The Annals of Statistics
1997, Vol. 25, No. 2, 662–682

TOWARDS A GENERAL ASYMPTOTIC THEORY FOR COX MODEL WITH STAGGERED ENTRY

BY YANNIS BILIAS, MINGGAO GU[1] AND ZHILIANG YING[2]

Iowa State University, McGill University and Rutgers University

A general asymptotic theory is established for the two-parameter Cox score process with staggered entry data. It extends in several directions the existing theory developed by Sellke and Siegmund, Slud and Gu and Lai. An essential tool employed here is a modern empirical process theory, as elucidated in a recent monograph by Pollard.

1. Introduction. Cox's (1972) proportional hazards regression model has been widely accepted in clinical trials and other follow-up studies. It relates the hazard rate of a failure time T to a possibly time-dependent $p \times 1$ covariate vector process Z through

$$\lambda(s \mid Z(u), u \le s) = \exp\{\beta' Z(s)\}\lambda_0(s), \tag{1.1}$$

where λ_0 is the (unspecified) baseline hazard function and β the unknown regression parameter vector of primary interest. When all the follow-ups start at the same time, say 0, the partial likelihood score computed at time t becomes a martingale with respect to a properly defined filtration and an elegant martingale-based asymptotic theory has been developed [cf. Andersen and Gill (1982)].

In many practical situations, however, subjects under study may be recruited at different times, from which the follow-ups begin. It is known that under these circumstances, the score process is no longer a martingale, thus theoretical treatment of it becomes much harder. In particular, the standard martingale central limit theorem of Rebolledo cannot be applied, at least directly, to obtain weak convergence of the score process. See Sellke and Siegmund (1983) for a fundamental breakthrough and Slud (1984) and Gu and Lai (1991) for some related work.

The design, model and corresponding statistic we shall consider herein may be described as follows. There are potentially infinitely many individuals, with whom are attached entry (recruiting) times $\tau_i(\ge 0)$, failure times $T_i(\ge 0)$, censoring times $C_i(\ge 0)$ which could take value ∞ and $p \times 1$ covariate vector processes Z_i. Suppose that (τ_i, T_i, C_i, Z_i') are independent random vectors and that the conditional hazard rate of T_i at s, given τ_i, C_i, and

Received January 1994; revised May 1996.

[1]Research supported by National Sciences and Engineering Research Council of Canada.

[2]Research supported by NSF and National Security Agency.

AMS 1991 *subject classifications*. Primary 62l10; secondary 62E20, 62J05.

Key words and phrases. Proportional hazards regression, partial likelihood score, staggered entry, empirical processes, functional central limit theorem, pseudodimension, manageability, and random fields.

$Z_i(u)$, $u \le s$, is $\exp\{\beta' Z_i(s)\}\lambda_0(s)$. The τ_i need not be ordered. Furthermore, suppose only n of them, $i = 1, \dots, n$ are being sampled. Thus at t, the current calendar time, the ith individual's failure time T_i is censored by $C_i \wedge (t - \tau_i)^+$. Throughout the sequel, $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, $a^+ = \max\{0, a\}$ and $a^- = \max\{0, -a\}$. Let $\tilde{T}_i(t) = T_i \wedge C_i \wedge (t - \tau_i)^+$, $\Delta_i(t) = I_{(T_i \le C_i \wedge (t - \tau_i)^+)}$ and $R_i(t) = \{j: 1 \le j \le n, \tilde{T}_j(t) \ge \tilde{T}_i(t)\}$. With the above notation, the Cox (1975) partial likelihood at calendar time t can then be written as

$$L_t(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\left(\beta' Z_i\left(\tilde{T}_i(t)\right)\right)}{\sum_{j \in R_i(t)} \exp\left(\beta' Z_j\left(\tilde{T}_i(t)\right)\right)} \right]^{\Delta_i(t)}. \tag{1.2}$$

In terms of the usual counting process representation for the score, it is necessary to consider simultaneously two types of times, the calendar time t and the survival time s. Unless otherwise stated, it will be assumed throughout the paper that $t \ge s$. Let

$$N_i(t, s) = \Delta_i(t) I_{(\tilde{T}_i(t) \le s)} \left(= I_{(T_i \le C_i \wedge (t - \tau_i)^+ \wedge s)}\right), \qquad Y_i(t, s) = I_{(\tilde{T}_i(t) \ge s)},$$

$$\bar{Z}(\beta; t, s) = \sum_{i=1}^{n} Z_i(s) \exp(\beta' Z_i(s)) Y_i(t, s) \Big/ \sum_{i=1}^{n} \exp(\beta' Z_i(s)) Y_i(t, s).$$

Define a two-parameter score process

$$U(\beta; t, s) = \sum_{i=1}^{n} \int_0^s \left[Z_i(u) - \bar{Z}(\beta; t, u) \right] N_i(t, du). \tag{1.3}$$

It is easy to see that $U(\beta; t, t)$ is the partial likelihood score $(\partial / \partial \beta) \log L_t(\beta)$.

Earlier, Sellke and Siegmund (1983) showed that, under certain regularity conditions, the diagonal process $U(\beta; t, t)$ is approximately a martingale and therefore converges weakly (with appropriate normalization) to a Brownian motion. A somewhat similar result, also for the diagonal process, can be found in Slud (1984), who considered only the two-sample problem and showed that weighted log-rank score processes can be approximated by a time-rescaled Brownian motion provided the weight functions are independent of the calendar time. Weak convergence of the two-parameter process was recently derived by Gu and Lai (1991), but only for the two-sample model. We refer to Andersen, Borgan, Gill and Keiding (1993) for a summary of their results.

Our main objective here is to derive functional central limit theorems for the basic two-parameter process $n^{-1/2} U(\beta; \cdot, \cdot)$ under the null hypothesis and the contiguous alternatives and for the corresponding maximum partial likelihood estimator. Our approach is entirely different from those of Sellke and Siegmund (1983) and Slud (1984) in that it basically ignores the martingale structure and relies, instead, on a modern empirical process theory, as elucidated in Pollard (1990). It will become clear in subsequent developments that the structure of U is ideally suitable for exploiting the powerful tools of this empirical process theory. Consequently, we are able to deal with a model far more general than that studied in Gu and Lai (1991), yet avoiding most of their heavy technicalities.

The paper is structured in a natural order. Sections 2 and 3 provide functional central limit theorems for U under the null hypothesis and the contiguous alternatives. The corresponding maximum partial likelihood estimator is treated in Section 4, where convergence for the cumulative baseline hazard estimator is also established. An application to sequential tests with covariate adjustment along with some discussions are given in Section 5. Some technical lemmas are put together and proved in the Appendix.

2. Convergence of the two-parameter score process. The main effort of this section is to show convergence of U to a Gaussian random field. Clearly, $U(\beta; t, s) = U(\beta; t, t)$ for $s \geq t$. So we only need to consider those (t, s) for which $s \leq t$. Furthermore, stability consideration of $\bar{Z}$ leads us to restrict t to $[0, t^*]$ with t^* satisfying

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} EY_i(t^*, t^*) > 0, \tag{2.1}$$

and λ_0 being bounded on $[0, t^*]$. Existence of such t^* is tantamount to requiring that there is a positive proportion of individuals whose entry times are 0, a rather restrictive assumption especially in designs of clinical trials and other prospective studies. It is not essential to require (2.1) for the subsequent results to hold, however. But without it, we would have to spend a substantial portion in our technical development controlling the so-called tail instability, as in Lai and Ying (1988), thereby distracting our attention from the main theme. Throughout the rest of the paper, D_* denotes $\{(t, s): 0 \leq s \leq t \leq t^*\}$ and, except in Section 3, β_0 denotes the true regression parameter; that is, $U(\beta_0; t, s)$ is under the null hypothesis.

The following regularity conditions will be needed as we proceed.

CONDITION 1. There exists a nonrandom constant B such that the total variation $|Z_i(0)| + \int_0^{t^*} |dZ_i(s)| \leq B$, where the first $|\cdot|$ denotes the L_1-norm for a p-dimensional vector and the second one the L_1-type total variation for a p-vector function.

CONDITION 2. For each $k = 0, 1$ and 2, there exists $\Gamma_k(t, s)$ such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E\left[Z_i^{\otimes k}(s) Y_i(t, s) \exp(\beta_0' Z_i(s))\right] = \Gamma_k(t, s) \quad \text{for all } (t, s) \in D_*,$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa'$ for a column vector a.

CONDITION 3. Let $K(t, s) = \Gamma_1(t, s)/\Gamma_0(t, s)$ and

$$K_n(t, s) = \sum_{i=1}^{n} E\left[Z_i(s) Y_i(t, s) \exp(\beta_0' Z_i(s))\right] \Big/ \sum_{i=1}^{n} E\left[Y_i(t, s) \exp(\beta_0' Z_i(s))\right].$$

Then, for each fixed s, $K(\cdot, s)$ is continuous on $[s, t^*]$ and

$$\sup_{0\le t\le t^*}\int_0^t \left[K_n(t,s) - K(t,s)\right]^2\, ds \to 0.$$

REMARK 1. Conditions 2 and 3 are analogous to conditions B and D in Andersen and Gill (1982), which provided the first martingale-based derivation of the asymptotic theory for the Cox model. Condition 1 assumes bounded variation for the covariate processes, a key assumption in applying the empirical process theory. Constancy on B is not essential, however, but simplifies our proofs; we believe that B could be random and depend on i, but needs to have finite moment generating functions.

REMARK 2. The work of Sellke and Siegmund (1983) requires that conditional on τ_i, random vectors (Z_i', C_i, T_i) are i.i.d. In this case, it is easy to see that $K_n(t, s)$ depends neither on n nor on t. So condition 3 is automatically satisfied.

Let d be the supremum norm for the space of bounded functions on $D_*: d(x, y) = \sup_{(t,s)\in D_*}|x(t,s) - y(t,s)|$. The functions could be p-dimensional vectors, for which $|\cdot|$ is in the sense of L_1, which is equivalent to the L_2-norm since the space R^p is finite dimensional. Let $B(D_*)$ be the space of bounded functions on D_* equipped with metric d. Following Pollard [(1990), Definition 9.1], a sequence of processes $\{X_n\}$ on D_* is said to converge in distribution to X if $\lim Ef(X_n) = Ef(X)$ for all f that are bounded and uniformly continuous on $B(D_*)$.

We first show that two basic processes converge in distribution to Gaussian random fields. Let

$$M_i(\beta; t, s) = N_i(t; s) - \textstyle\int_0^s Y_i(t,u)\exp(\beta' Z_i(u))\lambda_0(u)\, du.$$

Define $U_1(\beta; t, s) = \sum_{i=1}^n M_i(\beta; t, s)$ and

$$U_2(\beta; t, s) = \textstyle\sum_{i=1}^n \int_0^s Z_i(u) M_i(\beta; t, du).$$

Note that $U(\beta; t, s) = U_2(\beta; t, s) - \int_0^s \overline{Z}(\beta; t, u) U_1(\beta; t, du)$. When $\beta = \beta_0$, we shall omit β in M_i, U_1, U_2 and U, that is, $M_i(t,s) = M_i(\beta_o; t, s)$, $U_1(t,s) = U_i(\beta_0; t, s)$ and so on.

THEOREM 2.1. *Suppose Conditions* 1 *and* 2 *are satisfied. Then* $\{n^{-1/2}U_1(t,s), (t,s)\in D_*\}$ *and* $\{n^{-1/2}U_2(t,s), (t,s)\in D_*\}$ *converge in distribution to two Gaussian processes*, ξ_1 *and* ξ_2, *respectively, that have continuous sample paths, mean* 0 *and covariance functions specified by*

$$E\left[\xi_1(t_1, s_1)\xi_1'(t_2, s_2)\right] = \int_0^{s_1\wedge s_2} \Gamma_0(t_1\wedge t_2, u)\lambda_0(u)\, du,$$

$$E\left[\xi_2(t_1, s_2)\xi_2'(t_2, s_2)\right] = \int_0^{s_1\wedge s_2} \Gamma_2(t_1\wedge t_2, u)\lambda_0(u)\, du.$$

PROOF. Since U_1 is a special case of U_2 when the Z_i are replaced by 1, we shall only prove the convergence for $n^{-1/2}U_2$. Furthermore, in view of Condition 2, convergence of finite-dimensional distributions of $n^{-1/2}U_2$ to those of ξ_2 is clearly true by the classical multivariate central limit theorem for independent random vectors. So the issue becomes proving the so-called tightness, which can be done by dealing with each of the p components of $n^{-1/2}U_2$ separately. Thus, with no loss of generality, we may assume $p = 1$.

We shall invoke Theorem 10.7 (functional central limit theorem) of Pollard (1990) to get the desired convergence. Thus conditions (i)–(v) thereof need to be verified. Condition (ii) follows from Condition 2, the stability assumption, whereas (iii) and (iv) hold because, in view of Condition 1, envelopes can be chosen to be $B^*/\sqrt{n}$ for some constant B^*.

To verify (v), define for any $(t_k, s_k) \in D_*$, $k = 1, 2$, $\rho_n((t_1, s_1), (t_2, s_2)) = E[n^{-1/2}U_2(t_1, s_2) - n^{-1/2}U_2(t_2, s_2)]^2$ and $\rho((t_1, s_1), (t_2, s_2)) = E[\xi_2(t_1, s_1) - \xi_2(t_2, s_2)]^2$. Now

$$\begin{aligned}
\rho_n&((t_1, s_1), (t_2, s_2)) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\bigg\{\int_{s_1\wedge(t_1-\tau_i)^+\wedge C_i}^{s_2\wedge(t_2-\tau_i)^+\wedge C_i} Z_i(u) \\
&\qquad\qquad \times\Big[dI_{(T_i\le u)} - I_{(T_i\ge u)}\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\Big]\bigg\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left|\int_{s_1\wedge(t_1-\tau_i)^+}^{s_2\wedge(t_2-\tau_i)^+} Z_i^2(u) I_{(T_i\wedge C_i\ge u)}\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\right|.
\end{aligned}$$

Clearly $\{\rho_n\}$ is equicontinuous on $D_* \times D_*$. It is not difficult to verify that ρ_n also converges pointwise to ρ, a pseudometric on D_*. Thus ρ_n converges, uniformly on D_*, to ρ. Let $\{(t_1^{(n)}, s_1^{(n)})\}$ and $\{(t_2^{(n)}, s_2^{(n)})\}$ be any two sequences in D_*. It follows that if

$$\rho\big((s_1^{(n)}, t_1^{(n)}), (s_2^{(n)}, t_2^{(n)})\big) \to 0,$$

then $\rho_n((s_1^{(n)}, t_1^{(n)}), (s_2^{(n)}, t_2^{(n)}))$ converges to 0. Thus (v) holds.

It remains to verify (i). To do this, it suffices to show, in view of Lemma A.1, that both $\{\int_0^s Z_i(u)N_i(t, du)\}$ and $\{\int_0^s Z_i(u)Y_i(t, u)\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\}$ are manageable. Since $Z_i(u) = Z_i^+(u) - Z_i^-(u)$, we may furthermore assume $Z_i \ge 0$. Now

$$\begin{aligned}
\int_0^s Z_i(u)N_i(t, du) &= Z_i(T_i) I_{(T_i\le s\wedge C_i)} I_{(T_i\le(t-\tau_i)^+)} \\
&= \min\big\{Z_i(T_i) I_{(T_i\le s\wedge C_i)}, Z_i(T_i) I_{(T_i\le(t-\tau_i)^+)}\big\}.
\end{aligned} \tag{2.2}$$

For each i, $Z_i(T_i)I_{(T_i\le s\wedge C_i)}$ is nondecreasing in s and $Z_i(T_i)I_{(T_i\le(t-\tau_i)^+)}$ is nondecreasing in t. Thus both have pseudodimension at most 1. By Lemma 5.1 of Pollard (1990) and (2.2), $\{\int_0^s Z_i(u)N_i(t, du)\}$ has pseudodimension at most 10, and therefore must be Euclidean and certainly manageable. Manageability of $\{\int_0^s Z_i(u)Y_i(t,u)\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\}$ follows from the same argument since

$$\begin{aligned}\int_0^s & Z_i(u)Y_i(t,u)\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\\ &= \min\bigg\{\int_0^{s\wedge T_i\wedge C_i} Z_i(u)\exp(\beta_0 Z_i(u))\lambda_0(u)\,du,\\ &\qquad\qquad \int_0^{(t-\tau_i)^+\wedge T_i\wedge C_i} Z_i(u)\exp(\beta_0 Z_i(u))\lambda_0(u)\,du\bigg\}.\end{aligned}$$

Therefore, Theorem 10.7 of Pollard (1990) implies that $n^{-1/2}U_2$ converges in distribution to a Gaussian random field on D_* having continuous sample path with respect to pseudometric ρ. Since ρ is dominated by the Euclidean metric on D_*, the limiting Gaussian random field must also be continuous with respect to the Euclidean metric. □

We now state the main result of this section.

THEOREM 2.2. *Under Conditions* 1–3, $n^{-1/2}U$ *converges in distribution to a vector-valued Gaussian random field* ξ *on* D_* *with continuous sample path, mean* 0 *and covariance function*

$$\begin{aligned}E&[\xi(t_1,s_1)\xi'(t_2,s_2)]\\ &= \int_0^{s_1\wedge s_2}\big[\Gamma_2(t_1\wedge t_2,u) - \Gamma_1^{\otimes 2}(t_1\wedge t_2,u)/\Gamma_0(t_1\wedge t_2,u)\big]\lambda_0(u)\,du.\end{aligned}$$

REMARK 3. It is clear from its covariance function that ξ has independent increments in both s and t directions. Thus the diagonal process $\xi(t,t)$ is a time-rescaled Brownian motion when $\dim(Z_i)=1$, and a vector-valued Gaussian process with independent increments when $\dim(Z_i)>1$.

REMARK 4. Theorem 2.2 generalizes the result of Sellke and Siegmund (1983) in two ways: (i) convergence occurs on D_* instead of on the diagonal line $\{t=s\}$; (ii) it does not require that the entry times be independent of (Z_i', T_i, C_i). It also generalizes the result of Gu and Lai (1991), in which Z_i is either 0 or 1.

REMARK 5. Suppose that $w_n(t,s)$ is a sequence of weight functions on D_* such that for each t, they have uniformly bounded variations in s and that, uniformly on D_*, it converges in probability to a deterministic function $w(t,s)$. Then it follows from Theorem 2.2 and Lemma A.3 that the weighted log-rank process $n^{-1/2}\int_0^t w_n(t,u)U(t,du)$ converges in distribution to a Gaussian process. In the case of testing $\beta_0=0$, one may choose $w_n(t,s)=\sum Y_i(t,s)/n$, giving rise to the modified Gehan statistic of Slud and Wei (1982); one may also choose $w_n(t,s)$ to be the Kaplan–Meier estimator of the survival distribution from observations available at time t to get the sequential version of the Tarone–Ware statistic. For the latter, the limiting weight function w does not depend on t and therefore the limiting Gaussian process has independent increments. Formally, we have the following corollary.

COROLLARY 2.1. *Let w_n be a sequence of functions, converging uniformly on D_* to w. For each fixed t, $w_n(t,\cdot)$ and $w(t,\cdot)$ are left continuous and their total variations are bounded by a constant, independent of t and n. Define $U_w(t,s)=\int_0^s w_n(t,u)U(t,du)$. Then under Conditions* 1–3, *$n^{-1/2}U_w(t,s)$ converges in distribution to a Gaussian random field ξ_w with mean* 0 *and covariance function*

$$E[\xi_w(t_1,s_1)\xi_w'(t_2,s_2)] = \int_0^{s_1\wedge s_2} w(t_1,u)w(t_2,u) \times\left[\Gamma_2(t_1\wedge t_2,u) - \frac{\Gamma_1^{\otimes 2}(t_1\wedge t_2,u)}{\Gamma_0(t_1\wedge t_2,u)}\right]\lambda_0(u)\,du.$$

REMARK 6. Gill and Schumacher (1987) initiated use of different weighted log-rank score processes to obtain a simple goodness-of-fit test in the two-sample case. Their idea was based on an observation that if lack of fit is present, then there is a weight function to make the score process have a positive drift. See Lin (1991) for an extension to multiple regression. Similar tests can be constructed for the staggered entry model and their theoretical foundation is provided by Corollary 2.1.

PROOF OF THEOREM 2.2. We first show that with probability 1,

$$\sup_{(t,s)\in D_*}\left|\frac{1}{n}\sum_{i=1}^n Y_i(t,s)Z_i(s)\exp(\beta_0'Z_i(s)) - \frac{1}{n}\sum_{i=1}^n E[Y_i(t,s)Z_i(s)\exp(\beta_0'Z_i(s))]\right|\to 0, \tag{2.3}$$

$$\sup_{(t,s)\in D_*}\left|\frac{1}{n}\sum_{i=1}^n Y_i(t,s)\exp(\beta_0'Z_i(s)) - \frac{1}{n}\sum_{i=1}^n E[Y_i(t,s)\exp(\beta_0'Z_i(s))]\right|\to 0. \tag{2.4}$$

For (2.4), it suffices to show, in view of Theorem 8.3 of Pollard (1990), that $\{Y_i(t,s)\exp(\beta_0'Z_i(s))\}$ is manageable. From Condition 1, total variations of $\exp(\beta_0'Z_i(s))$ are bounded by some constant, implying that we can write $\exp(\beta_0'Z_i(s)) = \tilde{Z}_i^+(s) - \tilde{Z}_i^-(s)$, where both $\tilde{Z}_i^+$ and $\tilde{Z}_i^-$ are nonnegative, nonincreasing and bounded by some constant $\tilde{B}$. Write

$$Y_i(t,s)\exp(\beta_0'Z_i(s)) = I_{(T_i\wedge C_i\ge s)}\tilde{Z}_i^+(s)I_{((t-\tau_i)^+\ge s)} - I_{(T_i\wedge C_i\ge s)}\tilde{Z}_i^-(s)I_{((t-\tau_i)^+\ge s)}. \tag{2.5}$$

We now argue that the first term on the right-hand side of (2.5) is manageable. In view of (5.2) of Pollard (1990), it suffices to show that both $\{I_{(T_i \wedge C_i \geq s)} \tilde{Z}_i^+(s)\}$ and $\{I_{((t-\tau_i)+\geq s)}\}$ have finite pseudodimensions. We know from Lemma A.2 that the former has pseudodimension 1. The latter also has pseudodimension 1 because for any i, j, the set $\{(I_{((t-\tau_i)^+ \geq s)}, I_{((t-\tau_j)^+ \geq s)})$: $(t, s) \in D_*\}$ contains at most three points. Likewise, the second term on the right-hand side of (2.5) is also manageable. Hence $\{Y_i(t, s)\exp(\beta_0' Z_i(s))\}$ is manageable. The same argument can be used to prove (2.3). Note that, componentwise, $Z_i(s)\exp(\beta_0' Z_i(s))$ are of bounded variation with a bound independent of i.

Let $K_n(t, s)$ be defined as in Condition 3. From (2.3) and (2.4),

$$\sup_{(t,s)\in D_*} |\overline{Z}(\beta_0; t, s) - K_n(t, s)| \to 0 \quad \text{a.s.} \tag{2.6}$$

By Theorem 2.1 and the strong representation theorem [Pollard (1990), Theorem 9.4], we have, in another probability space, (2.6) and

$$\sup_{(t,s)\in D_*} |n^{-1/2} U_1(t, s) - \xi_1(t, s)| \to 0 \quad \text{a.s.}$$

In view of the preceding convergence, (2.6) and Lemma A.3,

$$\sup_{(t,s)\in D_*} \left| n^{-1/2} \sum_{i=1}^{n} \int_0^s \left[\overline{Z}(\beta_0; t, u) - K_n(t, u) \right] M_i(t, du) \right| = o_p(1), \tag{2.7}$$

which holds in the original probability space since the statement is now "in probability." Thus convergence of $n^{-1/2}U$ to ξ reduces to that of $n^{-1/2}\tilde{U}$ to ξ, where

$$\tilde{U}(t, s) = \sum_{i=1}^{n} \int_0^s [Z_i(u) - K_n(t, u)] M_i(t, du). \tag{2.8}$$

Convergence of finite-dimensional distributions of $n^{-1/2}\tilde{U}$ to those of ξ is straightforward by the classical multivariate central limit theorem, since $\tilde{U}$ is a sum of independent random variables. It remains to show tightness for $n^{-1/2}\tilde{U}$, or, equivalently, tightness for

$$n^{-1/2}\tilde{U}_K(t, s) = n^{-1/2} \sum_{i=1}^{n} \int_0^s K_n(t, u) M_i(t, du)$$

because $n^{-1/2}\sum_{i=1}^n \int_0^s Z_i(u) M_i(t, du)$ is tight by Theorem 2.1. In analogy with the proof of Theorem 2.1, it suffices to check that $n^{-1/2}\tilde{U}_K$ satisfies conditions (i)–(v) in Theorem 10.7 of Pollard (1990). As before, (ii)–(iv) are trivial and their verifications are omitted. For (v), let

$$
\begin{aligned}
&\tilde{\rho}_n((t_1, s_1), (t_2, s_2)) \\
&\quad = E\|n^{-1/2}\tilde{U}_K(t_1, s_1) - n^{-1/2}\tilde{U}_K(t_2, s_2)\|^2 \\
&\quad = \frac{1}{n}\sum_{i=1}^n E\int_0^{s_1}\|K_n(t_1,u)\|^2 \exp(\beta_0' Z_i(u)) Y_i(t_1,u)\lambda_0(u)\,du \\
&\qquad + \frac{1}{n}\sum_{i=1}^n E\int_0^{s_2}\|K_n(t_2,u)\|^2 \exp(\beta_0' Z_i(u)) Y_i(t_2,u)\lambda_0(u)\,du \\
&\qquad - \frac{1}{n}\sum_{i=1}^n E\int_0^{s_1\wedge s_2}[K_n'(t_1,u)K_n(t_2,u) + K_n'(t_2,u)K_n(t_1,u)] \\
&\qquad\qquad \times \exp(\beta_0' Z_i(u)) Y_i(t_1\wedge t_2, u)\lambda_0(u)\,du.
\end{aligned}
$$

By condition 3, it is easy to see that as $n \to \infty$, uniformly on $D_* \times D_*$, $\tilde{\rho}_n$ converges to a pseudometric $\tilde{\rho}$. Thus (v) holds. Furthermore, since it is not difficult to show that $\tilde{\rho}$ is continuous on $D_* \times D_*$, the limiting Gaussian random field, which has continuous sample path with respect to $\tilde{\rho}$, has continuous sample path with respect to the Euclidean metric.

It remains to verify (i), the manageability condition. Note that

$$
\begin{aligned}
&n^{-1/2}\sum_{i=1}^n \int_0^s K_n(t,u) M_i(t, du) \\
&\quad = n^{-1/2}\sum_{i=1}^n K_n(t, T_i) I_{(T_i \le C_i \wedge s)} I_{(T_i \le (t-\tau_i)^+)} \\
&\qquad - n^{-1/2}\sum_{i=1}^n \int_0^s K_n(t,u) I_{(T_i \wedge C_i \ge u)} I_{((t-\tau_i)^+ \ge u)} \\
&\qquad\qquad \times \exp(\beta_0' Z_i(u))\lambda_0(u)\,du.
\end{aligned} \tag{2.9}
$$

Because we can deal with K_n componentwise, we shall assume, without loss of generality, that the Z_i are scalars. We may further assume $Z_i \ge 0$ since $Z_i = Z_i^+ - Z_i^-$. In view of its definition,

$$
\begin{aligned}
K_n(t, T_i) &= \frac{E\left[\sum_{j=1}^n Z_j(s)\exp\left(\beta_0 Z_j(s)\right) I_{((t-\tau_j)+\wedge C_j \ge s)}\right]|_{s=T_i}}{E\left[\sum_{j=1}^n \exp\left(\beta_0 Z_j(s)\right) I_{((t-\tau_j)+\wedge C_j \ge s)}\right]|_{s=T_i}} \\
&= \frac{K_{n,1}(t, T_i)}{K_{n,2}(t, T_i)}, \quad \text{say.}
\end{aligned}
$$

Because of their monotonicity, by Lemma A.2, $\{K_{n,1}(t, T_i)\}$, $\{1/K_{n,2}(t, T_i)\}$, $\{I_{(T_i \le C_i \wedge s)}\}$ and $\{I_{(T_i \le (t-\tau_i)^+)}\}$ all have pseudodimension 1. Thus, in view of formula (5.2) of Pollard (1990), it is easy to see that the first term on the right-hand side of is manageable. On the other hand, the second term on the right-hand side of (2.9) can be written as

$$\sum_{i=1}^{n} \int_0^{t^*} K_n(t,u) I_{(T_i \wedge C_i \geq u)} I_{((t-\tau_i)^+ \geq u)} \exp(\beta_0 Z_i(u)) I_{(u \leq s)} \lambda_o(u)\, du.$$

By Theorem 6.2 of Pollard (1990), to show manageability of the preceding integral process it suffices to show that

$$\{K_n(t,u) I_{(u \leq s)} I_{(T_i \wedge C_i \geq u)} I_{((t-\tau_i)^+ \geq u)} \exp(\beta_0 Z_i(u))\} \tag{2.10}$$

is Euclidean. Since total variations of $\exp(\beta_0 Z_i(u))$ are uniformly bounded, they can be expressed as differences of increasing processes; thus they must be Euclidean by Lemmas A.1 and A.2. Furthermore, it is trivial that $\{K_n(t,u) I_{(u \leq s)}\}$ has pseudodimension 1 because it does not involve i, that $\{I_{(T_i \wedge C_i \geq u)}\}$ has pseudodimension 1 because of monotonicity and that $\{I_{((t-\tau_i)^+ \geq u)}\}$ has pseudodimension 1 since for any i, j, $\{(I_{((t-\tau_i)^+ \geq u)}, I_{((t-\tau_j)^+ \geq u)})\}$ takes at most three points. In view of these, (2.10) is Euclidean. Thus $n^{-1/2}\tilde{U}$ is tight. Finally, since $\tilde{\rho}$ is continuous on $D_* \times D_*$, the topology it induces is smaller than the usual one on $D_* \times D_*$ and therefore the limiting Gaussian random field has continuous sample path in the usual sense. □

3. Convergence under contiguous alternatives. Theorem 2.2 shows that the two-parameter score process $n^{-1/2}U(\beta_0; \cdot, \cdot)$ converges in distribution to a Gaussian random field, provided β_0 is the true regression parameter vector. Based on this, statistical tests with asymptotically correct significance level may be constructed. Calculation of their asymptotic powers, however, requires knowing distributional behavior of $n^{-1/2}U(\beta_0; \cdot, \cdot)$ under a sequence of contiguous alternatives. This section is devoted to developing a functional central theorem for $n^{-1/2}U(\beta_0; \cdot, \cdot)$ when the true parameter is $\beta_0 + b/\sqrt{n}$.

To be specific, let b be any fixed vector in R^p and $\beta_n = \beta_0 + b/\sqrt{n}$. For each fixed n, we assume that the true probability model is specified by (1.1) with $\beta = \beta_n$. Accordingly, P and E now stand for probability and expectation under this parameter specification, as do expectations in (2.1), (2.7) and Conditions 1–3 whenever we may refer to them in this section. Notation t^* and D_* remain the same. Let $U_1(\beta; t, s)$ and $U_2(\beta; t, s)$ be the same as in Section 2.

THEOREM 3.1. *Suppose for each n, β_n is the true parameter vector. Under Conditions* 1 *and* 2, *$n^{-1/2}U_1(\beta_n, \cdot, \cdot)$ and $n^{-1/2}U_2(\beta_n; \cdot, \cdot)$ converge in distribution to ξ_1 and ξ_2, where ξ_1 and ξ_2 are the same zero-mean random fields as those specified in Theorem* 2.1.

As explained in the proof of Theorem 2.1, the key is to show the so-called tightness, which reduces to verification of manageability. It is easy to see that proof of Theorem 2.1 can be borrowed, word for word, to show this, since β_0 does not play any role. We will not, however, repeat the same argument.

THEOREM 3.2. *Under the same assumptions as those of Theorem* 3.1, $n^{-1/2}U(\beta_0;\cdot,\cdot)$ *converges in distribution to* $\xi+\mu b$, *where* ξ *is the same zero-mean Gaussian random field as given in Theorem* 2.2 *and*

$$\mu(t,s)=\int_0^s\left[\Gamma_2(t,u)-\Gamma_1^{\otimes 2}(t,u)/\Gamma_0(t,u)\right]\lambda_0(u)\,du. \tag{3.1}$$

PROOF. We can write

$$\begin{aligned}
&n^{-1/2}U(\beta_0;t,s)\\
&\quad=n^{-1/2}\sum_{i=1}^n\int_0^s\left[Z_i(u)-\bar{Z}(\beta_0;t,u)\right]M_i(\beta_n;t,du)\\
&\qquad+n^{-1/2}\sum_{i=1}^n\int_0^s\left[Z_i(u)-\bar{Z}(\beta_0;t,u)\right]\\
&\qquad\qquad\times\left[\exp(\beta_n'Z_i(u))-\exp(\beta_0'Z_i(u))\right]Y_i(t,u)\lambda_0(u)\,du.
\end{aligned} \tag{3.2}$$

Since we are dealing with convergence in distribution, we may, by constructing a new probability space, assume that for all n, the relevant random variables $\{(T_i,C_i,\tau_i,Z_i): 1\le i\le n\}$, which are double arrays, that is, $T_i=T_{i,n}$ and so on live in the same probability space.

The first term on the right-hand side of (3.2) can be expressed by

$$n^{-1/2}U_2(\beta_n;t,s)-n^{-1/2}\int_0^s\bar{Z}(\beta_0;t,u)U_1(\beta_n;t,du). \tag{3.3}$$

By using the exponential inequality (7.3) instead of Theorem 8.3 [both in Pollard (1990)], we can show that (2.6) still holds here. In view of Theorem 3.1, we can apply Lemma A.3 and the argument involving the strong representation, as in proving (2.7), to show that (3.3) is asymptotically equivalent to

$$n^{-1/2}U_2(\beta_n;t,s)-n^{-1/2}\int_0^s K_n(t,u)U_1(\beta_n;t,du). \tag{3.4}$$

The last part of the proof Theorem 2.2 for showing convergence of $n^{-1/2}\tilde{U}$ to ξ is applicable to (3.4) and thus we have the convergence of the first term on the right-hand side of (3.2) to ξ.

Finally, by the Taylor series expansion, the second term on the right-hand side of (3.2) is

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n\int_0^s\left[Z_i(u)-\bar{Z}(\beta_0;t,u)\right]Z_1'(u)\exp(\beta_0'Z_i(u))\\
&\qquad\times Y_i(t,u)\lambda_0(u)\,du\,b+O(n^{-1/2}),
\end{aligned} \tag{3.5}$$

where O is uniform in $(t, s) \in D_*$. By Lemma A.4 and exponential inequality (7.3) of Pollard (1990), it follows easily that, with probability 1,

$$\sup_{(t,s)\in D_*}\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^s Z_i^{\otimes k}(u)\exp(\beta_0' Z_i(u))Y_i(t,u)\lambda_0(u)\,du - \frac{1}{n}\sum_{i=1}^{n}E\int_0^s Z_i^{\otimes k}(u)\exp(\beta_0' Z_i(u))Y_i(t,u)\lambda_0(u)\,du\right| \to 0, \quad k = 1, 2. \tag{3.6}$$

From Condition 3, (2.6) and (3.6) with $k = 1$, we have

$$\sup_{(t,s)\in D_*}\left|\int_0^s \bar{Z}(\beta_0; t, u)\frac{1}{n}\sum_{i=1}^{n} Z_i'(u)\exp(\beta_0' Z_i(u))Y_i(t,u)\lambda_0(u)\,du - \int_0^s K(t,u)\frac{1}{n}\sum_{i=1}^{n}E\left[Z_i'(u)\exp(\beta_0' Z_i(u))Y_i(t,u)\right]\lambda_0(u)\,du\right| \to 0. \tag{3.7}$$

However, it is not difficult to see that

$$\int_0^s K(t,u)\frac{1}{n}\sum_{i=1}^{n}E\left[Z_i'(u)\exp(\beta_0' Z_i(u))Y_i(t,u)\right]\lambda_0(u)\,du = \frac{1}{n}\sum_{i=1}^{n}E\int_0^{s\wedge(t-\tau_i)+\wedge T_i\wedge C_i} K(t,u)Z_i'(u)\exp(\beta_0' Z_i(u))\lambda_0(u)\,du$$

is equicontinuous in $(t, s) \in D_*$ and converges pointwise to

$$\int_0^s \Gamma_1^{\otimes 2}\frac{(t,u)}{\Gamma_0(t,u)\lambda_0(u)\,du}.$$

This, combined with (3.5)–(3.7), shows that the second term in (3.2) converges uniformly in D_* to

$$\int_0^s \left[\Gamma_2(t,u) - \Gamma_1^{\otimes 2}(t,u)/\Gamma_0(t,u)\right]\lambda_0(u)\,du\, b.$$

Hence the theorem holds. □

4. Convergence of the maximum partial likelihood estimator. In this section, we prove uniform strong consistency and functional CLT for the sequentially computed maximum partial likelihood estimator of β_0 and Nelson–Aalen estimator of Λ_0. For each $(t, s) \in D_*$, define the Cox partial likelihood estimator $\hat{\beta}(t, s)$ as a solution to $U(\beta; t, s) = 0$. As in Section 2, β_0 denotes the true regression parameter. We need the following condition for the consistency of $\hat{\beta}$.

CONDITION 4. There exists t_* in $(0, t^*)$ such that

$$\liminf_{n\to\infty} \lambda_{\min}\left(\frac{1}{n} \sum_{i=1}^{n} E\int_0^{t_*} [Z_i(s) - K_n(t_*, s)]^{\otimes 2} Y_i(t_*, s) \times \exp(\beta_0' Z(s)) \lambda_0(s)\, ds \right) = r_0 > 0,$$

where $\lambda_{\min}(A)$ of a symmetric matrix A denotes its minimum eigenvalue.

THEOREM 4.1. *Suppose that Conditions* 1 *and* 4 *are satisfied. Then* $\hat\beta(t, s)$ *is uniformly strongly consistent in the sense*

$$\lim_{n\to\infty} \sup_{t_* \le s \le t \le t^*} \| \hat\beta(t, s) - \beta_0 \| = 0 \quad \text{a.s.}$$

PROOF. From (2.3) and (2.4) and an obvious extension of them to cover $k = 2$, we know that under Condition 1,

$$\lim_{n\to\infty} \sup_{(t,s)\in D_*} \left| \frac{1}{n} \sum_{i=1}^{n} Y_i(t, s) Z_i^{\otimes k}(s) \exp(\beta_0' Z_i(s)) - \frac{1}{n} \sum_{i=1}^{n} E\big[Y_i(t, s) Z_i^{\otimes k}(s) \exp(\beta_0' Z_i(s)) \big] \right| = 0 \quad \text{a.s.,} \tag{4.1}$$

for $k = 0, 1, 2$. Now

$$\frac{1}{n} U(\beta_0; t, s) = \frac{1}{n} \sum_{i=1}^{n} \int_0^s Z_i(u) M_i(t, du) - \int_0^s \bar Z(\beta_0; t, u) \frac{1}{n} \sum_{i=1}^{n} M_i(t, du). \tag{4.2}$$

In view of the manageability in the proof of Theorem 2.1 and the uniform strong law of large numbers of Pollard [(1990), Theorem 8.3], the first term on the right-hand side of (4.2) goes to 0 uniformly in D_*. Likewise, $n^{-1}\sum_{i=1}^{n} M_i$ also goes to 0 uniformly in D_*. Consequently, Lemma A.3 can be applied to show that the second term on the right-hand side of (4.2) also goes to 0. Thus

$$\lim_{n\to\infty} \sup_{(t,s)\in D_*} \left| \frac{1}{n} U(\beta_0; t, s) \right| = 0 \quad \text{a.s.} \tag{4.3}$$

By (4.1), we can easily show that with probability 1,

$$\sup_{(t,s)\in D_*} \left| \frac{1}{n} \frac{\partial}{\partial\beta} U(\beta_0; t, s) - \int_0^s \frac{1}{n} \sum_{i=1}^{n} E\Big[(Z_i(u) - K_n(t, u))^{\otimes 2} \times Y_i(t, u) \exp(\beta_0' Z(u)) \Big] \lambda_0(u)\, du \right| \to 0. \tag{4.4}$$

Since $\{(t,s): t_* \le s \le t \le t^*\} \subset D_*$, Condition 4 and (4.4) entail

$$\liminf_{n\to\infty} \inf_{t_* \le s \le t \le t^*} \lambda_{\min}\left(\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta_0; t, s)\right) = r_0 > 0 \quad \text{a.s.} \tag{4.5}$$

Since $n^{-1}(\partial/\partial\beta)U(\beta; t, s)$ has uniformly bounded derivative (with respect to β), (4.5) implies that there exists a neighborhood of β_0, $\mathscr{N}(\beta_0)$, such that

$$\liminf_{n\to\infty} \inf_{t_* \le s \le t \le t^*} \inf_{\beta\in\mathscr{N}(\beta_0)} \lambda_{\min}\left(\frac{1}{n}\frac{\partial}{\partial\beta}U(\beta; t, s)\right) \ge \frac{r_0}{2} > 0. \tag{4.6}$$

From (4.3), (4.6) and Lemma A.5 follows the desired uniform convergence. □

THEOREM 4.2. *Under Conditions* 1–4, $\{\sqrt{n}(\hat\beta(t,s) - \beta_0), t_* \le s \le t \le t^*\}$ *converges in distribution to a vector-valued Gaussian random field* η, *which has mean* 0 *and covariance*

$$E(\eta(t_1, s_1)\eta'(t_2, s_2)) = \mu^{-1}(t_1, s_1)\mu(t_1 \wedge t_2, s_1 \wedge s_2)\mu^{-1}(t_2, s_2),$$

where μ *is defined by* (3.1).

An immediate corollary of this is the convergence of the diagonal process $\{\sqrt{n}(\hat\beta(t,t) - \beta_0), t \in [t_*, t^*]\}$ to $\{\eta(t,t), t \in [t_*, t^*]\}$.

PROOF OF THEOREM 4.2. Since $\hat\beta(t,s)$ converges to β_0 uniformly on $\{t_* \le s \le t \le t^*\}$, we have, via the usual Taylor expansion,

$$\begin{aligned} U(\hat\beta(t,s); t, s) &= U(\beta_0; t, s) \\ &\quad + \left[\frac{\partial}{\partial\beta}U(\beta_0, t, s) + o_p(n)\right]\left[\hat\beta(t,s) - \beta_0\right], \end{aligned} \tag{4.7}$$

where o_p is uniform for $t_* \le s \le t \le t^*$. From (4.7), (4.4) and Theorem 2.1, we easily conclude that $\sqrt{n}(\hat\beta(\cdot,\cdot) - \beta_0)$ converges to η. □

We now discuss estimation of $\Lambda_0(s)$. If β_0 were known, then the Nelson–Aalen estimator would be $\hat\Lambda(\beta_0; t, s)$, where

$$\hat\Lambda(\beta; t, s) = \int_0^s \frac{\sum_{i=1}^n N_i(t, du)}{\sum_{i=1}^n Y_i(t,u)\exp(\beta' Z_i(u))}. \tag{4.8}$$

Since at time t, the best available estimator for β_0 is $\hat\beta(t,t)$, it is natural to use $\hat\Lambda(\hat\beta(t,t); t, s)$. Convergence of this estimator is given by the following theorem.

THEOREM 4.3. *Let* $\tilde D_* = \{(t,s): t_* \le t \le t^*, s \le t\}$. *Under Conditions* 1–4, $\{\sqrt{n}[\hat\Lambda(\hat\beta(t,t); t, s) - \Lambda_0(s)], (t,s) \in \tilde D_*\}$, *converges in distribution to a*

Gaussian random field ζ *with mean* 0 *and covariance function specified by*

$$E[\zeta(t_1, s_1)\zeta(t_2, s_2)] = \int_0^{s_1 \wedge s_2} \frac{d\Lambda_0(u)}{\Gamma_0(t_1 \vee t_2, u)} + a'(t_2, s_2)\mu^{-1}(t_1 \vee t_2, t_1 \vee t_2)a(t_1, s_1),$$

where $a(t, s) = \int_0^s \Gamma_1(t, u)/\Gamma_0(t, u)\, d\Lambda_0(u)$.

PROOF. Again taking the Taylor expansion of $\hat{\Lambda}(\beta; t, s)$ at $\beta = \beta_0$, we can get

$$\begin{aligned} &\hat{\Lambda}(\hat{\beta}(t,t); t, s) - \Lambda_0(s) \\ &\quad = \hat{\Lambda}(\beta_0; t, s) - \Lambda_0(s) - \int_0^s \frac{\sum_{i=1}^n Y_i(t,u)\exp(\beta_0' Z_i(u)) Z_i'(u)}{(\sum_{i=1}^n Y_i(t,u)\exp(\beta_0' Z_i(u)))^2} \\ &\qquad \times \sum_{i=1}^n N_i(t, du)(\hat{\beta}(t,t) - \beta_0) + o_p(n^{-1/2}), \end{aligned} \tag{4.9}$$

where o_p is uniform in $(t, s) \in \tilde{D}_*$. By appealing to the uniform strong law of large numbers as in the proof of Theorem 4.1, it follows that

$$\int_0^s \frac{\sum_{i=1}^n Y_i(t,u)\exp(\beta_0' Z_i(u)) Z_i'(u)}{[\sum_{i=1}^n Y_i(t,u)\exp(\beta_0' Z_i(u))]^2} \sum_{i=1}^n N_i(t, du) = a(t, s) + o_p(1).$$

This, together with (4.8) and (4.9), gives

$$\begin{aligned} n^{1/2}(4.9) = &\int_0^s \frac{n^{-1/2}\sum M_i(t, du)}{n^{-1}\sum Y_i(t,u)\exp(\beta_0' Z_i(u))} \\ &- a'(t, s)\mu^{-1}(t,t)n^{-1/2}U(\beta_0; t, t) + o_p(1). \end{aligned}$$

From Theorem 2.2, we know that the second term on the right-hand side of the preceding equation is tight. The tightness of the first term is easily seen in view of Theorem 1 and Lemma A.3. Convergence of the finite-dimensional distributions of (4.8) to those of ζ is straightforward by the multivariate central limit theorem along with a routine variance–covariance calculation. □

5. Score process with covariate adjustment and concluding remarks. As an application of the preceding general theory, we consider the problem of testing the effect of one covariate component while adjusting for effects of other components. In addition to the previous notation, introduce δ, γ, X_i and W_i specified through $\beta' = (\delta, \gamma')$ and $Z_i'(s) = (X_i(s), W_i'(s))$. Suppose, for simplicity, it is desired to test H_0: $\delta_0 = 0$, while γ is treated as a nuisance parameter.

If γ_0, the true value of γ, is known, then a relevant two-parameter Cox score process with staggered entry is $V(\gamma_0; t, s)$, where

$$V(\gamma; t, s) = \sum_{i=1}^{n} \int_0^s \left(X_i(t) - \frac{\Sigma X_j(u) \exp(\gamma' W_j(u)) Y_j(t, u)}{\Sigma \exp(\gamma' W_j(u)) Y_j(t, u)} \right) N_i(t, du). \tag{5.1}$$

Since γ_0 is unknown in this setup, we propose to use $V(\hat{\gamma}_t; t, s)$, where $\hat{\gamma}_t$ is an estimator of γ_0 from the available data at time t and is defined as a solution to

$$\sum_{i=1}^{n} \int_0^t \left(W_i(u) - \frac{\Sigma W_j(u) \exp(\gamma' W_j(u)) Y_j(t, u)}{\Sigma \exp(\gamma' W_j(u)) Y_j(t, u)} \right) N_i(t, du) = 0. \tag{5.2}$$

Tests that incorporate covariate adjustment are useful in reducing bias and increasing efficiency. Tsiatis, Rosner and Tritchler (1985) showed, under the assumption that W_i and X_i are independent, that the finite-dimensional distributions of $n^{-1/2} V(\hat{\gamma}_t; t, t)$ and those of $n^{-1/2} V(\gamma_0; t, t)$ are asymptotically equivalent and the limiting process is a time-rescaled Brownian motion. The more general case in which W_i and X_i may be dependent is studied in Gu and Ying (1995), where it is shown that the limiting process of $n^{-1/2} V(\hat{\gamma}_t; t, t)$ is still a time-rescaled Brownian motion.

The preceding asymptotic theory enables us now to derive a functional central limit theorem for the more general two-parameter process $\{n^{-1/2} V(\hat{\gamma}_t; t, s), (t, s) \in \tilde{D}_*\}$. Let

$$\overline{X}(t, s) = \sum_{i=1}^{n} X_i(s) \exp\{\gamma_0' W_i(s)\} Y_i(t, s) \Big/ \sum_{i=1}^{n} \exp\{\gamma_0' W_i(s)\} Y_i(t, s),$$

$$\overline{W}(t, s) = \sum_{i=1}^{n} W_i(s) \exp\{\gamma_0' W_i(s)\} Y_i(t, s) \Big/ \sum_{i=1}^{n} \exp\{\gamma_0' W_i(s)\} Y_i(t, s).$$

It is clear that, under Condition 2, the following limits are well defined with probability 1:

$$\mu_{xw}(t, s) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^s \left(X_i(u) - \overline{X}(t, u) \right) \times \left(W_i(u) - \overline{W}(t, u) \right)' \exp(\gamma_0' W_i(u)) Y_i(t, u) \lambda_0(u)\, du,$$

$$\mu_{ww}(t) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left(W_i(u) - \overline{W}(t,u) \right)^{\otimes 2} \exp(\gamma_0' W_i(u)) Y_i(t,u) \lambda_0(u)\, du,$$

$$\mu_{xx}(t, s) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \int_0^s \left(X_i(u) - \overline{X}(t, u) \right)^2 \exp(\gamma_0' W_i(u)) Y_i(t, u) \lambda_0(u)\, du.$$

THEOREM 5.1. *Suppose that Conditions* 1–3 *are satisfied and* $\mu_{ww}(t) > 0$. *Then* $\{n^{-1/2}V(\hat{\gamma}_t; t, s), (t, s) \in \tilde{D}_*\}$ *converges in distribution to a Gaussian process* ξ_{adj} *with mean* 0 *and covariance function* $\sigma^2_{\text{adj}}(t_1 \wedge t_2, s_1 \wedge s_2)$, *where*

$$\sigma^2_{\text{adj}}(t, s) = \mu_{xx}(t, s) - \mu_{xw}(t, s)\mu^{-1}_{ww}(t)\mu'_{xw}(t, s). \tag{5.3}$$

It is clear from (5.3) that the "diagonal" process $\xi_{\text{adj}}(t, t)$ has independent increments. In fact with some elementary algebra, one can show directly that its variance function $\sigma^2_{\text{adj}}(t, t)$ is increasing in t. From Theorem 5.1 follows immediately Corollary 5.1 for the convergence of weighted log-rank statistics.

COROLLARY 5.1. *Let* $w_n(t, s)$ *be a sequence of weight functions as in Corollary* 2.1. *Define weighted log-rank process* $V_w(t) = \int_0^t w_n(t, s)V(t, ds)$. *Then under the same conditions as in Theorem* 5.1, $\{n^{-1/2}V_w(t): t_* \le t \le t^*\}$ *converges in distribution to a Gaussian process with mean* 0 *and covariance function*

$$\sigma^2_{w,\text{adj}}(t, \tilde{t}) = \int_0^{t\wedge\tilde{t}} w(t, s)w(\tilde{t}, s)\sigma^2(t \wedge \tilde{t}, ds),$$

recalling that w *is the limit of* w_n. *In particular, if* $w(t, s)$ *does not involve* t, *then the limiting process* V_w *has independent increments.*

PROOF OF THEOREM 5.1. Proof of Theorem 5.1 becomes rather straightforward in view of Theorems 2.2 and 4.2. Expanding $V(\hat{\gamma}_t; t, s)$ at γ_0 and in view of Theorem 4.2 and (4.7),

$$\begin{aligned} V(\hat{\gamma}_t; t, s) &= V(\gamma_0; t, s) - \mu_{xw}(t, s)\mu^{-1}_{ww}(t) \\ &\quad\times \sum_{i=1}^n \int_0^t \big(W_i(u) - \overline{W}(t, u)\big)M_i(t, du) + o_p(n^{1/2}), \end{aligned} \tag{5.4}$$

where o_p is uniform in $(t, s) \in \tilde{D}_*$. From (5.4) we can apply the same argument as in Section 2 to show tightness of $V(\hat{\gamma}_t; t, s)$. Convergence of its finite-dimensional distributions follows from a simple covariance calculation. □

We have developed in this paper a general asymptotic theory useful for statistical inference related to Cox's proportional hazards regression with staggered entry data. It has been known that the usual approach via counting processes and their associated martingales is in general not suitable. Our method makes use of a modern empirical process theory, which treats uni- and multi-parameter processes in a unified way. It greatly simplifies proofs of the so-called tightness as well as producing asymptotic results in great generality. The results we obtained include functional central limit theorems for the two-parameter Cox score process, the related maximum partial likelihood estimator of the regression parameter vector and the Nelson–Aalen estimator of the cumulative baseline hazard function. Convergence under contiguous alternatives is also derived. In addition, we have shown how these results may be used to derive convergence of a weighted log-rank score process for testing one covariate component while adjusting for others.

The main application of the results and the tools developed in this paper is to sequential tests under the Cox model. However, there are other situations which could benefit from the present investigation. For example, Lin, Shen, Ying and Breslow (1996) recently proposed a one-arm sequential design whose asymptotic behavior can be derived by applying the techniques developed here. Furthermore, a parallel theory for the accelerated failure time (AFT) model may be developed along the same line. Note that there appears to be no result for the AFT model parallel to that of Sellke and Siegmund (1983). Another area in which our approach may be adopted is the type of truncation model arising from analysis of warranty data as described in Kalbfleisch, Lawless and Robinson (1991).

APPENDIX

LEMMA A.1. *Suppose that* $\{f_i\}$ *and* $\{g_i\}$ *are manageable with respect to a common envelope* $\{F_i\}$. *Then* $\{f_i + g_i\}$ *are manageable with respect to* $\{2F_i\}$.

The proof is trivial in view of the definition of manageability.

LEMMA A.2. *Let* T *be a subset of the real line. Suppose that* $f_i(t)$ *is nondecreasing in* t *for every* $i = 1, \dots, n$. *Then* $\{f_i(t): t \in T\}$ *has pseudodimension* 1.

PROOF. For any fixed i, j, following Pollard [(1990), Definition 4.3], we need to show that the set $\{(f_i(t), f_j(t)), t \in T\}$ cannot surround any point in R^2. This is clear because $\{\cdot\}$ is a well-ordered subset in R^2. □

LEMMA A.3. *Let* $D = [a, b] \times [c, d] \subset R^2$. *Suppose that*

$$\lim_{n\to\infty} \sup_{(t,s)\in D} \left\{ |h_n(t,s) - h(t,s)| + |J_n(t,s) - \tilde{J}_n(t,s)| \right\} = 0,$$

where h *is continuous on* D, *and*, *for each fixed* t, $J_n(t, \cdot)$ *and* $\tilde{J}_n(t, \cdot)$ *are left continuous*, *with their total variations bounded by a constant* $\overline{B}$, *independent of* n *and* t. *Then*

$$\lim_{n\to\infty} \sup_{(t,s)\in D} \left| \int_c^s h_n(t,u) J_n(t,du) - \int_c^s h(t,u) \tilde{J}_n(t,du) \right| = 0, \tag{A.1}$$

$$\lim_{n\to\infty} \sup_{(t,s)\in D} \left| \int_c^s h_n(t,u) J_n(t,du) - \int_c^s h_n(t,u) \tilde{J}_n(t,du) \right| = 0. \tag{A.2}$$

REMARK. The lemma also holds when $D = \{(t,s): s \le t, t \in [a,b]$ and $s \in [c,d]\}$, because we can extrapolate the functions via $h_n(t,s) = h_n(t,t)$, whenever $s \ge t$.

PROOF. First, since h_n converges uniformly to h and $J_n(t, \cdot)$ has bounded variation,

$$\lim_{n\to\infty} \sup_{(t,s)\in D} \left| \int_c^s h_n(t,u) J_n(t,du) - \int_c^s h(t,u) J_n(t,du) \right| = 0, \tag{A.3}$$

$$\lim_{n\to\infty} \sup_{(t,s)\in D} \left| \int_c^s h_n(t,u) \tilde{J}_n(t,du) - \int_c^s h(t,u) \tilde{J}_n(t,du) \right| = 0. \tag{A.4}$$

From (A.4) we know that (A.1) implies (A.2).

Since h is continuous, we can find partitions $a = t_0 < t_1 \cdots < t_{n_0} = b$ and $c = s_0 < s_1 \cdots < s_{m_0} = d$ and constants $h_{ij} (= h(t_i, s_j))$ such that the simple function

$$h_\varepsilon(t,s) = \sum_{j=1}^{m_0} \sum_{i=1}^{n_0} h_{ij} I_{((t,s)\in(t_{i-1},t_i]\times(s_{j-1},s_j])}$$

satisfies $\sup_{(t,s)\in D} |h_\varepsilon(t,s) - h(t,s)| < \varepsilon$. Thus

$$\begin{aligned}
&\left| \int_c^s h(t,u) J_n(t,du) - \int_c^s h(t,u) \tilde{J}_n(t,du) \right| \\
&\quad \le \left| \int_c^s [h(t,u) - h_\varepsilon(t,u)] J_n(t,du) \right| \\
&\qquad + \left| \int_c^s h_\varepsilon(t,u) \big[J_n(t,du) - \tilde{J}_n(t,du) \big] \right| \\
&\qquad + \left| \int_c^s [h(t,u) - h_\varepsilon(t,u)] \tilde{J}_n(t,du) \right| \\
&\quad \le 2\varepsilon \bar{B} + 2 \sum_{j=1}^{m_0} \sum_{i=1}^{n_0} |h_{ij}| \sup_{(v,u)\in D} |J_n(v,u) - \tilde{J}_n(v,u)| \\
&\quad \to 2\varepsilon \bar{B} \quad \text{as } n \to \infty.
\end{aligned}$$

This in conjunction with (A.3) implies (A.1). □

LEMMA A.4. *Let T be a compact subset in R^d. Suppose that the set of triangular array $\{f_{ni}\}$ with envelopes F_{ni} satisfies, for all t, $s \in T$, $|f_{ni}(t) - f_{ni}(s)| \le F_{ni} \|t - s\|$. Then $\{f_{ni}\}$ is Euclidean (therefore manageable) with respect to $\{F_{ni}\}$.*

PROOF. For ε small enough, the number of points in T that are ε distance apart must be less than $(1/\varepsilon)^{d+1}$. But for any $\{\alpha_i\}$,

$$\left[\sum_{i=1}^n \big(\alpha_i f_{ni}(t) - \alpha_i f_{ni}(s) \big)^2 \right]^{1/2} \le \left[\sum_{i=1}^n \alpha_i^2 F_{ni}^2 \right]^{1/2} \|t - s\|.$$

Thus the number of points in $\{f_{ni}\}$ that are $\varepsilon(\sum_{i=1}^n \alpha_i^2 F_{ni}^2)^{1/2}$ apart must be less than $(1/\varepsilon)^{d+1}$. □

LEMMA A.5. *Let* $\{f_{n,\alpha}: n \geq 1, \alpha \in A\}$ *be a set of functions from* R^d *to* R^d. *Suppose that* (i) $(\partial/\partial\theta)f_{n,\alpha}(\theta)$ *are nonnegative definite for all* n, α *and* θ; (ii) $\sup_\alpha \|f_{n,\alpha}(\theta_0)\| \to 0$ *as* $n \to \infty$; (iii) *there exists a neighborhood of* θ_0, *denoted by* $\mathscr{N}(\theta_0)$, *such that*

$$\liminf_{n\to\infty} \inf_{\theta\in\mathscr{N}(\theta_0)} \inf_{\alpha\in\mathscr{A}} \lambda_{\min}\left(\frac{\partial f_{n,\alpha}(\theta)}{\partial\theta}\right) > 0,$$

where, as in Condition 4, $\lambda_{\min}$ *denotes the minimum eigenvalue. Then there exists* n_0 *such that for every* $n \geq n_0$ *and* $\alpha \in \mathscr{A}$, $f_{n,\alpha}$ *has a unique root* $\theta_{n,\alpha}$ *and* $\sup_{\alpha\in\mathscr{A}}\|\theta_{n,\alpha} - \theta_0\| \to 0$.

PROOF. We first show the existence of the root. From (iii), we can find n_0 and $r > 0$ such that

$$\min_{\|\theta-\theta_0\|\leq r} \lambda_{\min}\left(\frac{\partial f_{n,\alpha}(\theta)}{\partial\theta}\right) \geq r \quad \text{for all } n \geq n_0 \text{ and } \alpha \in \mathscr{A}.$$

Thus for all θ such that $\|\theta - \theta_0\| = r$,

$$\begin{aligned}(\theta - \theta_0)'[f_{n,\alpha}(\theta) - f_{n,\alpha}(\theta)] \\ &= (\theta-\theta_0)'[f_{n,\alpha}(\theta_0 + u(\theta-\theta_0))|_{u=1} - f_{n,\alpha}(\theta_0)] \\ &= (\theta-\theta_0)'\frac{\partial}{\partial\theta}f_{n,\alpha}(\theta_0 + u^*(\theta-\theta_0))(\theta-\theta_0) \geq r^3,\end{aligned}$$

implying $\|f_{n,\alpha}(\theta) - f_{n,\alpha}(\theta_0)\| \geq r^2$. By Theorem 2 (and its proof) of Goffman (1965), we know that the image $f_{n,\alpha}(\{\theta: \|\theta - \theta_0\| \leq r\})$ contains $\{y: \|y - f_{n,\alpha}(\theta_0)\| \leq r^2/3\}$. From this and (ii) we have the existence of the root. Furthermore, because r may be chosen arbitrarily small, we can select a sequence of roots $\theta_{n,\alpha}$ such that $\sup_\alpha\|\theta_{n,\alpha} - \theta_0\| \to 0$.

To prove uniqueness, suppose for some $n \geq n_0$ and α, there is another root $\theta^*_{n,\alpha}$. Define

$$q(t) = (\theta^*_{n,\alpha} - \theta_{n,\alpha})'f_{n,\alpha}(\theta_{n,\alpha} + t(\theta^*_{n,\alpha} - \theta_{n,\alpha})).$$

Clearly $q(0) = 0$ and $q'(t) \geq 0$ with $q'(0) > 0$. So $q(1) > 0$, contradicting the assumption that $\theta^*_{n,\alpha}$ is a root of $f_{n,\alpha}$. □

Acknowledgments. The authors thank the referee for a careful reading and helpful comments. The initial version of the paper was written while the first and the third authors were at the University of Illinois and Columbia University, respectively.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

GILL, R. D. and SCHUMACHER, M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74** 289–300.

GOFFMAN, C. (1965). *Calculus of Several Variables*. Harper and Row, New York.

GU, M. G. and LAI, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Ann. Statist.* **19** 1403–1433.

GU, M. G. and YING, Z. (1995). Group sequential methods for survival data using partial likelihood score processes with covariate adjustment. *Statist. Sinica* **5** 793–804.

KALBFLEISCH, J. D., LAWLESS, J. F. and ROBINSON, J. A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics* **33** 273–285.

LAI, T. L. and YING, Z. (1988). Stochastic integral of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27** 334–358.

LIN, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J. Amer. Statist. Assoc.* **86** 725–728.

LIN, D. Y., SHEN, L., YING, Z. and BRESLOW, N. (1996). Group sequential designs for monitoring survival probabilities. *Biometrics* **52** 1033–1041.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. IMS, Hayward, CA.

SELLKE, T. and SIEGMUND, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70** 315–326.

SLUD, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *Ann. Statist.* **12** 551–571.

SLUD, E. and WEI, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Amer. Statist. Assoc.* **77** 862–868.

TSIATIS, A. A., ROSNER, G. L. and TRITCHLER, D. L. (1985). Group sequential tests with censored survival data adjusting for covariates. *Biometrika* **72** 365–373.

YANNIS BILLIS
DEPARTMENTS OF ECONOMICS
AND STATISTICS
266 HEADY HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50011

MINGGAO GU
DEPARTMENT OF MATHEMATICS
AND STATISTICS
MCGILL UNIVERSITY
MONTREAL, QUEBEC
CANADA H3A 2K6

ZHILIANG YING
DEPARTMENT OF STATISTICS
HILL CENTER, BUSCH CAMPUS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08855
E-MAIL: zying@stat.rutgers.edu