β in the parametric subfamily $\{g_{i,\gamma}\}$ defined by (2.36) and (2.38) for the conditional densities of t_i^* given x_i^* is EW_n^2, and there is no loss in information if we replace L_n by the following partial likelihood (cf. Cox (1975), Wong (1986)):

$$L_n^*(\beta) = \prod_{j=1}^n \{f^{\delta_j}(\tilde{y}_j - \beta^T x_j)[1 - F(\tilde{y}_j - \beta^T x_j)]^{1-\delta_j} / [1 - F(t_j - \beta^T x_j)]\}, \tag{2.42}$$

which ignores the other factors that involve the term $\big(1 - F(t - \beta^T x)\big) g_{i,\gamma}(t|x)$ in the full likelihood (2.35). Indeed

$$\frac{\partial}{\partial\beta} \log L_n^*(\beta)|_{\beta=\beta_0} = \sum_{i=1}^n x_i \int_{-\infty}^{\infty} \frac{-\lambda'(t)}{\lambda(t)} [dN_i(\beta_0, t) - J_i(\beta_0, t)\lambda(t)dt] = W_n.$$

Note that the partial likelihood $L_n^*(\beta)$ can be regarded as the likelihood function of β for the family of conditional probability measures P_β^* given (τ_i, t_i, x_i), $i = 1, \ldots, n$. Since the factor $\big(1 - F(t - \beta^T x)\big) g_i(t|x)$ that involves β in the joint density of (τ_i, t_i, x_i) contains the unknown g_i, it is intuitively clear that restricting to the conditional probability measures P_β^* should not result in loss of information, especially in the case where the g_i can be arbitrary densities that change with i so that the sample contains negligible information about them. We have shown that even if the g_i can be embedded in a d-dimensional parametric family $\{g_{i,\gamma}\}$ whose hazard functions are parametrized by (2.36), there is a "least favorable" choice of ψ in (2.36) such that the Fisher information of the parametric family for the estimation of β agrees with that based on the conditional probability measures P^*_β.

We now consider the case of unknown f. Suppose that f can be embedded in a parametric family $\{f_\theta\}$ of the form in Subsection 2.2. A simple modification of the preceding argument can be used to show that using the conditional probability measures $P^*_{\theta,\beta}$ given $(\tau_i, t_i, x_i), i = 1, \dots, n$, again results in essentially no loss of information for estimating (θ, β). Hence we can estimate (θ, β) by maximizing the partial likelihood

$$L^*_n(\theta,\beta) = \prod_{j=1}^n \Big\{ f_\theta^{\delta_j}(\tilde{y}_j - \beta^T x_j)[1-F_\theta(\tilde{y}_j - \beta^T x_j)]^{1-\delta_j} / [1-F_\theta(t_j - \beta^T x_j)] \Big\}. \tag{2.43}$$

For the parametric family $\lambda_\theta(t) = \lambda(t) + \theta^T \eta_A(t)$, where λ_θ is the hazard function of f_θ and η_A satisfies (2.26a)–(2.26c), the following theorem shows that the parametric family of conditional probability measures $\{P^*_{u/\sqrt{n},\ \beta_0 + v/\sqrt{n}}\}$ has the LAN property for any $u, v \in \mathbf{R}^d$, i.e.,

$$\begin{aligned} &\log L^*_n(u/\sqrt{n}, \beta_0 + v/\sqrt{n}) - \log L^*_n(0, \beta_0) \\ &\xrightarrow{\mathcal{L}} N\Big(-\frac{1}{2}(u^T, v^T)\mathbf{I}_{\eta_A}\binom{u}{v}, (u^T, v^T)\mathbf{I}_{\eta_A}\binom{u}{v}\Big). \end{aligned} \tag{2.44}$$

From this LAN property and the preceding discussion that shows no loss of information in using the partial likelihood instead of the full likelihood, we can use that same arguments as those in the proof of Theorem 1 to show that the conclusions of parts (i) and (ii) of Theorem 1 still hold in the present setting.

Theorem 2. *Suppose that (t_i^*, c_i^*, x_i^{*T}), $i = 1, 2, \dots$, are independent random vectors which are independent of $\{\epsilon_n^*\}$ and whose distributions are not specified and do not depend on β. Suppose that (2.31)–(2.34) hold and that the commom density f and hazard function λ of the ϵ_n^* satisfy conditions (2.7) and (2.12). For $k = 0, 1, 2$, define*

$$\Gamma_k(s) = \Gamma_k^*(s) / \int_{-\infty}^{\infty} \{\Gamma_0^*(t) + \Gamma^*(t)\} dF(t). \tag{2.45}$$

With Γ_k defined by (2.45), define $\mathbf{I}_\eta$ by (2.23) for $\eta : \mathbf{R} \to \mathbf{R}^d$ and define $I_f(\beta_0)$ by (2.25). Assume that $I_f(\beta_0)$ is positive definite, and choose $\eta_A : \mathbf{R} \to \mathbf{R}^d$ so that

(2.26a)–(2.26c) *are satisfied for every* $A \geq A_0$. *Let* $\lambda_{\theta,A}(t) = \lambda(t) + \theta^T \eta_A(t)$, $\phi_{\theta,A}(t) = \lambda_{\theta,A}(t) \exp\{-\int_{-\infty}^t \lambda_{\theta,A}(s)ds\}$. *Let* $0 < \epsilon < 1/2$ *and let* $\mathcal{F}_n$ *be a family of densities satisfying* (2.27). *Then for the parametric subfamily* $\{\phi_{\theta,A}\}$, *the* LAN *property* (2.44) *holds for all* $d \times 1$ *vectors* u, v *and every* $A \geq A_0$. *Consequently, the conclusions of parts* (i) *and* (ii) *of Theorem* 1 *still hold under the present setting and notation.*

Proof. Let $\tau_0 = 0$, $\tau_j = \inf\{i > \tau_{j-1} : \tilde{y}_i^* \geq t_i^*\}, n^* = \tau_n$. Then with probability 1, as $n \to \infty$,

$$n^*/n \to \Big(\lim_{m\to\infty} m^{-1} \sum_{i=1}^m P\{t_i^* \leq \tilde{y}_i^*\}\Big)^{-1} = \Big\{\int_{-\infty}^{\infty} [\Gamma_0^*(t) + \Gamma^*(t)] dF(t)\Big\}^{-1} \tag{2.46}$$

(cf. Lai and Ying (1991b)). Letting $N_i^*(\beta, t) = I_{\{\tilde{y}_i^* - \beta^T x_i^* \leq t,\ \delta_i^* = 1\}}$, $J_i^*(\beta, t) = I_{\{t_i^* - \beta^T x_i^* \leq t \leq \tilde{y}_i^* - \beta^T x_i^*\}}$, we can rewrite (2.43) as

$$\begin{aligned}\log L_n^*(\theta, \beta) &= \sum_{i=1}^{n^*} \{\delta_i^* \log \lambda_\theta(\tilde{y}_i^* - \beta^T x_i^*) - \Lambda_\theta(\tilde{y}_i^* - \beta^T x_i^*) + \Lambda_\theta(t_i^* - \beta^T x_i^*)\} I_{\{t_i^* \leq \tilde{y}_i^*\}} \\ &= \sum_{i=1}^{n^*} \{\int_{-\infty}^{\infty} \log \lambda_\theta(s) dN_i^*(\beta, s) + \int_{-\infty}^{\infty} \Lambda_\theta(s) dJ_i^*(\beta, s)\}. \end{aligned} \tag{2.47}$$

Fix $A \geq A_0$. We shall make use of (2.47) to show that for the parametric family $\lambda_\theta(t) = \lambda_{\theta,A}(t) = \lambda(t) + \theta^T \eta_A(t)$, the LAN property (2.44) holds for any $u, v \in \mathbf{R}^d$.

Let $\zeta_{i,n} = \{[\lambda(\tilde{y}_i^* - \beta_0^T x_i^* - v^T x_i^*/\sqrt{n})/\lambda(\tilde{y}_i^* - \beta_0^T x_i^*)]^{\delta_i^*} - 1\} I_{\{t_i^* \leq \tilde{y}_i^*\}}$. Using (2.46) and an argument similar to the proof of (2.17), it can be shown that $\max_{i \leq n^*} |\zeta_{i,n}| \xrightarrow{P} 0$. Note that by (2.26a), $\eta_A(s) = 0$ and therefore $\lambda_\theta(s) = \lambda(s)$ if $f(s) \leq 1/A$. Let $A_{n,\epsilon} = \{\max_{i \leq n^*} |\zeta_{i,n}| < \epsilon\}$ for $0 < \epsilon < 1/2$. On $A_{n,\epsilon}$, we have from (2.47) that analogous to (2.18),

$$\begin{aligned} &\log L_n^*(u/\sqrt{n}, \beta_0 + v/\sqrt{n}) - \log L_n^*(0, \beta_0) \\ &= \sum_{i=1}^{n^*} \int_{-\infty}^{\infty} I_{\{\lambda(s) > 0, |\lambda(s - v^T x_i/\sqrt{n})/\lambda(s) - 1| < \epsilon\}} \\ &\cdot \Big\{ \log\Big[\frac{\lambda(s - v^T x_i/\sqrt{n}) + n^{-1/2} u^T \eta_A(s - v^T x_i/\sqrt{n})}{\lambda(s)}\Big] dM_i^*(s) \\ &+ \log\Big[1 + \frac{\lambda(s - v^T x_i/\sqrt{n}) - \lambda(s) + n^{-1/2} u^T \eta_A(s - v^T x_i/\sqrt{n})}{\lambda(s)}\Big] \lambda(s) J_i^*(\beta_0, s) ds \\ &- J_i^*(\beta_0, s)[\lambda(s - v^T x_i/\sqrt{n}) - \lambda(s) + n^{-1/2} u^T \eta_A(s - v^T x_i/\sqrt{n})] ds \Big\}, \end{aligned}$$

where $M_i^*(t) = N_i^*(\beta_0, t) - \int_{-\infty}^t J_i^*(\beta_0, s)\lambda(s)ds$. Let $\mathcal{B}_s^*$ be the complete σ-field generated by (2.40). Then $\{M_i^*(s), \mathcal{B}_s^*, -\infty < s < \infty\}$ is a martingale with predictable variation process $\langle M_i^*\rangle(t) = \int_{-\infty}^t J_i^*(\beta_0, s)\lambda(s)ds$, cf. Lemma 5 of Lai and Ying (1991a). Moreover, n^* is $\mathcal{B}_{-\infty}^*$-measurable, and analogous to (2.9),

$$\sum_{i=1}^{n^*} \int_{-\infty}^{t} I_{\{\lambda(s)>0, |\lambda(s-v^T x_i/\sqrt{n})/\lambda(s)-1|<\epsilon\}}$$
$$\times \log^2 \left[1 + \frac{\lambda(s - v^T x_i/\sqrt{n}) - \lambda(s) + n^{-1/2} u^T \eta_A(s - v^T x_i/\sqrt{n})}{\lambda(s)}\right]$$
$$\times \lambda(s) I_{\{\tilde{y}_i^* - \beta_0^T x_i^* \geq s \geq t_i^* - \beta_0^T x_i^*\}} ds \xrightarrow{P} (u^T, v^T) \mathbf{I}_{\eta_A} \begin{pmatrix} u \\ v \end{pmatrix},$$

by (2.12), (2.26a), (2.31), (2.32) and (2.46), noting that $I_{\{\tilde{y}_i^* - \beta_0^T x_i^* \geq s \geq t_i^* - \beta_0^T x_i^*\}}$ are independent random variables to which the law of large numbers is applicable. The rest of the proof of (2.44) is essentially the same as that of (2.13).

3. Asymptotically Efficient Estimators of β

In this section we first review the development in Lai and Ying (1991b) of rank estimators $\widehat{\beta}_n$ of β based on left truncated and right censored data such that $\sqrt{n}(\widehat{\beta}_n - \beta)$ has a limiting normal distribution with mean 0 and covariance matrix $(I_f(\beta))^{-1}$, which is asymptotically minimal in the sense of Theorem 1 (or 2). We then provide some refinements and extensions of these results. Earlier, for the censored regression model (i.e., with $t_i^* \equiv -\infty$) and under the assumption of i.i.d. covariates x_i and i.i.d. censoring variables $c_i^*(= c_i)$, Ritov (1984) introduced another method, which is much more complicated and involves quite stringent assumptions, to construct asymptotically normal estimators that attain the information bound $(I_f(\beta_0))^{-1}$ suggested by the general theory of Begun et al. (1983) and others in semiparametric estimation based on i.i.d. observations.

A starting point of the development in Lai and Ying (1991b) is the following general class of rank statistics formed from the residuals $e_i(b) = \tilde{y}_i - b^T x_i$. Let $e_{(1)}(b) \leq \cdots \leq e_{(k)}(b)$ denote all the ordered uncensored residuals. For $i = 1, \ldots, k$, let

$$J(i,b) = \{j \leq n : t_j - b^T x_j \leq e_{(i)}(b) \leq \tilde{y}_j - b^T x_j\},\ n_i(b) = \#J(i,b),$$
$$\bar{x}(i,b) = \Big(\sum_{j \in J(i,b)} x_j\Big)/n_i(b), \tag{3.1}$$

where we use the notation $\#A$ to denote the number of elements of a set A. Let ψ be a twice continuously differentiable function on (0,1) such that $\sup_x |\psi''(x)| < \infty$. Let p be a nondecreasing and twice continuously differentiable function on the real line such that

$$p(y) = 0 \text{ for } y \le 0 \text{ and } p(y) = 1 \text{ for } y \ge 1. \tag{3.2}$$

Take $0 < \lambda < 1/18$ and define $p_n(z) = p(n^\lambda(z - cn^{-\lambda}))$ for $0 \le z \le 1$. Define the product-limit estimator $\widehat{F}_{n,b}$ and the rank statistic associated with ψ by

$$1 - \widehat{F}_{n,b}(u) = \prod_{i:e_{(i)}(b)<u,\delta_{(i)}=1} \{1 - p_n(n^{-1}n_i(b))/n_i(b)\}, \tag{3.3}$$

$$S_n(b) = \sum_{i=1}^{k} \psi(\widehat{F}_{n,b}(e_{(i)}(b)))p_n(n^{-1}n_i(b))\{x_{(i)} - \bar{x}(i,b)\}. \tag{3.4}$$

A rank estimator $\tilde{\beta}_n$ of β is defined as a minimizer of $\|S_n(b)\|$ for $\|b\| \le \rho$, assuming knowledge of an upper bound $\rho > \|\beta\|$.

For the setting in which (t_i^*, c_i^*, x_i^{*T}) are independent random vectors, Lai and Ying (1991b) showed that the rank estimators $\tilde{\beta}_n$ are consistent and asymptotically normal under (2.31)–(2.34) and some additional assumptions. In particular, the limiting normal distribution of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ has covariance matrix $(I_f(\beta_0))^{-1}$ if the score function ψ is so chosen that $\psi \circ \tilde{F}$ is a scalar multiple of λ'/λ, where β_0 denotes the true value of the unknown parameter β, $\tilde{F}(t) = P\{\epsilon_1^* \le t | \epsilon_1^* \ge \sigma_0\}$ and σ_0 is defined in (2.34). Since the hazard function λ of the ϵ_i^* is typically unknown, Lai and Ying (1991b) also proposed the following modification of (3.4) that involves an adaptive choice of the score function.

An important idea in the modification is to divide the sample into two disjoint subsets, the first of which is $\{(t_i, x_i, \delta_i, \tilde{y}_i) : i \le n/2\}$. From the first subsample define the residuals $e_i(b) = \tilde{y}_i - b^T x_i (i \le n/2)$ and order the uncensored ones among them as $e_{(1)}(b) \le \cdots \le e_{(k_1)}(b)$. Let $n_1 = [n/2]$, i.e., the largest integer $\le n/2$, and define $J(i,b), n_i(b), \bar{x}(i,b)$ as in (3.1) but with n_1 replacing n (i.e., on the basis only of the first subsample). Let $\psi_{n,2}(s)$ be an estimate of $\lambda'(s)/\lambda(s)$ based on the second subsample of $n_2 = n - n_1$ observations and define in analogy with (3.4)

$$S_{n,1}(b) = \sum_{i=1}^{k_1} \psi_{n,2}(e_{(i)}(b))p_n(n^{-1}n_i(b))\{x_{(i)} - \bar{x}(i,b)\}. \tag{3.5}$$

Likewise from the second subsample define the residuals $e_i^*(b) = \tilde{y}_{n_1+i} - b^T x_{n_1+i}$ $(i \le n_2)$ and order the uncensored ones among them as $e_{[1]}^*(b) \le \cdots \le e_{[k_2]}^*(b)$.

As in (3.1), let

$$J^*(i,b) = \{n_1 < r \le n : t_r - b^T x_r \le e^*_{[i]}(b) \le \tilde{y}_r - b^T x_r\},\ n^*_i(b) = \#J^*(i,b),$$
$$\bar{x}^*(i,b) = \Big(\sum_{r\in J^*(i,b)} x_r\Big)/n^*_i(b).$$

Let $\psi_{n,1}(s)$ be an estimate of $\lambda'(s)/\lambda(s)$ based on the first subsample, and define

$$S_{n,2}(b) = \sum_{i=1}^{k_2} \psi_{n,1}(e^*_{[i]}(b)) p_n(n^{-1} n^*_i(b))\{x_{n_1+[i]} - \bar{x}^*(i,b)\}. \tag{3.6}$$

Combining the two subsample statistics (3.5) and (3.6) gives the adaptive rank statistics

$$S^*_n(b) = S_{n,1}(b) + S_{n,2}(b). \tag{3.7}$$

From the jth subsample, starting with a preliminary consistent estimate $b_{n,j}$ of β, Lai and Ying (1991b) showed (i) how to construct from the uncensored residuals $\tilde{y}_i - b^T_{n,j} x_i$ in the jth subsample a smooth consistent estimate $\widehat{\lambda}_{n,j}$ of the hazard function λ, and (ii) how to smooth $\widehat{\lambda}'_{n,j}/\widehat{\lambda}_{n,j}$ to obtain a smooth consistent estimate $\psi_{n,j}$ of λ'/λ. In view of their consistency results on rank estimators defined from rank statistics of the form (3.4), one can use such rank estimators for the preliminary estimates $b_{n,j}$. Using smooth consistent estimates $\psi_{n,1}, \psi_{n,2}$ of λ'/λ in the adaptive rank statistics $S^*_n(b)$, they also showed that the adaptive rank estimator $\widehat{\beta}_n$, which is defined as a minimizer of $S^*_n(b)$ in some small neighborhood of $(b_{n,1}+b_{n,2})/2$, is asymptotically normal; in fact, $\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{\mathcal{L}} N(0, (I_f(\beta_0))^{-1})$ under (2.31)–(2.34) and the following assumptions (in which σ_0 and σ_1 are given in (2.34)):

$$\begin{gathered} f \text{ is twice continuously differentiable with } \sup_s |f''(s)| < \infty \text{ and} \\ \int_{-\infty}^{\infty} \sup_{|t|\le\eta} [f'(s+t)/f(t)]^2 dF(s) < \infty \text{ for some } \eta > 0; \end{gathered} \tag{3.8}$$

$$I_f(\beta_0), \text{ defined by (2.25) and (2.45), is nonsingular;} \tag{3.9}$$

$$\sup_i (E|c^*_i|^q I_{\{c^*_i \le 0\}} + E|t^*_i|^q I_{\{-\infty < t^*_i \le 0\}}) < \infty \text{ for some } q > 0; \tag{3.10}$$

$$\begin{aligned} \lim_{m\to\infty} m^{-(1-\lambda)} \sum_{i=1}^m \big[&P\{t^*_i - \beta_0^T x_i < \sigma_0 - \delta\} I_{\{F(\sigma_0)>0\}} \\ &+ P\{c^*_i - \beta_0^T x^*_i > \sigma_1 + \delta\} I_{\{F(\sigma_1)<1\}}\big] = 0 \text{ for every } \delta > 0; \end{aligned} \tag{3.11}$$

$$\sup_{\|b\|\le\rho,-\infty<s<\infty}\sum_{i=1}^{m}[P\{s\le t_i^*-b^Tx_i^*\le s+h\}+P\{s\le c_i^*-b^Tx_i^*\le s+h\}]$$
$$=O(mh)\quad\text{as } m\to\infty \text{ and } h\to 0 \text{ such that } mh\to\infty. \tag{3.12}$$

A refinement of the proof of the asymptotic normality of the adaptive rank estimator $\widehat{\beta}_n$ in Lai and Ying (1991b) gives the following stronger uniformity result which is related to the information bounds in Theorem 2. For every $A \ge A_0$ (sufficiently large), choose a twice continuously differentiable function $\eta_A : \mathbf{R} \to \mathbf{R}^d$ satisfying (2.26a)–(2.26c) and the additional condition

$$\sup_s \left(\|\frac{d}{ds}\eta_A(s)\| + \|\frac{d^2}{ds^2}\eta_A(s)\|\right) \le A. \tag{3.13}$$

Let $\lambda_{\theta,A}(t) = \lambda(t) + \theta^T\eta_A(t)$ and $\phi_{\theta,A}(t) = \lambda_{\theta,A}(t)\exp\{-\int_{-\infty}^t \lambda_{\theta,A}(s)ds\}$, as in Theorem 2. Let $0<\epsilon<1/2$. From (2.26a), it follows that $\lambda_{\theta,A}\ge 0$ for $\|\theta\|\le n^{-\epsilon}$ and $An^{-\epsilon}\le 1/A$. Take $0<\delta\le\epsilon/2$ and let

$$\mathcal{F}_n = \{\phi_{\theta,A} : \|\theta\| \le n^{-\epsilon},\ A_0 \le A \le n^{\delta}\}. \tag{3.14}$$

Clearly, $\mathcal{F}_n$ satisfies condition (2.27) for every fixed $A \ge A_0$. Making use of (2.26a)–(2.26c) together with (3.8) and (3.13), we can modify the arguments of Lai and Ying (1991b) to show that for all sufficiently small $\delta > 0$, as $n \to \infty$,

$$\sqrt{n}(\widehat{\beta}_n-\beta) \xrightarrow{\mathcal{L}} N(0,(I_f(\beta_0))^{-1}) \text{ under } P_{\beta,\phi},$$
$$\text{uniformly in } \|\beta-\beta_0\|\le n^{-\epsilon} \text{ and } \phi\in\mathcal{F}_n, \tag{3.15}$$

as will be explained in the Appendix. This implies that for any bounded continuous function $w : \mathbf{R}^d \to \mathbf{R}$,

$$\lim_{n\to\infty}\left\{\sup_{\|\beta-\beta_0\|\le n^{-\epsilon},\phi\in\mathcal{F}_n} E_{\beta,\phi}w(\sqrt{n}(\widehat{\beta}_n-\beta))\right\} = E\{w(N(0,(I_f(\beta_0))^{-1}))\},$$

and therefore $\widehat{\beta}_n$ attains the asymptotic lower bound on minimax risks given in Theorem 1(i).

Using Lemma 1 above to replace Lemma 2 of Lai and Ying (1991b), we can also extend the arguments of that paper to prove that (3.15) holds for the adaptive rank estimator $\widehat{\beta}_n$ in the setting of (1.3) with independent (t_i, c_i, x_i), under the assumptions (2.6), (2.8) and (3.8)–(3.12) in which we replace t_i^*, c_i^*, x_i^* by t_i, c_i, x_i. It is interesting to compare the limiting covariance matrix $(I_f(\beta_0))^{-1}$ with that of the maximum likelihood estimator which assumes knowledge of f, given by $(I(\beta_0))^{-1}$ in (2.11). Since

$$I(\beta_0) = \int_{-\infty}^{\infty} (\lambda'/\lambda)^2 \Gamma_2 dF \text{ and } I_f(\beta_0) = \int_{-\infty}^{\infty} (\lambda'/\lambda)^2 (\Gamma_2 - \Gamma_1^2/\Gamma_0) dF, \quad (3.16)$$

it follows that $\widehat{\beta}_n$ is asymptotically as efficient as the maximum likelihood estimator when $\Gamma_1 \equiv 0$, which by (2.8) is the case if

$$Ex_i = 0 \text{ and } x_i \text{ is independent of } \left(t_i - \beta_0^T x_i, c_i - \beta_0^T x_i\right) \text{ for every } i. \quad (3.17)$$

Thus, under (3.17), adaptive estimation (cf. Bickel (1982) for the definition) can be accomplished by using the rank estimators $\widehat{\beta}_n$. In general, (3.16) implies that $I(\beta_0) - I_f(\beta_0)$ is nonnegative definite and adaptive estimation may not be possible; however, Theorem 1 shows that $(I_f(\beta_0))^{-1}$ is still asymptotically minimal when f is unknown.

4. Asymptotically Efficient Estimators of Both the Slope and the Intercept in the Absence of Truncation and Censoring

In this section we first specialize the results of Sections 2 and 3 to the case $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$, i.e., the y_i^* are completely observable. Since $(x_i^*, y_i^*) = (x_i, y_i)$ in this case, we will simply write x_i, y_i instead of x_i^*, y_i^* in the sequel. Under the assumption (2.7), let

$$J_f = \int_{-\infty}^{\infty} (f'/f)^2 dF \quad (4.1)$$

denote the Fisher information number. Since $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$, the assumption (2.33) is automatically satisfied with $\Gamma^*(s) = 0$, while $\Gamma_0^*(s) = 1$ in (2.32) which then reduces to the assumption

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Ex_i = \Gamma_1, \ \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E\left(x_i x_i^T\right) = \Gamma_2. \quad (4.2)$$

Since $\Gamma_0^*(s) = 1$, (2.34) clearly holds with $\sigma_0 = -\infty$ and $\sigma_1 = F^{-1}(1)$. Moreover, the assumptions (3.10)–(3.12) are trivally satisfied since $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$. Since $\int_{-\infty}^{\infty} (\lambda'/\lambda)^2 dF = \int_{-\infty}^{\infty} (f'/f)^2 dF$ (cf. Efron and Johnstone (1990)), the definition (2.25) of $I_f(\beta_0)$ reduces to

$$I_f(\beta_0) = J_f\left(\Gamma_2 - \Gamma_1^2\right), \text{ where } \Gamma_1^2 = \Gamma_1 \Gamma_1^T. \quad (4.3)$$

Hence, under the assumptions (2.7), (2.12), (2.31) and (4.2), the conclusions of Theorem 1 hold, assuming that $I_f(\beta_0)$ defined in (4.3) is positive definite. Moreover, if (2.7) and (2.12) are replaced by the stronger assumption (3.8), then we can construct adaptive rank estimators $\widehat{\beta}_n$ of β as in Section 3 so that (3.15) holds.

In the present setting of completely observable y_i^*, it is usually assumed that the ϵ_i^* in (1.1) has a finite variance and (1.1) is usually written in the form

$$y_i = \alpha + \beta^T x_i + \epsilon_i, \tag{4.4}$$

where ϵ_i are i.i.d. random variables with mean 0 and variance σ^2. Letting $\epsilon_i^* = \alpha + \epsilon_i$, the preceding discussion gives asymptotically efficient estimates of β when the underlying distribution of the ϵ_i (and therefore of the ϵ_i^* also) is unknown. We now consider asymptotically efficient estimation of the intercept α.

Since $\alpha = E(y_i - \beta^T x_i)$, an obvious estimate of α is

$$\widehat{\alpha}_n = n^{-1} \sum_{i=1}^{n} (y_i - \widehat{\beta}_n^T x_i), \tag{4.5}$$

where $\widehat{\beta}_n$ is the adaptive rank estimator described in Section 3 such that

$$\sqrt{n}(\widehat{\beta}_n - \beta) \xrightarrow{\mathcal{L}} N(0, (\Gamma_2 - \Gamma_1^2)^{-1}/J_f). \tag{4.6}$$

From (4.4) and (4.5), it follows that

$$\widehat{\alpha}_n - \alpha = n^{-1}\Big\{ \sum_{i=1}^{n} \epsilon_i - (\widehat{\beta}_n - \beta)^T \sum_{i=1}^{n} x_i \Big\}. \tag{4.7}$$

A refinement of the analysis of $\widehat{\beta}_n$ in Lai and Ying (1991b) shows that

$$\begin{pmatrix} \sum_1^n \epsilon_i/\sqrt{n} \\ \sqrt{n}(\widehat{\beta}_n - \beta) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & (\Gamma_2 - \Gamma_1^2)^{-1}/J_f \end{pmatrix}\right). \tag{4.8}$$

An outline of the proof is given in the Appendix. From (4.7) and (4.8), it follows that

$$\sqrt{n}(\widehat{\alpha}_n - \alpha) \xrightarrow{\mathcal{L}} N(0, \sigma^2 + \Gamma_1^T(\Gamma_2 - \Gamma_1^2)^{-1}\Gamma_1/J_f). \tag{4.9}$$

Let g be the common density function of the ϵ_i. Note that $\int_{-\infty}^{\infty}(g'/g)^2 g dt = J_f$. If g is known, the maximum likelihood estimators of α and β are asymptotically normal and the covariance matrix of their limiting normal distribution is

$$\begin{pmatrix} 1 & \Gamma_1^T \\ \Gamma_1 & \Gamma_2 \end{pmatrix}^{-1} / J_f = \begin{pmatrix} \Gamma_1^T(\Gamma_2 - \Gamma_1^2)^{-1}\Gamma_1 & -\Gamma_1^T(\Gamma_2 - \Gamma_1^2)^{-1} \\ -(\Gamma_2 - \Gamma_1^2)^{-1}\Gamma_1 & (\Gamma_2 - \Gamma_1^2)^{-1} \end{pmatrix} / J_f. \tag{4.10}$$

In view of (4.6), this implies that $\widehat{\beta}_n$ is asymptotically as efficient as the maximum likelihood estimator of β that assumes knowledge of g. Hence adaptive estimation of β is possible, as was already noted by Bickel (1982) under the additional assumption that the covariates x_i are i.i.d.

From (4.9) and (4.10), it follows that the limiting distribution of $\sqrt{n}(\widehat{\alpha}_n - \alpha)$ has a larger variance than that of the maximum likelihood estimator which assumes knowledge of g, and the difference between the two limiting variances is $\sigma^2(= E\epsilon_1^2)$. We now show that the limiting variance in (4.9) is still asymptotically minimal when f is unknown by establishing information bounds for this semiparametric estimation problem.

Let α_0, β_0 be the unknown values of the parameters α, β. We shall embed f in a parametric family of the form

$$f_\alpha(t) = f(t)\{1 + (\alpha - \alpha_0)\psi(t)\}, \tag{4.11}$$

so that $f_{\alpha_0} = f$, where ψ is a function satisfying

$$\int_{-\infty}^{\infty} \psi(t)f(t)dt = 0, \quad \int_{-\infty}^{\infty} t\psi(t)f(t)dt = 1. \tag{4.12}$$

Note that (4.12) ensures that $\int_{-\infty}^{\infty} f_\alpha(t)dt = 1$ and $\int_{-\infty}^{\infty} tf_\alpha(t)dt = \alpha$. For this parametric family the log-likelihood function is

$$\ell_n(\alpha, \beta) = \sum_{i=1}^{n} \log f_\alpha(y_i - \beta^T x_i), \tag{4.13}$$

and therefore by the central limit theorem, under (2.7), (4.2) and suitable regularity conditions on ψ,

$$\frac{1}{\sqrt{n}}\begin{pmatrix}(\partial/\partial\alpha)\ell_n(\alpha_0,\beta_0)\\(\partial/\partial\beta)\ell_n(\alpha_0,\beta_0)\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\psi(\epsilon_i)\\ [(-f'/f)(\epsilon_i)]x_i\end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \mathbf{I}_\psi(\theta)),$$
$$\text{where } \mathbf{I}_\psi(\theta) = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta}^T \\ I_{\alpha\beta} & I_{\beta\beta}\end{pmatrix}, \tag{4.14}$$
$$I_{\alpha\alpha} = \int_{-\infty}^{\infty} \psi^2 dF, \; I_{\alpha\beta} = \Gamma_1 \int_{-\infty}^{\infty} (-f'\psi/f)dF, \; I_{\beta\beta} = J_f\Gamma_2.$$

Assuming $\mathbf{I}_\psi$ to be nonsingular, the top right entry of $\mathbf{I}_\psi^{-1}$ is

$$\begin{aligned}&(I_{\alpha\alpha} - I_{\alpha\beta}^T I_{\beta\beta}^{-1} I_{\alpha\beta})^{-1}\\ =&\Big\{\int_{-\infty}^{\infty} \psi^2 dF - (\Gamma_1^T\Gamma_2^{-1}\Gamma_1)\Big[\int_{-\infty}^{\infty}(-f'\psi/f)dF\Big]^2/J_f\Big\}^{-1}.\end{aligned} \tag{4.15}$$

The constraint (4.12) is satisfied by the functions $\psi(t) = -f'(t)/f(t)$ and $\psi(t) = (t-\alpha_0)/\sigma^2$, and therefore also by

$$\psi(t) = c\frac{-f'(t)}{f(t)} + (1-c)\frac{t-\alpha_0}{\sigma^2}, \tag{4.16}$$

where the constant c will be specified later. For ψ given by (4.16), we have

$$\int_{-\infty}^{\infty} \psi^2 dF = c^2 J_f + (1-c^2)\sigma^{-2}, \quad \int_{-\infty}^{\infty} (-f'\psi/f) dF = cJ_f + (1-c)\sigma^{-2}.$$

Putting these values in (4.15) and then maximizing (4.15) with respect to c gives the maximizing value

$$c^* = \frac{\Gamma_1^T\Gamma_2^{-1}\Gamma_1}{(1-\Gamma_1^T\Gamma_2^{-1}\Gamma_1)\sigma^2 J_f + \Gamma_1^T\Gamma_2^{-1}\Gamma_1}, \tag{4.17}$$

for which (4.15) reduces to $\sigma^2 + \Gamma_1^T(\Gamma_2 - \Gamma_1^2)^{-1}\Gamma_1/J_f$, which is the limiting variance in (4.9).

Let ψ^* denote the function (4.16) in which c is given by (4.17). Under the assumptions (2.7) and $0 < \sigma^2 < \infty$, we can choose for every $A \ge A_0$ (sufficiently large) a continuous function $\psi_A : \mathbf{R} \to \mathbf{R}$ such that

$$\psi_A(s) = 0 \text{ if } f(s) \le 1/A, \ \sup_s |\psi_A(s)| \le A, \tag{4.18a}$$

$$\mathbf{I}_{\psi_A} \text{ is positive definite, where } \mathbf{I}_{\psi_A} \text{ is defined in (4.14)}, \tag{4.18b}$$

$$\int_{-\infty}^{\infty} \psi_A(t)f(t)dt = 0, \ \int_{-\infty}^{\infty} t\psi_A(t)f(t)dt = 1, \tag{4.18c}$$

$$\lim_{A\to\infty} \int_{-\infty}^{\infty} (\psi_A - \psi^*)^2 dF = 0. \tag{4.18d}$$

Take any ϵ satisfying $0 < \epsilon < 1/2$. For $|\alpha - \alpha_0| \le n^{-\epsilon}$ and all sufficiently large n, $f_{\alpha,A}(t) = f(t)\{1 + (\alpha-\alpha_0)\psi_A(t)\}$ is nonnegative and in view of (4.18c) is also a density function. For fixed $A \ge A_0$, denoting $f_{\alpha,A}$ by f_α and defining $\ell_n(\alpha,\beta)$ as in (4.13), the same argument as the proof of (2.13) can be used to show that for any $u \in \mathbf{R}$ and $v \in \mathbf{R}^d$,

$$\ell_n(\alpha_0+u/\sqrt{n}, \beta_0+v/\sqrt{n}) - \ell_n(\alpha_0,\beta_0) \xrightarrow{\mathcal{L}} N\Big(-\frac{1}{2}(u,v^T)\mathbf{I}_{\psi_A}\binom{u}{v}, (u,v^T)\mathbf{I}_{\psi_A}\binom{u}{v}\Big).$$

Hence, as in Theorem 1, we have the following.

Theorem 3. *Suppose that in the linear regression model (4.4) the ϵ_i are* i.i.d. *random variables with mean 0, variance $\sigma^2 (< \infty)$ and a continuously differentiable density function g_0 such that f satisfies (2.7) and (2.12), where $f(u) =$*

$g_0(u-\alpha_0)$. *Suppose that the* x_i *are independent random vectors that are independent of* $\{\epsilon_n\}$ *and such that* (2.31) *and* (4.2) *hold. Defining* J_f *by* (4.1) *and assuming that* $\Gamma_2 - \Gamma_1^2$ *is nonsingular, let*

$$v_f(\alpha_0) = \sigma^2 + \Gamma_1^T(\Gamma_2 - \Gamma_1^2)^{-1}\Gamma_1/J_f. \tag{4.19}$$

Let $0 < \epsilon < 1/2$ *and let* $\mathcal{G}_n$ *be a family of densities of* ϵ_1 *such that for every given* $A \geq A_0$,

$$f_{\alpha,A}(\cdot + \alpha) \in \mathcal{G}_n \text{ for } |\alpha - \alpha_0| \leq n^{-\epsilon} \text{ and all large } n. \tag{4.20}$$

(i) *Let* $w : \mathbf{R} \to \mathbf{R}$ *be a bounded continuous function satisfying* (2.14). *For any estimator* T_n *of* α,

$$\begin{aligned}&\liminf_{n\to\infty}\left\{\sup_{|\alpha-\alpha_0|\vee\|\beta-\beta_0\|\leq n^{-\epsilon},\ g\in\mathcal{G}_n} E_{\alpha,\beta,g}w(\sqrt{n}(T_n-\alpha))\right\}\\ &\geq Ew(Z), \text{ where } Z \sim N(0, v_f(\alpha_0)).\end{aligned} \tag{4.21}$$

(ii) *Suppose that an estimator* T_n *is regular in the following sense: For any sequence* $(\alpha_n, \beta_n, g_n)_{n\geq 1}$ *with* $\sqrt{n}(\alpha_n - \alpha_0)$ *and* $\sqrt{n}(\beta_n - \beta_0)$ *converging to some limits and* $g_n \in \mathcal{G}_n$, $\sqrt{n}(T_n - \alpha_n)$ *has a limiting distribution* Q *under* P_{α_n,β_n,g_n}, *where* Q *does not depend on the particular sequence* $(\alpha_n, \beta_n, g_n)_{n\geq 1}$. *Then* $Q = N(0, v_f(\alpha_0)) * H$ *for some distribution* H.

In particular, let $\mathcal{G}_n = \{f_{\alpha,A}(\cdot + \alpha) : |\alpha - \alpha_0| \leq n^{-\epsilon}, A_0 \leq A \leq n^\delta\}$ with $0 < \delta \leq \epsilon$. It can be shown that for the estimator $\widehat{\alpha}_n$ defined by (4.5), the asymptotic normality result (4.9) can be strengthened to:

$$\begin{aligned}&\sqrt{n}(\widehat{\alpha}_n - \alpha) \xrightarrow{\mathcal{L}} N(0, v_f(\alpha_0)) \text{ under } P_{\alpha,\beta,g},\\ &\qquad\text{uniformly in } |\alpha-\alpha_0| \vee \|\beta-\beta_0\| \leq n^{-\epsilon} \text{ and } g \in \mathcal{G}_n,\end{aligned} \tag{4.22}$$

as $n \to \infty$, for all sufficiently small δ, as will be explained in the Appendix. This implies that the estimator $\widehat{\alpha}_n$ attains the asymptotic lower bound (4.21) on minimax risks for any bounded continuous loss function w, and therefore $\widehat{\alpha}_n$ is asymptotically efficient.

Acknowledgement

This research was supported by the National Science Foundation, the National Security Agency and the Air Force Office of Scientific Research.

Appendix

We outline here the proof of (4.8) and of the uniformity results (4.22) and (3.15). The arguments are basically modifications and extensions of those in Lai and Ying (1991b), which we shall abbreviate by LY.

Proof of (4.8). Specializing the proof of Theorem 2 and Corollary 2 in LY to the present setting in which $t_i^* \equiv -\infty$ and $c_i^* \equiv \infty$, it can be shown that

$$\begin{aligned}
&nI_f(\beta)(\widehat{\beta}_n - \beta) = -S_n^*(\beta) + o_p(\sqrt{n}) \\
&= -\sum_{i=1}^{n_i} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} p_n\Big(n^{-1}\sum_{j=1}^{n_1} I_{\{\epsilon_j^* \geq s\}}\Big)\Big(x_i - \frac{\sum_{j=1}^{n_1} x_j I_{\{\epsilon_j^* \geq s\}}}{\sum_{j=1}^{n_1} I_{\{\epsilon_j^* \geq s\}}}\Big) dM_i(s) \\
&- \sum_{i=n_1+1}^{n} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} p_n\Big(n^{-1}\sum_{j=n_1+1}^{n} I_{\{\epsilon_j^* \geq s\}}\Big)\Big(x_i - \frac{\sum_{j=n_1+1}^{n} x_j I_{\{\epsilon_j^* \geq s\}}}{\sum_{j=n_1+1}^{n} I_{\{\epsilon_j^* \geq s\}}}\Big) dM_i(s) \\
&+ o_p(\sqrt{n}),
\end{aligned} \tag{A.1}$$

where $n_1 = [n/2]$ and $M_i(s) = I_{\{\epsilon_i^* \leq s\}} - \int_{-\infty}^{s} \lambda(u) I_{\{\epsilon_i^* \geq u\}} du$. Since $\{M_i(s), \mathcal{B}_s, -\infty < s < \infty\}$ is a martingale by Lemma 1, Lenglart's (1977) inequality can be applied in conjunction with empirical process theory (cf. Lai and Ying (1988, 1991b)) to show that (A.1) can be further approximated by the more tractable expression

$$\begin{aligned}
nI_f(\beta)(\widehat{\beta}_n - \beta) &= -\sum_{i=1}^{n_1} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} (x_i - \bar{x}_{n_1}) dM_i(s) \\
&\quad - \sum_{i=n_1+1}^{n} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} (x_i - \bar{x}_{n,n_1}) dM_i(s) + o_p(\sqrt{n}) \\
&= U_{n_1} + U_{n,n_1} + o_p(\sqrt{n}), \text{ say,}
\end{aligned} \tag{A.2}$$

where $\bar{x}_{n_1} = n^{-1}\sum_{i=1}^{n_1} x_i$, $\bar{x}_{n,n_1} = (n - n_1)^{-1}\sum_{i=n_1+1}^{n} x_i$. Note that $(\sum_1^{n_1} \epsilon_i^*, U_{n_1})$ is a sum of independent random vectors and is independent of $(\sum_{i=n_1+1}^{n} \epsilon_i^*, U_{n,n_1})$ which is another sum of independent random vectors. Moreover, since $\{x_i\}$ and $\{\epsilon_i^*\}$ are independent,

$$\begin{aligned}
\mathrm{Cov}\Big(\sum_1^n \epsilon_i^*, U_{n_1}\Big) &= -\sum_{i=1}^{n_1} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} E\{(x_i - \bar{x}_{n_1})(\epsilon_i^* - \alpha) dM_i(s)\} \\
&= -\sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\lambda'(s)}{\lambda(s)} E(x_i - \bar{x}_{n_1}) E\{(\epsilon_i^* - \alpha) dM_i(s)\} = 0.
\end{aligned}$$

A similar argument shows that $\mathrm{Cov}(\sum_{n_1+1}^{n} \epsilon_i^*, U_{n,n_1}) = 0$. The desired conclusion then follows from (A.2) and the central limit theorem.

Proof of (4.22). Let $\mathcal{F}_n = \{f_{\alpha,A} : |\alpha - \alpha_0| \leq n^{-\epsilon}, A_0 \leq A \leq n^{\delta}\}$, which is a family of densities of $\epsilon_1^* = \epsilon_1 + \alpha$. A refinement of the proof of (A.1) shows that the $o_p(\sqrt{n})$ term in (A.1) under (1.1), in which ϵ_i^* has density ϕ, is uniform in

$||\beta - \beta_0|| \le n^{-\epsilon}$ and $\phi \in \mathcal{F}_n$ if $\delta > 0$ is sufficiently small. Also empirical process theory and Lenglart's (1977) inequality show that the $o_p(\sqrt{n})$ term in (A.2) under (1.1) in which ϵ_i^* has density ϕ is uniform in $||\beta - \beta_0|| \le n^{-\epsilon}$ and $\phi \in \mathcal{F}_n$. The desired conclusion then follows from (4.7) by applying a uniform central limit theorem for sums of independent random vectors satisfying a uniform Lindeberg condition (cf. Theorem 1 of Lai (1977)).

Proof of (3.15). Modify the proof of Theorem 2 and Corollary 2 of LY to establish (A.1) and (A.2) in which the $o_p(\sqrt{n})$ terms under $P_{\beta,\phi}$ are uniform in $||\beta - \beta_0|| \le n^{-\epsilon}$ and $\phi \in \mathcal{F}_n$. Then apply a uniform central limit theorem to $U_{n_1} + U_{n,n_1}$.

References

Amemiya, A. (1985). *Advanced Econometrics.* Harvard University Press.

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric nonparametric models. *Ann. Statist.* **11**, 432-452.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647-671.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **64**, 269-276.

Efron, B. and Johnstone, I. M. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18**, 38-62.

Goldberger, A. S. (1981). Linear regression after selection. *J. Econometrics* **15**, 357-366.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14**, 323-330.

Hájek J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **1**, 175-194. University California Press.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York-Berlin-Heidelberg.

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

Koshevnik, Yu. A. and Levit, B. Ya. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21**, 738-753.

Lai, T. L. (1977). First exit times from moving boundaries for sums of independent random variables. *Ann. Probab.* **5**, 210-221.

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. (1991a). Estimating a distribution function with truncated and censored data. *Ann. Statist.* **19**, 417-442.

Lai, T. L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

Lai, T. L. and Ying, Z. (1991c). Linear rank statistics in regression analysis with censored or truncated data. *J. Multiavariate Anal.* **39**.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley, New York.

Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré* **13**, 171-179.

Nicoll, J. F. and Segal, I. E. (1980). Nonparametric estimation of the observational cutoff bias. *Astronom. and Astrophys.* **82**, L3-L6.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51**, 269-286.

Ritov, Y. (1984). Efficient and unbiased estimation in nonparametric linear regression. Technical Report, Department of Statistics, University of California at Berkeley.

Ritov, Y. and Wellner, J. A. (1988). Censoring, martingales and the Cox model. In *Contemporary Mathematics* **80**: *Statistical Inference for Stochastic Processes* (Edited by N. U. Prabhu), 191-220. Amer. Math. Soc., Providence.

Segal, I. E. (1975). Observational validation of the chronometric cosmology: I. Preliminaries and the red-shift magnitude relation. *Proc. Nat. Acad. Sci. USA* **72**, 2437-2477.

Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* 1, 187-196. University California Press.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290-295.

Wong, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14**, 88-123.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.
Department of Statistics, University of Illinois at Urbana, IL 61820, U.S.A.

(Received February 1991; accepted June 1991)