JOURNAL OF MULTIVARIATE ANALYSIS **49**, 179–201 (1994)

Linear Regression with Censoring

C. SRINIVASAN AND MAI ZHOU

Koul, Susarla and Van Ryzin (1981, *Ann. Statist.* **9**, 1276–1288) proposed a generalization of the ordinary least squares estimator in linear models with censored data. This paper uses counting processes and martingale techniques to provide a proof of the asymptotic normality of the estimator. A detailed analysis of the asymptotic variance is presented. © 1994 Academic Press, Inc.

1. INTRODUCTION

Suppose patients' survival times, or their logarithms, Y_i, under study, are random variables that are independent of each other and follow the linear model

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1, 2, ..., n, \tag{1.1}$$

where the X_i's are observable covariates and ε_i are iid random variables with zero mean and finite variance σ^2. The parameters (α, β) are to be estimated. We, however, observe only the censored data (Z_i, δ_i, X_i) with $Z_i = \min\{Y_i, C_i\}$, $\delta_i = I_{[Y_i \leqslant C_i]}$ where C_i's are censoring times, iid random variables that are independent of ε_i's. Let $P(Y_i \leqslant t) = F_i(t)$ and $P(C_i \leqslant t) = G(t)$.

Koul, Susarla, and Van Ryzin (hereafter abbreviated KSV) (1981) suggested the following estimate (α, β) based on the censored data (Z_i, δ_i, X_i), $i = 1, 2, ..., n$,

$$\hat{\alpha} = \sum_{i=1}^{n} a_{ni} \frac{\delta_i Z_i}{1 - \hat{G}(Z_i)} I_{[Z_i \leqslant M_n]}, \qquad \hat{\beta} = \sum_{i=1}^{n} b_{ni} \frac{\delta_i Z_i}{1 - \hat{G}(Z_i)} I_{[Z_i \leqslant M_n]}, \tag{1.2}$$

where

$$b_{ni} = \frac{X_i - \bar{X}}{\sum_{i=1}^{n} (X_i - \bar{X})^2}, \; a_{ni} = \frac{1}{n} - \bar{X} b_{ni}, \qquad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

Received September 25, 1991; revised May 30, 1993.
AMS 1980 subject classifications: primary 62G10, secondary 62P10, 62N05.
Key words and phrases: Censored data, regression, martingale, asymptotic distribution.

0047-259X/94 $6.00

and $\hat{G}$ is a Kaplan–Meier product-limit type estimator of the distribution $G(t)$ of the C_i's that are defined later. Here M_n, $n = 1, 2, \ldots$, is a sequence of truncation constants tending to infinity at certain rate. This estimator is particularly easy to implement on a computer.

In their 1981 paper, KSV used a U-statistic representation to derive the asymptotic distribution of the estimates by appealing to Hoeffding (1948). Our approach uses the martingale structure of the counting processes associated with the underlying problem and develops a martingale representation for the estimator, modulo an error term which is $o_p(1)$. This obviously has many strengths and advantages. The first is that the martingale representation is easily adaptable for a sequential analysis of the model. The second concerns the truncation constants. As noted by Miller and Halpern (1982) and Gill (1983), constants M_n depend on the unknown $F_i(t)$ and $G(t)$ in a complicated way, and therefore the choice of their values requires further guidelines and experience. In our approach, as opposed to theirs, the truncation sequence is simply taken to be the observable $T^n = \max\{Z_i\}$ and, hence, there is really no truncation in (1.2). In addition, our discussion on the asymptotic variance of $(\hat{\alpha}, \hat{\beta})$ reveals that the KSV formula (3.7) needs an extra n factor in the second term there. Thus the second part of their (3.7) (negative part) cannot be neglected, and therefore their asymptotic variance estimator needs to be adjusted.

Miller and Halpern (1982) also find that the KSV estimator performs poorly compared to other estimators in the context of the Stanford Heart Transplant Data. However, Fygenson and Zhou (1992, 1994) have suggested some simple modifications to overcome this shortcoming and showed that the modified KSV estimator performs equally well in these situations. Our theoretical study of the asymptotic variance also indicates that under certain scenarios the KSV (modified) can be quite good. It is also known that in the special case of two samples, the KSV estimator is identical to the Buckley and James (1979) estimator which is known to be efficient under the normal errors. See Lai and Ying (1991).

Another, perhaps the most popular, approach to the regression problem is the Cox model and partial likelihood analysis. Although the Cox model has many nice features it cannot, unlike in a classical linear regression model, adapt for unexplained heterogeneity. See Struthers and Kalbfleisch (1986) for details. The KSV linear regression method that we study in this paper can easily handle heterogeneity, in the sense that the ε_i's need not be iid but only independent and satisfy certain moment conditions.

Section 2 below contains additional notation as well as some counting process results that are useful later. Section 3 contains the asymptotic normality theorem with proof, but the more technical treatment of the high-order terms is deferred to Section 5 so that one can better concentrate on the martingale representation part of the proof. Extension to multiple regression is straightforward. Also, the method can be readily extended to handle the heteroscedastic case. In Section 4 we take a closer look at the asymptotic variance derived in Section 3 and point out a correction to a formula in KSV and discuss some consequences of that.

2. Notation and Counting Processes

We now introduce some additional notation and establish some needed facts. For $i = 1, 2, ..., n$, let

$$1 - H_i(t) = P(Z_i > t) = [1 - F_i(t)]\,[1 - G(t)],$$

where $F_i(t)$ and $G(t)$ are as defined in Section 1. In addition, set

$$1 - \hat{H}_i(t) = I_{[Z_i \geqslant t]};\ R^+(t) = \sum_{i=1}^{n} [1 - \hat{H}_i(t)] \qquad \text{and} \qquad T^n = \max\{Z_i\}. \tag{2.1}$$

Also, let

$$\Lambda_i^D(t) = \int_{[-\infty, t]} \frac{dF_i(s)}{1 - F_i(s-)},\ \Lambda^C(t) = \int_{[-\infty, t]} \frac{dG(s)}{1 - G(s-)},$$

and

$$\Lambda_i^+(t) = \int_{[-\infty, t]} \frac{dH_i(s)}{1 - H_i(s-)}.$$

It is well known that the three processes

$$M_i^D(t) = I_{[Z_i \leqslant t;\, \delta_i = 1]} - \int_{-\infty}^{t} I_{[Z_i \geqslant s]}\, d\Lambda_i^D(s);$$

$$M_i^C(t) = I_{[Z_i \leqslant t,\, \delta_i = 0]} - \int_{-\infty}^{t} I_{[Z_i \geqslant s]}\, d\Lambda^C(s)$$

and (2.2)

$$M_i^+(t) = I_{[Z_i \leqslant t]} \int_{-\infty}^{t} I_{[Z_i \geqslant s]}\, d\Lambda_i^+(s)$$

are all square integrable martingales on $[-\infty, +\infty]$ with respect to the filtration

$$\mathscr{F}_t = \sigma\{Z_k I_{[Z_k \leqslant t]}, \delta_k I_{[Z_k \leqslant t]}, 1 \leqslant k \leqslant n\},$$

and their predictable variations are, respectively,

$$\langle M_i^D\rangle(t)=\int_{-\infty}^{t} I_{[Z_i\geqslant s]}\,d\Lambda_i^D(s),$$

$$\langle M_i^C\rangle(t)=\int_{-\infty}^{t} I_{[Z_i\geqslant s]}\,d\Lambda^C(s),$$

and (2.3)

$$\langle M_i^+\rangle(t)=\int_{-\infty}^{t} I_{[Z_i\geqslant s]}\,d\Lambda_i^+(t).$$

Furthermore,

$$\langle M_i^D, M_i^C\rangle=0. \tag{2.4}$$

See, e.g., Gill (1980) for a proof of these facts.

It is not hard to see that $\Lambda_i^+(t)=\Lambda_i^D(t)+\Lambda^C(t)$, and as a consequence we have

$$M_i^+(t)=M_i^D(t)+M_i^C(t). \tag{2.5}$$

Furthermore, define

$$M_C^+(t)=\sum_{i=1}^{n} M_i^C(t). \tag{2.6}$$

The Kaplan–Meier estimator of $G(t)$ is then given by

$$\hat{G}(t)=\hat{G}_k(t)=1-\prod_{s\leqslant t}\left(1-\frac{\Delta N_C^+(s)}{R^+(s)}\right), \tag{2.7}$$

where $N_C^+(s)=\sum I_{[Z_i\leqslant s,\delta_i=0]}$ and $\Delta N_C^+(s)=N_C^+(s+)-N_C^+(s-)$. And we have the following representations:

$$\frac{\hat{G}(T^n\wedge t)-G(T^n\wedge t)}{1-G(T^n\wedge t)}=\frac{\hat{G}(t)-G(t)}{1-G(t)}$$

$$=\int_{-\infty}^{t}\frac{1-\hat{G}(s-)}{1-G(s)}\frac{I_{[R^+>0]}}{R^+(s)}\,dM_C^+(s)$$

$$\text{for}\quad t\in[-\infty, T^n], \tag{2.8}$$

$$\frac{\hat{H}_i(t)-H_i(t)}{1-H_i(t)}=\int_{-\infty}^{t}\frac{1}{1-H_i(s)}\,dM_i^+(s)$$

$$\text{for } t \text{ such that } 1-H_i(t)>0. \tag{2.9}$$

Therefore, these two processes (2.8) and (2.9) are local martingales in their domains.

3. Asymptotic Normality

Throughout the rest of this paper, we take the definition of the estimator $(\hat{\alpha}, \hat{\beta})$ in (1.2) with $M_n = T^n$ and $\hat{G}(\cdot)$ to be the left continuous version of the Kaplan–Meier estimator $\hat{G}_k(\cdot-)$ in (2.8); i.e.,

$$\hat{\alpha} = \sum a_{ni} \frac{\delta_i Z_i}{1 - \hat{G}_k(Z_i -)} I_{[Z_i \leqslant T^n]}, \qquad \hat{\beta} = \sum b_{ni} \frac{\delta_i Z_i}{1 - \hat{G}_k(Z_i -)} I_{[Z_i \leqslant T^n]}. \tag{3.1}$$

We drop the subscript k in $\hat{G}_k$ in the subsequent formulae.

It is perhaps worth noting that KSV (1981) used a variation of $\hat{G}$ in their proof for technical reasons, but this does not alter the limiting behavior of the estimator. On the other hand, the Kaplan–Meier estimator is probably more convenient in practice as it is most widely available on computing packages. Note, also, that the restriction $[Z_i \leqslant T^n]$ is in fact an empty requirement.

We begin by rewriting the estimator (3.1). Note

$$\begin{aligned} \frac{Z_i \delta_i}{1 - \hat{G}(Z_i -)} &= \int_{-\infty}^{T^n} \frac{t}{1 - \hat{G}(t-)} \, dI_{[Z_i \leqslant t, \delta_i = 1]} \\ &= \int_{-\infty}^{T^n} \frac{t}{1 - \hat{G}(T^n \wedge t -)} \, dI_{[Z_i \leqslant t, \delta_i = 1]}. \end{aligned} \tag{3.2}$$

The integrals are Lebesgue–Steiltjes integrals. Note we can and did stop above integral at T^n since $Z_i \leqslant T^n$ $\forall i$.

Using (3.2), the KSV estimator (3.1) can be written as

$$\begin{aligned} \hat{\alpha} &= \sum a_{ni} \int_{-\infty}^{T^n} \frac{t}{1 - \hat{G}(T^n \wedge t -)} \, dI_{[Z_i \leqslant t, \delta_i = 1]}; \\ \hat{\beta} &= \sum b_{ni} \int_{-\infty}^{T^n} \frac{t}{1 - \hat{G}(T^n \wedge t -)} \, dI_{[Z_i \leqslant t, \delta_i = 1]}. \end{aligned} \tag{3.3}$$

The centering quantities for (3.3) are

$$\alpha(T^n) = \sum a_{ni} \int_{-\infty}^{T^n} t \, dF_i(t), \qquad \beta(T^n) = \sum b_{ni} \int_{-\infty}^{T^n} t \, dF_i(t).$$

Therefore, we can write

$$\begin{aligned} \hat{\alpha} - \alpha(T^n) &= (\hat{\alpha} - \alpha)(T^n) = \sum a_{ni} \int_{-\infty}^{T^n} \left(\frac{t \, dI_{[Z_i \leqslant t, \delta_i = 1]}}{1 - \hat{G}(T^n \wedge t -)} - t \, dF_i(t) \right) \\ \hat{\beta} - \beta(T^n) &= (\hat{\beta} - \beta)(T^n) = \sum b_{ni} \int_{-\infty}^{T^n} \left(\frac{t \, dI_{[Z_i \leqslant t, \delta_i = 1]}}{1 - \hat{G}(T^n \wedge t -)} - t \, dF_i(t) \right) \end{aligned} \tag{3.4}$$

More generally, we define two processes $(\hat{\alpha}-\alpha)(\tau)$ and $(\hat{\beta}-\beta)(\tau)$ for any constant τ to be the same expression in the right-hand side of (3.4) except that T^n is replaced by τ.

Remark. The centering quantities are not exactly equal to α and β but are truncated at T^n. In general, some extra conditions are needed before one can center the estimator by the exact value α, β. See also Remark 4.4 of KSV (1981).

Now we are ready to formulate the main theorem. For simplicity of presentation, we assume a random design in the censored linear model; i.e., $\mathbf{X}_i$ in (1.1) are iid. The definitions and the proofs of this paper should then be understood to be conditional on $\mathbf{X}_i$. The case of a fixed design can be treated similarly. Whatever the type of design one chooses to work with, we need one of the following conditions.

D1. For the random design, $0<\operatorname{Var}(X)=\sigma^2$ and $EX^4<\infty$.

D2. For the fixed design, for sufficiently large n, $0<K_1<(1/n)\sum_{i=1}^n (X_i-\bar{X})^2$, $(1/n)\sum X_i^4<K_2<\infty$.

Since we are working with the random design, we assume Condition D1 throughout the rest of the paper. Before we state the main theorem, we pause here to observe that Condition D1 implies, among other things, that $\max_{1\leqslant i\leqslant n} nb_{ni}^2\to 0$ a.e. and $\max_{1\leqslant i\leqslant n} nb_{ni}^2X_i^2\to 0$ a.e. (see, e.g., Problem 8 of Chow and Teicher (1987, Sect. 5.2)).

In addition to the design conditions, we need the following regularity conditions to handle the tail of the remainder terms in martingale representation.

Towards this, let

$$C(t)=\int_0^t \frac{1}{1-\bar{F}}\frac{dG}{(1-G)^2},$$

where $\bar{F}=\lim_{n\to\infty}(1/n)\sum F_i$. The conditions are

R1. For some $\tau>0$,

$$\sup_n \int_\tau^\infty \left[\sum b_{n_i} tf_i(t)\right]^2 dC(t)<\infty. \tag{3.5}$$

R2. For some $\tau>0$,

$$E\int_\tau^\infty |B(C(t))\,\bar{h}(t)|\,dt<\infty, \tag{3.6}$$

where $\bar{h}(t)=\lim\sum b_{ni} tf_i(t)$.

R3.

$$\lim_{t\to\infty}\lim_{n\to\infty} n\sum b_{ni}^2 \frac{[\int_t^\infty s\,dF_i(s)]^2}{1-H_i(t)}=0. \tag{3.7}$$

R4. For any $\tau>0$, $f_i(t)/(1-H_i(t))$ is of bounded variation on $[-\tau,\tau]$ uniformly in $i=1,2,...$.

R5. For any $\tau>0$

$$\int_\tau^\infty \sqrt{n\sum b_{ni}^2 \frac{t^2f_i^2(t)}{1-H_i}}\,dt\leqslant C<\infty. \tag{3.8}$$

A few comments are in order regarding Conditions R1–R5 before we proceed with the statement of the theorem. Condition R1 is equivalent to requiring that the variance of $\int_\tau^\infty \sum b_{ni}tf_i(t)\,dB(C(t))$ be finite uniformly in n. Here the time changed Brownian motion $B(C(t))$ appears naturally as the limit of the Kaplan–Meier process. A condition of this type also appears in Gill (1983). A sufficient condition for R1 is that X be bounded and for some $C>0$ and $\varepsilon>0$,

$$\sup_{t>\tau}\frac{t^2\bar{f}^2(t)}{(1-\bar{F})(1-G)^{1+\varepsilon}}<C,$$

where $\bar{f}$ is the density of $\bar{F}$. The condition R4 is seen to hold if either (i) the hazard function is of bounded variation on $\mathbb{R}^1$ or (ii) the covariate X is bounded with probability one. The tail conditions R2, R3, and R5 will hold if for all $i=1,2,...,$

$$1-G(t)>K(1-F_i)^\beta \qquad \text{for} \quad K>0,\ 0<\beta<1, \tag{3.9}$$

and uniformly in i

$$\int_\tau^\infty \frac{t\,dF_i(t)}{(1-F_i)^\alpha}<\infty \qquad \text{with} \quad \alpha=\frac{1+\beta}{2}. \tag{3.10}$$

The condition R3 will hold if, in addition to Assumption 2 of Theorem 3.1, $t^2(1-F_i(t))/(1-G(t))\to 0$ as $t\to\infty$. Since

$$\begin{aligned} n\sum b_{ni}^2\frac{[\int_t^\infty s\,dF_i(s)]^2}{1-H_i(t)} &= n\sum b_{ni}^2\frac{[t(1-F_i)+\int_t^\infty (1-F_i)\,ds]^2}{(1-F_i)(1-G)} \\ &\leqslant 2t^2\sum nb_{ni}^2\frac{1-F_i}{1-G}+2\sum nb_{ni}^2\frac{[\int_t^\infty (1-F_i)\,ds]^2}{(1-F_i)^2}\frac{1-F_i}{1-G} \\ &\leqslant K\sum nb_{ni}^2\frac{1-F_i}{1-G}+Kt^2\sum nb_{ni}^2\frac{1-F_i}{1-G}. \end{aligned}$$

It is clear that the above will $\to 0$ as $t \to \infty$ if the said condition holds. Roughly speaking, these conditions require two properties: that F_i have reasonably small tails and G have a suitably heavier tail compared to F_i. For example, if $1 - F_i(t) = e^{-\lambda_i t}$ and $1 - G(t) = e^{-\eta t}$ with $\lambda_i < M < \eta$ then these conditions hold.

THEOREM 3.1. *In the censored linear model* (1.1), *the KSV estimator* (3.1) *is asymptotically normally distributed, i.e.,*

$$\sqrt{n}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix}(T^n) \xrightarrow{\mathscr{D}} N\left(0, \sum(\infty)\right),$$

provided

1. *the variance–covariance matrix* $\sum(\tau)$ (*see definition below*) *is well-defined and finite for* $\tau \in [0, +\infty]$ *and* $\sum(\tau) \xrightarrow{\tau\to\infty} \sum(\infty)$;
2. $\sup_{0<t} E[\varepsilon_i - t \mid \varepsilon_i > t] < \infty$;
3. *the tail conditions* R1–R5 *given above hold.*

The variance–covariance matrix $\sum(\tau) = (V_{ij}(\tau))$ *is*

$$\begin{aligned} V_{22}(\tau) &= \lim n \sum b_{ni}^2 \int_{-\infty}^{\tau} \left(\frac{t}{1-G(t)} - \frac{\int_s^\tau s\, dF_i(s)}{1-H_i(t)}\right)^2 [1-G(t)]\, dF_i(t) \\ &\quad + \lim n \sum \int_{-\infty}^{\tau} \left(\frac{\sum b_{nj} \int_s^\tau s\, dF_j(s)}{\sum [1-H_j(t)]}\right. \\ &\quad \left. - \frac{b_{ni}\int_t^\tau s\, dF_i(s)}{1-H_i(t)}\right)^2 [1-F_i(t)]\, dG(t) \end{aligned} \tag{3.11}$$

$$V_{11}(\tau) = \textit{same as } V_{22} \textit{ except } b_{ni} b_{nj} \textit{ are now } a_{ni} \textit{ and } a_{nj} \tag{3.12}$$

$$\begin{aligned} V_{12}(\tau) &= V_{21}(\tau) = \lim n \sum a_{ni} b_{ni} \int_{-\infty}^{\tau} \left(\frac{t}{1-G(t)}\right. \\ &\quad \left. - \frac{\int_t^\tau s\, dF_i(s)}{1-H_i(t)}\right)^2 [1-G(t)]\, dF_i(t) \\ &\quad + \lim n \sum \int_{-\infty}^{\tau} \prod_{c_{ni} = a_{ni};\, b_{ni}} \left(\frac{\sum c_{nj}\int_t^\tau s\, dF_j(s)}{\sum [1-H_j(t)]}\right. \\ &\quad \left. - \frac{c_{ni}\int_t^\tau s\, dF_i(s)}{1-H_i(t)}\right) [1-F_i(t)]\, dG(t). \end{aligned} \tag{3.13}$$

Proof. We start with expression (3.4) and write the detailed proof only for $(\hat{\beta} - \beta)(T^n)$, since the proof for $(\hat{\alpha} - \alpha)(T^n)$ is very similar. To this end, we first derive a martingale representation for $(\hat{\beta} - \beta)$.

Adding and subtracting the following term to (3.4),

$$\sum b_{ni} \int_{-\infty}^{T^n} \frac{tI_{[Z_i \geq t]}}{1-\hat{G}(T^n \wedge t-)} \frac{dF_i(t)}{1-F_i(t)}$$

and regrouping terms, we have

$$\begin{aligned}(\hat{\beta}-\beta)(T^n) &= \sum b_{ni} \int_{-\infty}^{T^n} \frac{t}{1-\hat{G}(T^n \wedge t-)} \\ &\quad \times \left(dI_{[Z_i \leq t,\, \delta_i = 1]} - I_{[Z_i \geq t]} \frac{dF_i(t)}{1-F_i(t)} \right) \\ &\quad + \sum b_{ni} \int_{-\infty}^{T^n} \left(\frac{I_{[Z_i \geq t]}}{1-\hat{G}(T^n \wedge t-)} \cdot \frac{1}{1-F_i(t)} - 1 \right) t\, dF_i(t), \\ &= \sum b_{ni} \int_{-\infty}^{T^n} \frac{t\, dM_i^D(t)}{1-\hat{G}(T^n \wedge t-)} \\ &\quad + \sum b_{ni} \int_{-\infty}^{T^n} \left(\frac{1-\hat{H}_i(t)}{1-\hat{G}(T^n \wedge t-)} \cdot \frac{1}{1-F_i(t)} - 1 \right) t\, dF_i(t), \end{aligned} \tag{3.14}$$

by (2.1) and (2.2).

Now let us work on the second integrand of (3.14). The bracket can be rewritten as

$$\begin{aligned}&\left(\frac{1-\hat{H}_i(t)}{1-\hat{G}(T^n \wedge t-)} \cdot \frac{1}{1-F_i(t)} - 1 \right) \\ &\qquad = \frac{1-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} \cdot \frac{1-\hat{H}_i(t)}{[1-F_i(t)]\,[1-G(T^n \wedge t-)]} - 1. \end{aligned} \tag{3.15}$$

Since

$$\frac{1-\hat{H}_i(t)}{[1-F_i(t)]\,[1-G(T^n \wedge t-)]} = 1 + \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)},$$

and

$$\frac{1-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} = 1 + \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)},$$

it is not hard to see that after some algebraic manipulations,

$$
\begin{aligned}
(3.15) &= \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} + \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} \\
&\quad + \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} \cdot \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} \\
&= \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} + \frac{\hat{G}-G}{1-G} + \frac{H_i(t)-\hat{H}_i(t)}{1-H_i} \\
&\quad \cdot \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} \\
&\quad + \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)} \cdot \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} \\
&= \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} + \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)} + \xi_i(t) + \eta(t) \qquad (3.16)
\end{aligned}
$$

(say). Now plug this back into (3.14) to obtain

$$
\begin{aligned}
\hat{\beta} - \beta(T^n) &= \sum b_{ni} \int_{-\infty}^{T^n} \frac{t}{1-\hat{G}(T^n \wedge t-)} dM_i^D(t) \\
&\quad + \sum b_{ni} \int_{-\infty}^{T^n} \left(\frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} \right. \\
&\quad \left. + \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)} \right) t\, dF_i(t) \\
&\quad + \sum b_{ni} \int_{-\infty}^{T^n} (\xi_i(t) + \eta(t))\, t\, dF_i(t). \qquad (3.17)
\end{aligned}
$$

Integrating by parts on the middle term above, we have

$$
\begin{aligned}
(\hat{\beta} - \beta)(T^n) &= \sum b_{ni} \int_{-\infty}^{T^n} \frac{t}{1-\hat{G}(T^n \wedge t-)} dM_i^D(t) \\
&\quad + \sum b_{ni} \left(\int_{-\infty}^{T^n} \left[\int_t^{\infty} s\, dF_i(s) \right] d \frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)} \right. \\
&\quad \left. + \int_{-\infty}^{T^n} \left[\int_t^{\infty} s\, dF_i(s) \right] d \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)} \right) \\
&\quad + \sum b_{ni} \int_{-\infty}^{T^n} (\xi_i(t) + \eta(t))\, t\, dF_i(t) \\
&\quad - \sum b_{ni} h_i(T^n) \left[\frac{H_i(T^n)-\hat{H}_i(T^n)}{1-H_i(T^n)} + \frac{\hat{G}(T^n-)-G(T^n-)}{1-G_i(T^n-)} \right]
\end{aligned}
$$

where $h_i(t)=\int_t^\infty s\,dF_i(s)$. We also let

$$\zeta_{ni}(T^n)=\frac{H_i(T^n)-\hat{H}_i(T^n)}{1-H_i(T^n)}+\frac{\hat{G}(T^n-)-G(T^n-)}{1-G_i(T^n-)}.$$

Using the representations (2.8) and (2.9), we have

$$\begin{aligned}\hat{\beta}-\beta(T^n)&=\sum b_{ni}\int_{-\infty}^{T^n}\frac{t}{1-\hat{G}(T^n\wedge t-)}\,dM_i^D(t)\\&+\sum b_{ni}\int_{-\infty}^{T^n}h_i(t)\frac{(-1)}{1-H_i(t)}\,dM_i^+(t)\\&+\sum b_{ni}\int_{-\infty}^{T^n}h_i(t)\frac{1-\hat{G}(t-)}{1-G(t)}\frac{I_{[R^+(t)>0]}}{R^+(t)}\,dM_C^+(t)\\&+\sum b_{ni}\int_{-\infty}^{T^n}(\xi_i(t)+\eta(t))\,t\,dF_i(t)-\sum b_{ni}h_i(T^n)\,\zeta_{ni}(T^n).\end{aligned}$$

Recalling the definition of M_i^+ and M_C^+, we can rewrite this as

$$\begin{aligned}(\hat{\beta}-\beta)(T^n)&=\sum b_{ni}\int_{-\infty}^{T^n}\left(\frac{t}{1-\hat{G}(T^n\wedge t-)}-\frac{h_i(t)}{1-H_i(t)}\right)dM_i^D(t)\\&+\sum b_{ni}\int_{-\infty}^{T^n}\frac{-h_i(t)}{1-H_i(t)}\,dM_i^C(t)\\&+\int_{-\infty}^{T^n}\left[\sum_i b_{ni}h_i(t)\right]\frac{1-\hat{G}(t-)}{1-G(t)}\frac{I_{[R^+(t)>0]}}{R^+(t)}\,d\sum_j M_j^C(s)\\&+\int_{-\infty}^{T^n}(\xi_i(t)+\eta(t))\,t\,dF_i(t)-\sum b_{ni}h_i(T^n)\,\zeta_{ni}(T^n)\\&=\sum b_{ni}\int_{-\infty}^{T^n}\left(\frac{t}{1-\hat{G}(T^n\wedge t-)}-\frac{h_i(t)}{1-H_i(t)}\right)dM_i^D(t)\\&+\sum_{i=1}^{n}\int_{-\infty}^{T^n}\left(\left[\sum_j b_{nj}h_j(t)\right]\frac{1-\hat{G}(t-)}{1-G(t)}\frac{I_{[R^+(t)>0]}}{R^+(t)}\right.\\&\left.-\frac{b_{ni}h_i(t)}{1-H_i(t)}\right)dM_i^C(t)+\sum b_{ni}\int_{-\infty}^{T^n}(\xi_i(t)+\eta(t))\,t\,dF_i(t)\\&-\sum b_{ni}h_i(T^n)\,[\zeta_{ni}(T^n)].\end{aligned}\tag{3.18}$$

This is the required martingale representation for $(\hat{\beta}-\beta)(T^n)$. The first two terms above are clearly martingales. We show in Section 5 that $\sqrt{n}\sum b_{ni}\int_{-\infty}^{T^n}(\xi_i(t)+\eta(t))\,t\,dF_i(t)=o_p(1)$ and $\sqrt{n}\sum b_{ni}h_i(T^n)\,\zeta_{ni}(T^n)=$

$o_p(1)$. In the following discussions we focus on the martingale part M_n = first two terms in (3.18).

To show that the martingales converge weakly, we need to check that (i) their predictable variation converges in probability and (ii) the Lindberg condition holds (see, e.g., Anderson and Borgan, 1985). The predictable variation process for the martingale $\sqrt{n}\, M_n$ in (3.18) evaluated up to τ, after some calculation, is

$$n\langle M_n\rangle = n\sum b_{ni}^2 \int_{-\infty}^{\tau} \left(\frac{t}{1-\hat{G}(T'' \wedge t-)} - \frac{h_i(t)}{1-H_i(t)}\right)^2 I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)}$$
$$+ n\sum_{i=1}^{n} \int_{-\infty}^{\tau} \left(\left[\sum b_{nj} h_j(t)\right] \frac{1-\hat{G}(t-)}{1-G(t)} \frac{I_{[R^+(t)>0]}}{R^+(t)}\right.$$
$$\left. - \frac{b_{ni} h_i(t)}{1-H_i(t)}\right)^2 I_{[Z_i \geqslant t]} \frac{dG(t)}{1-G(t)}. \tag{3.19}$$

Appealing to Lemma 5.1 of Zhou (1992) and noticing that $1-\hat{G}$ is uniformly consistent on $[-\infty, \tau]$, it is not hard to see that the above two terms both converge in probability, and the limit is

$$V_\beta(\tau) = V_{22}(\tau)$$
$$= \lim n \sum b_{ni}^2 \int_{-\infty}^{\tau} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)}\right)^2 [1-G(t)]\, dF_i(t)$$
$$+ \lim n \sum \int_{-\infty}^{\tau} \left(\left[\sum b_{nj} h_j(t)\right] \frac{1}{[1-G(t)] \sum [1-F_j(t)]}\right.$$
$$\left. - \frac{b_{ni} h_i(t)}{1-H_i(t)}\right)^2 [1-F_i(t)]\, dG(t). \tag{3.20}$$

To check the Lindberg condition, we need to verify

$$\sum \int_{-\infty}^{\tau} \left(\frac{\sqrt{n}\, t b_{ni}}{1-\hat{G}(T'' \wedge t-)} - \frac{\sqrt{n}\, h_i(t)\, b_{ni}}{1-H_i(t)}\right)^2 I_{[|*|>\varepsilon]} I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)} \xrightarrow{P} 0, \tag{3.21}$$

and

$$\sum \int_{-\infty}^{\tau} \left(\left[\sum b_{nj} h_j(t)\right] \frac{1-\hat{G}(t-)}{1-G(t)} \frac{\sqrt{n}\, I_{[R^+(t)>0]}}{R^+(t)}\right.$$
$$\left. - \frac{\sqrt{n}\, b_{ni} h_i(t)}{1-H_i(t)}\right)^2 I_{[|*|>\varepsilon]} I_{[Z_i \geqslant t]} \frac{dG(t)}{1-G(t)} \xrightarrow{P} 0 \tag{3.22}$$

for any $\varepsilon > 0$. Here the $*$ stands for the quantity in the parentheses.

The square term inside (3.21) is bounded by

$$2nb_{ni}^2 \frac{t^2}{[1-\hat{G}(T^n \wedge t-)]^2} + 2nb_{ni}^2 \frac{[h_i(t)]^2}{[1-H_i(t)]^2}. \tag{3.23}$$

When t is restricted to $[-\tau, \tau]$, $t^2[1-\hat{G}(T^n \wedge t-)]^{-2}$ is bounded except for a set A (not depending on i) with probability less than any given δ. Therefore, aside from A, the first term above goes to zero uniformly for all $i \leqslant n$ since $\max_i(nb_{ni}^2) \to 0$ as $n \to \infty$ by our design. As for the second term, first observe

$$\frac{h_i(t)}{1-F_i(t)} = E(Y_i \mid Y_i > t) = \alpha + \beta X_i + E(\varepsilon_i \mid \varepsilon_i > t - \alpha - \beta X_i).$$

Since $E(\varepsilon \mid \varepsilon > u)$ is nonnegative and monotone in u, the above is

$$\leqslant \alpha + \beta X_i + E(\varepsilon_i \mid \varepsilon_i > |t| + |\alpha + \beta X_i|)$$
$$= \alpha + \beta X_i + |t| + |\alpha + \beta X_i| + E(\varepsilon_i - |t| - |\alpha + \beta X_i| \; |\varepsilon_i > |t| + \alpha + \beta X_i)$$

which in turn

$$\leqslant 2|\alpha + \beta X_i| + |t| + K$$

by Assumption 2; i.e., $\sup E[\varepsilon_i - t \mid \varepsilon_i > t] < \infty$. This implies, when t is restricted to $[-\tau, \tau]$,

$$\max_i 2nb_{ni}^2 \frac{[h_i(t)]^2}{[1-H_i(t)]^2} \to 0.$$

Therefore (3.21), when restricted to $[-\tau, \tau]$,

$$\max_i 2nb_{ni}^2 \int_{-\tau}^{\tau} \left(\frac{t}{1-\hat{G}(T^n \wedge t-)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 I_{[|*|>\varepsilon]} I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)},$$

will converge to zero in probability since, aside from the set A of small probability, every term in the sum will be zero when n is large because the indicator $I_{[|*|>\varepsilon]}$ will be zero.

To handle the lower tail of (3.21), we first replace the square term inside (3.21) by its bound (3.23) and treat them as two integrals. Except the set A, these integrals are bounded by

$$\sum \int_{-\infty}^{-\tau} 2Knb_{ni}^2 t^2 I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)}$$
$$+ \sum \int_{-\infty}^{-\tau} 2nb_{ni}^2 \frac{[h_i(t)]^2}{[1-H_i(t)]^2} I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)}.$$

Taking expectation, this is bounded by

$$\leqslant \sum K \int_{-\infty}^{-\tau} 2nb_{ni}^2 t^2 \, dF_i(t)$$

$$+ \sum K \int_{-\infty}^{-\tau} 2nb_{ni}^2 (4(\alpha + \beta X_i)^2 + 2t^2 + 2K^2) \, dF_i(t) < \infty.$$

Therefore, as $\tau \to \infty$ the above terms go to zero.

The proof of (3.22) is similar, so it is not repeated here (cf. also Zhou (1992, p. 11)). Therefore, the martingale part in (3.18), when evaluated at τ, is asymptotically normally distributed.

Arguments similar to those in Section 5 (cf. (5.5), (5.8)) show that, $\forall \varepsilon > 0$, we can choose $\tau > 0$ such that the probability $P(\sqrt{n}\, |M_n(T^n) - M_n(\tau)| > \varepsilon)$ can be made arbitrarily small when $n \to \infty$. Thus the weak convergence of $\sqrt{n}\, M_n(\tau)$ is still valid with τ replaced by T^n. Since $T^n \xrightarrow{\text{a.s.}} \infty$, our assumption A1 and the fact that a normal random variable is continuous in its variance imply that the limiting normal distribution for $\sqrt{n}\, M(T^n)$ will have variance $V_{22}(\infty)$. ∎

Extension to multiple regression is straightforward. The least squares estimate of the parameters for the multiple regression is given by

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}^*$$

where $\mathbf{Y}^* = (Y_1^*, ..., Y_n^*)^T$ with $Y_i^* = \delta_i Z_i / (1 - \hat{G}(Z_i))$, X is the design matrix, and we assume $(X^T X)^{-1}$ exists for large n.

It is clear that the estimate $\hat{\beta}_j$ is also a weighted average of Y_i^*. In fact, $\hat{\beta}_j = \sum_i u_{ji} Y_i^*$ where u_{ji} are the jth row, ith column element of the matrix $(X^T X)^{-1} X^T$. Hence the same technique used in proving Theorem 3.1 also works here.

THEOREM 3.2. *If, in the censored linear model with random design,*

1. *the variance–covariance matrix* $\Sigma(\tau)$ *(see ((3.24)) below) is well-defined and finite for* $\tau \in [k, \infty]$, *and* $\Sigma(\tau) \to \Sigma(\infty)$ *as* $t \to \infty$;
2. $\sup_{0<t} E[\varepsilon_i - t \mid \varepsilon_i > t] < \infty$; *and*
3. R1–R5 *hold with* $c_i = u_{ji}$, $j = 1, 2, ..., p$,

then we have

$$\sqrt{n}\, (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{D} N(0, \Sigma(\infty)) \qquad \text{as} \quad n \to \infty,$$

where $\boldsymbol{\beta}^* = \{\beta_1^*, ..., \beta_p^*\}$ *with* $\beta_j^* = \sum_i u_{ji} \int_{-\infty}^{T_n} s\, dF_i(s)$. *And* $\Sigma(\tau) = (\sigma_{kl}(\tau))$ *with*

$$\sigma_{kl}(\tau) = \lim n \sum_{i=1}^{n} u_{kl} u_{li} \int_{-\infty}^{\tau} \left[\frac{1}{1-G(t)} - \frac{\int_t^\tau s\, dF_i(s)}{1-H_i(t)}\right]^2 [1-G(t)]\, dF_i(t)$$
$$+ \lim n \sum_{i=1}^{n} \int_{-\infty}^{\tau} \prod_{c_i = u_{ki}, u_{li}} \left[\frac{\Sigma c_j \int_t^\tau s\, dF_j(s)}{\Sigma(1-H_j(t))}\right.$$
$$\left. - \frac{c_i \int_t^\tau s\, dF_i(s)}{1-H_i(t)}\right] (1-F_i)\, dG(t). \tag{3.24}$$

Remark. All the preceding results will be valid, subject to mild conditions, if the ε_i are independent but not identically distributed.

4. A Detailed Look at The Asymptotic Variance

The asymptotic variance we had obtained above, (3.11)–(3.13), can be rewritten in a different and more interesting form as shown below in (4.1). We use this new variance form in our subsequent discussions. By developing the big square in the second term of (3.11), and sum termwise, we find that the first and third terms combine to give

$$(3.11) = \lim n \sum b_{ni}^2 \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)}\right)^2 [1-H_i(t)] \frac{dF_i(t)}{1-F_i(t)}$$
$$\lim n \int_{-\infty}^{\infty} \sum \frac{b_{ni}^2 h_i^2(t)}{1-H_i(t)} \frac{dG(t)}{1-G(t)}$$
$$- \lim n \int_{-\infty}^{\infty} \frac{(\sum_j b_{nj} h_j(t))^2}{\sum_j [1-H_j(t)]} \frac{dG}{1-G}$$
$$= \lim n \sum b_{ni}^2 \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)}\right)^2$$
$$\times [1-H_i(t)] \frac{dF_i(t)}{1-F_i(t)}$$
$$+ \lim n \int_{-\infty}^{\infty} \sum b_{ni}^2 \frac{h_i^2(t)}{[1-H_i(t)]^2} [1-F_i(t)]\, dG$$
$$- \lim n \int_{-\infty}^{\infty} \frac{(\sum b_{nj} h_j(t))^2}{\sum_j [1-H_j(t)]} \frac{dG(t)}{1-G(t)}. \tag{4.1}$$

The first two positive terms above are nothing but the limit of $\mathrm{Var}(\sqrt{n}\sum b_{ni}(\delta_i Z_i/(1-G(Z_i)))$, as the following lemma shows.

Note that the third term above is negative! Also note that $\hat{\beta}=\sum b_{ni}(\delta_i Z_i/(1-\hat{G}(Z_i)))$, which means we are actually doing better by substituting $\hat{G}$ into $\sum b_{ni}(\delta_i Z_i/(1-G(Z_i)))$ in the sense that the resulting estimator has a smaller asymptotic variance. Also, this third term is a little different from Koul *et al.*'s (1981) formula (3.7), and in general it does not vanish. The interesting consequences of this fact are discussed after proving the lemma.

Remark. There is another way of writing the variances in (4.1) based on the following identity. Recalling the definition of $h_i(t)$ and integrating by parts we have

$$\frac{t}{1-G(t)}-\frac{h_i(t)}{1-H_i(t)}=\frac{t}{1-G(t)}-\frac{t(1-F_i(t))+\int_t^\infty (1-F_i)\,ds}{1-H_i(t)}$$

$$=-\frac{\int_t^\infty (1-F_i)\,ds}{1-H_i(t)}$$

Plug this identity into (4.1) and note the negative sign disappears after the square. Thus we have

$$(3.11)=\lim n\sum b_{ni}^2\int_{-\infty}^{\infty}\left(\frac{\int_t^\infty (1-F_i)\,ds}{1-H_i(t)}\right)^2[1-H_i(t)]\frac{dF_i(t)}{1-F_i(t)}+\cdots.$$

LEMMA 4.1.

$$\begin{aligned}\mathrm{Var}\left(\frac{\delta_i Z_i}{1-G(Z_i)}\right)&=\int_{-\infty}^{\infty}\left(\frac{t}{1-G(t)}-\frac{h_i(t)}{1-H_i(t)}\right)^2[1-G(t)]\,dF_i(t)\\&\quad+\int_{-\infty}^{\infty}\frac{h_i^2(t)}{[1-H_i(t)]^2}[1-F_i]\,dG(t)\\&=\int_{-\infty}^{\infty}\left(\frac{\int_t^\infty (1-F_i)\,ds}{1-H_i(t)}\right)^2[1-G(t)]\,dF_i(t)\\&\quad+\int_{-\infty}^{\infty}\frac{h_i^2(t)}{[1-H_i(t)]^2}[1-F_i]\,dG(t).\end{aligned}$$

Proof. Similar to the proof of Theorem 3.1, we first derive a martingale representation. Note first $E(\delta_i Z_i/(1-G(Z_i)))=\int_{-\infty}^{\infty} t\,dF_i(t)$ and therefore,

$$\mathrm{Var}\left(\frac{\delta_i Z_i}{1-G(Z_i)}\right)=E\left(\frac{\delta_i Z_i}{1-G(Z_i)}-\int_{-\infty}^{\infty} t\,dF_i(t)\right)^2. \qquad (4.2)$$

On the other hand, note that $\delta_i Z_i/(1-G(Z_i)) = \int_{-\infty}^{\infty} t/(1-G(t))\, dI_{[Z_i \leqslant t,\, \delta_i = 1]}$, therefore

$$\frac{\delta_i Z_i}{1-G(Z_i)} - \int_{-\infty}^{\infty} t\, dF_i(t)$$

$$= \int_{-\infty}^{\infty} \frac{t}{1-G(t)}\, dI_{[Z_i \leqslant t,\, \delta_i = 1]} - \int_{-\infty}^{\infty} t\, dF_i(t). \tag{4.3}$$

Adding and subtracting the term $\int_{-\infty}^{\infty} t/(1-G(t))\, I_{[Z_i \geqslant t]}\, dF_i(t)/(1-F_i(t))$ in the above and regrouping, we have

$$= \int_{-\infty}^{\infty} \frac{t}{1-G(t)} \left(dI_{[Z_i \leqslant t,\, \delta_i = 1]} - I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)} \right)$$

$$+ \int_{-\infty}^{\infty} \left(\frac{I_{[Z_i \geqslant t]}}{1-G(t)} \frac{1}{1-F_i(t)} - 1 \right) t\, dF_i(t)$$

$$= \int_{-\infty}^{\infty} \frac{t}{1-G(t)}\, dM_i^D(t) + \int_{-\infty}^{\infty} \frac{H_i(t) - \hat{H}_i(t)}{1-H_i(t)}\, t\, dF_i(t). \tag{4.4}$$

Integrating the second term above by parts and recalling the fact $h_i(t) = \int_t^{\infty} s\, dF_i(s)$, we get

$$= \int_{-\infty}^{\infty} \frac{t}{1-G(t)}\, dM_i^D(t) + \int_{-\infty}^{\infty} h_i(t)\, d\frac{H_i - \hat{H}_i(t)}{1-H_i(t)}.$$

Therefore, by (2.9), we have

$$(4.3) = \int_{-\infty}^{\infty} \frac{t}{1-G(t)}\, dM_i^D(t) - \int_{-\infty}^{\infty} h_i(t) \frac{dM_i^+(t)}{1-H_i(t)}$$

$$= \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)} \right) dM_i^D(t) - \int_{-\infty}^{\infty} h_i(t) \frac{dM_i^C(t)}{1-H_i(t)}.$$

It is not hard to see that the above integral is a martingale evaluated at $t = \infty$. The predictable variation process of the martingale is now easy to compute, and the expectation of the predictable variation process gives the desired variance

$$(4.2.) = E\left\langle \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)} \right) dM_i^D(t) \right.$$

$$\left. - \int_{-\infty}^{\infty} h_i(t) \frac{dM_i^C(t)}{1-H_i(t)} \right\rangle$$

$$= E\left\{ \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 I_{[Z_i \geqslant t]} \frac{dF_i(t)}{1-F_i(t)} \right.$$

$$\left. + \int_{-\infty}^{\infty} \frac{h_i^2(t)}{[1-H_i(t)]^2} I_{[Z_i \geqslant t]} \frac{dG}{1-G} \right\}$$

$$= \int_{-\infty}^{\infty} \left(\frac{t}{1-G(t)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 [1-G]\, dF_i(t)$$

$$+ \int_{-\infty}^{\infty} \frac{h_i^2(t)}{[1-H_i(t)]^2} [1-F_i]\, dG.$$

This completes the proof. ∎

To see that the negative term in the new variance form (4.2) is, in general, nonzero, we rewrite it as

$$\int \left(\sum b_{nj} h_j(t) \right)^2 \frac{1}{(1/n) \sum [1-H_i(t)]} \frac{dG}{1-G}$$

$$\geqslant \int \left(\sum b_{nj} h_j(t) \right)^2 \frac{dG}{1-G} \geqslant \int \left(\sum b_{nj} h_j(t) \right)^2 dG(t).$$

The integrand is a continuous nonnegative function of t, and for $t = -\infty$, it s $[\sum b_{nj} h_j(-\infty)]^2 = [\sum b_{nj}(\alpha + \beta X_j)]^2 = \beta^2 > 0$. Therefore, unless $G(t)$ is completely flat at places where $[\sum b_{nj} h(t)]^2 > 0$, the term is positive.

The variance estimator proposed by Koul *et al.* [1981, (4.8)] only estimates the first part of (4.1). In light of the discussion above, it needs to include an extra negative term. We suggest the use of

$$-n \int^{\infty} \frac{[\sum b_{nj} \hat{h}_j(t)]^2}{R^+(t) - 1} \frac{dN_c^+(t)}{R(t)}$$

as an estimator of the negative term where

$$N_c^+(t) = \sum I_{[Z_i \leqslant t,\, \delta_i = 0]} \qquad \text{and} \qquad \hat{h}_j(t) = \int_t^{\infty} s\, d\left(\frac{I_{[Z_j \geqslant s]}}{1 - \hat{G}(s)} \right).$$

The fact that the negative term in (4.2) does not vanish in general means that the asymptotic variance of $\hat{\beta}$ is smaller than $\lim n \operatorname{Var}(\sum b_{ni}(\delta_i T_i/(1-G(T_i))))$. The improvement comes from the fact that we are estimating G rather than using the true G (assumed known), a fact that is somewhat counterintuitive. To use a different estimator of G to result in an even larger negative term in (4.2) is indeed possible. The idea is to use a "local" estimator of G based only on part of the data. For more details, see Fygenson and Zhou (1994).

5. Higher Order Terms

In order to complete the proof of the main theorem (Theorem 3.1), we need to show that the higher order terms are $o_p(1)$. We now proceed to establish this.

Recall that these high-order terms are

$$\sqrt{n} \sum b_{ni} \int_{-\infty}^{T^n} (\xi_i(t) + \eta(t))\, t\, dF_i(t) \qquad \text{and} \qquad \sqrt{n} \sum b_{ni} h_i(T^n)[\zeta_n(T^n)]. \tag{5.1}$$

Let us first look at the integral term. Putting back what $\xi_i(t)$ and $\eta(t)$ stand for and writing them as two separate integrals, we have

$$\sqrt{n} \sum b_{ni} \int_{-\infty}^{T^n} \left(\frac{H_i(t) - \hat{H}_i(t)}{1 - H_i(t)} \cdot \frac{\hat{G}(T^n \wedge t-) - G(T^n \wedge t-)}{1 - G(T^n \wedge t-)} \right) t\, dF_i(t) \tag{5.2}$$

$$+ \sqrt{n} \sum b_{ni} \int_{-\infty}^{T^n} \frac{[\hat{G}(T^n \wedge t-) - G(T^n \wedge t-)]^2}{[1 - G(T^n \wedge t-)]\,[1 - \hat{G}(T^n \wedge t-)]}\, t\, dF_i(t). \tag{5.3}$$

It is easier to show that $(5.3) = o_p(1)$, so we first prove this. We accomplish this by breaking the integral into $\int_{-\infty}^{\tau} + \int_{\tau}^{\infty}$. In the interval $[-\infty, \tau]$, we have

$$\sup_{t \leqslant \tau} \left| \frac{\hat{G}(T^n \wedge t-) - G(T^n \wedge t-)}{1 - \hat{G}(T^n \wedge t-)} \right| = o_p(1)$$

and

$$\sqrt{n} \left| \frac{\hat{G}(T^n \wedge t-) - G(T^n \wedge t-)}{1 - G(T^n \wedge t-)} \right| \xrightarrow{\mathscr{D}} |B(C(t))| \qquad \text{in space } D[-\infty, \tau],$$

where $B(\cdot)$ is a standard brownian motion. The proof of the first claim is a simple application of the Lenglart inequality and the latter due to the martingale central limit theorem.

Theorem,

$$\int_{-\infty}^{\tau} \sqrt{n}\,\frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t)}\,\frac{\hat{G}-G}{1-\hat{G}}\,t d\left(\sum b_{ni}F_i(t)\right)$$

$$\leqslant o_p(1)\int_{-\infty}^{\tau}\left|\sqrt{n}\,\frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)}\right| |t| \sum |b_{ni}|\, dF_i$$

$$\leqslant o_p(1)\times O_p(1)\times \int_{-\infty}^{\tau} \sum |b_{ni}|\, |t|\, dF_i$$

$$\leqslant o_p(1)\sum |b_{ni}|\, E|Y_i|, \tag{5.4}$$

and the sum in the last expression above is uniformly bounded by our assumption of a random design,

$$\sum |b_{ni}|\, E|Y_i| \leqslant \frac{1}{\sqrt{\sum (X_i-\bar{X})^2}}\cdot \sqrt{\sum \operatorname{Var} Y_i + (EY_i)^2}$$

$$= \sqrt{\frac{n\sigma^2}{\sum (X_i-\bar{X})^2}+\frac{n\alpha^2}{\sum (X_i-\bar{X})^2}+\frac{\beta^2 \sum X_i^2}{\sum (X_i-\bar{X})^2}+\frac{2\alpha\beta \sum X_i}{\sum (X_i-\bar{X})^2}},$$

which is easily seen to be bounded in view of the design condition D1. Therefore $(5.4)=o_p(1)$ for any finite τ.

The upper tail of the integral, $\int_{\tau}^{T^n}$, has to be treated differently since $\sqrt{n}\,(\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)/1-G(T^n \wedge t-))$ is no longer $O_p(1)$ there, and $|(\hat{G}-G)/(1-\hat{G})|$ is no longer $o_p(1)$ though it remains $O_p(1)$ (Zhou, 1991). This calls for

$$\int_{\tau}^{T^n}\left|\sqrt{n}\,\frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-G(T^n \wedge t-)}\, t \sum b_{ni} f_i(t)\right| dt$$

$$\to \int_{\tau}^{\infty} |B(C(t))\,\bar{h}(t)|\, dt, \tag{5.5}$$

which follows from R1 if we use an extension of Gill's (1983) Theorem 2.1 (see Zhou, 1986). Finally, the tail is negligible for large τ by Assumption R2.

We now turn to (5.2). First treat the integral

$$\int_{-\tau}^{\tau} \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)}\sqrt{n}\sum b_{ni}\,\frac{H_i(t)-\hat{H}_i(t)}{1-H_i(t)}\, tf_i(t)\, dt \tag{5.6}$$

for arbitrary but fixed $\tau > 0$. It is easy to show

$$\sup_{|t|<\tau} \sqrt{n}\,\frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} = O_p(1)$$

and

$$\sup_{|t|<\tau} \sum b_{ni}[H_i(t)-\hat{H}_i(t)]\,\frac{tf_i(t)}{1-H_i(t)} = o_p(1)$$

provided

$$\sup_{|t|\leqslant\tau} \sum b_{ni}^2\,\frac{t^2 f_i^2(t)}{[1-H_i(t)]^2} \to 0,$$

which follows from D1 and R3. It then is an easy exercise to show that $(5.3) = o_p(1)$ by using Lemma 5.1 of Zhou (1992).

The lower tail can be handled separately. Note that

$$\begin{aligned}
&\int_{-\infty}^{-\tau} \sum |b_{ni}| \left|\frac{H_i(t)-\hat{H}_i(t)}{1-H_i}\right| |t|\, dF_i(t)\\
&\quad\leqslant \int_{-\infty}^{-\tau} \sum |b_{ni}|\,\frac{1}{1-H_i(-\tau)}\,|t|\, dF_i(t)\\
&\quad= \sum |b_{ni}|\,\frac{1}{1-H_i(-\tau)} \int_{-\infty}^{-\tau} |t|\, dF_i(t),
\end{aligned}$$

which is bounded as shown before. It implies that we can make the integral $\int_{-\infty}^{-\tau}$ small by choosing a large τ.

Now let us look at the upper tail of (5.2). Since by Zhou (1991)

$$\sup_{t\leqslant T^n} \frac{\hat{G}(T^n \wedge t-)-G(T^n \wedge t-)}{1-\hat{G}(T^n \wedge t-)} = O_p(1), \tag{5.7}$$

we have, for any $\tau > 0$,

$$\begin{aligned}
&\left|\int_{\tau}^{T^n} \sqrt{n} \sum b_{ni}\left(\frac{H_i-\hat{H}_i}{1-H_i}\right)\frac{\hat{G}-G}{1-\hat{G}}\, t\, dF_i(t)\right|\\
&\quad\leqslant O_p(1)\cdot \int_{\tau}^{T^n} \left|\sqrt{n}\sum b_{ni}\,\frac{H_i-\hat{H}_i}{1-H_i}\, t\, f_i(t)\right| dt, \qquad (5.8)
\end{aligned}$$

and

$$P\left(\int_\tau^{T^n}\left|\sqrt{n}\sum b_{ni}\frac{H_i-\hat{H}_i}{1-H_i}tf_i(t)\right|dt>\varepsilon\right)$$
$$\leqslant\frac{1}{\varepsilon}E\int_\tau^\infty\left|\sqrt{n}\sum b_{ni}\frac{H_i-\hat{H}_i}{1-H_i}tf_i(t)\right|dt$$
$$\leqslant\frac{1}{\varepsilon}\int_\tau^\infty\sqrt{E\left(\sqrt{n}\sum b_{ni}\frac{H_i-\hat{H}_i}{1-H_i}tf_i(t)\right)^2}\,dt$$
$$=\frac{1}{\varepsilon}\int_\tau^\infty\sqrt{nt^2\sum b_{ni}^2f_i^2(t)\frac{H_i(1-H_i)}{[1-H_i]^2}}\,dt$$
$$\leqslant\frac{1}{\varepsilon}\int_\tau^\infty\sqrt{n\sum b_{ni}^2\frac{t^2f_i^2(t)}{1-H_i}}\,dt. \tag{5.9}$$

Condition R5 implies that (5.9) can be made arbitrarily small by choosing a large τ.

We now turn to the last term in (5.1). This term, with $\zeta_n(t)$ plugged back in, is

$$\sqrt{n}\sum b_{ni}h_i(T^n)\frac{H_i(T^n)-\hat{H}_i(T^n)}{1-H_i(T^n)}+\sqrt{n}\sum b_{ni}h_i(T^n)\frac{\hat{G}(T^n)-G(T^n)}{1-G(T^n)}. \tag{5.10}$$

The second term in (5.10) will be $o_p(1)$ because of Assumption R2 and Gill's (1983) Theorem 2.1. As for the first term in (5.10), since $T^n\to\infty$ a.s., it is $o_p(1)$ by Assumption R3. This completes the treatment of the tails.

Acknowledgments

We thank the editor and the referee for valuable suggestions which improved the presentation of the paper.

References

[1] Andersen, P. K., and Borgan, O. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12** 97–158.

[2] Buckley, and James, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

[3] Chow, Y. S., and Teicher, H. (1978). *Probability Theory*, Springer, New York.

[4] Fygenson, M., and Zhou, M. (1994). On using stratification in the analysis of linear regression models with right censoring, *Ann. Statist.*, to appear.

[5] FYGENSON, M., AND ZHOU, M. (1992). Modification of the Koul-Susarla-Van Ryzin estimator for linear regression models with right censoring. *Statist. Probab. Lett.* **13** 295–299.

[6] GILL, R. (1980). Censoring and stochastic integrals. In *Mathematical Centre Tracts*, Vol. 124. Mathematisch Centrum, Amsterdam.

[7] GILL, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

[8] KOUL, H., SUSARLA, V. AND VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288.

[9]. LAI, T. L., AND YING, Z. (1991). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann. Statist.* **19** 1370–1402.

[10] MILLER, R. G., AND HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.

[11] STRUTHERS, C. A., AND KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369.

[12] ZHOU, M. (1986). *Some Nonparametric Two Sample Tests with Randomly Censored Data*. Ph.D. thesis, Columbia University.

[13] ZHOU, M. (1991). Some properties of the Kaplan-Meier estimator for independent, nonidentically distributed random variables. *Ann. Statist.* **19** 2266–2274.

[14] ZHOU, M. (1992). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20** 1002–1021.