The Annals of Statistics
1993, Vol. 21, No. 4, 1721–1759

SOME NEW ESTIMATORS FOR COX REGRESSION[1]

BY PETER SASIENI

Imperial Cancer Research Fund

New estimators for Cox regression are considered. Their asymptotic properties, both on and off the model, are established. Corollaries include conditions under which the maximum partial likelihood estimator defines a parameter in the population and the asymptotics of the case-cohort estimator. Robust estimators that minimize the asymptotic variance subject to a bound on the maximal bias on infinitesimal neighborhoods are discussed. The estimators are illustrated with medical data.

1. Introduction. The proportional hazards model is often used to assess the influence of various factors on survival. Although a model should generally only be regarded as a useful approximation to the underlying process from which the data are generated, it is somewhat disturbing that if a proportional hazards model holds when several covariates are included, the omission of a covariate (even one that is orthogonal to the others) will lead to nonproportional hazards. In this paper we present asymptotic results that enable one to study the behavior of proportional hazard estimators in a general setup.

Alternatives to the maximum partial likelihood estimator (mple) for estimating the regression parameters in the Cox model are considered. The mple is known to be efficient if the Cox model holds [Begun, Hall, Huang and Wellner (1983) and Efron (1977)], but in practice one never knows whether the data are sampled from a population that conforms to the Cox model. A data analyst may perform a test to see whether particular model assumptions are valid, but this will only detect departures that are of at least a certain magnitude. (A classical goodness-of-fit test will not reject the model if the distance from the true distribution to the model is of order no larger than $n^{-1/2}$.) For this reason it seems sensible to adopt the position considered by Bickel (1984) and to look for an estimator that has minimum variance subject to a bound on the bias on an infinitesimal neighborhood of the model. On such neighborhoods variance and squared bias are of the same order. Thus the approach is equivalent to minimizing the maximum mean squared error over the neighborhood.

An applied statistician may, for want of something better, use the Cox model even after a formal test has rejected it. Since any model is only supposed to be a convenient approximation of the population, the Cox model may be worthwhile even when the model assumptions fail to hold. For this reason one would like to know something about the mple when the model is misspecified,

Received August 1990; revised November 1992.

[1]Supported in part by NIH grant 5T32CA09168.

AMS 1991 *subject classifications*. Primary 62F12; secondary 62F35, 62G05.

Key words and phrases. Case-cohort, contiguity, Cox model, influence function, partial likelihood, robust estimators, survival analysis.

and to compare its properties to that of other estimators. In particular, it would be useful to know: the variance of the estimator; how much it is influenced by small subgroups of the data; and, if possible, whether the functional defined by the estimator has a descriptive interpretation (such as a measure of relative risk). This paper starts to address these issues, but does not provide a satisfactory answer to all of them. Sasieni (1993) introduces a class of weighted estimators and compares them under the assumption that the data are a sample from a member of the Cox model. That paper also considers the effect of model outliers on the estimators. Lin (1991) proposes a goodness-of-fit test based on the same class of estimators. Here we consider a slightly larger family of estimators and prove their asymptotic normality under a general probability model. When the Cox model has been correctly specified, the estimators are all consistent for the regression parameter with the usual relative risk interpretation ($e^{\beta'(Z_1-Z_2)}$ is the relative risk between two individuals with covariates Z_1 and Z_2), but when it is misspecified different estimators will converge to different quantities and it is difficult to provide a physical interpretation to these "parameters." Direct comparisons can be made on contiguous alternatives to the Cox model, since then we can evaluate the asymptotic bias and variance of the estimators.

We start by considering a class of estimators that generalize maximum weighted partial likelihood estimators [Lin (1991) and Sasieni (1993)]. Section 3 gives sufficient conditions for the functionals associated with these estimators to be well defined. As a special case we give conditions under which the Cox functional (mple) is well defined, thus providing the "missing step" in the consistency proof of Struthers and Kalbfleisch (1986), who omit to show whether the β^* to which $\hat{\beta}$ converges actually exists. The next section considers the asymptotic properties of these estimators. Proofs are given at the end of the paper (Section 7). The special case of the mple has been considered by Bickel, Klaassen, Ritov and Wellner (1993) and by Lin and Wei (1989). Our treatment is more general than the first reference and more rigorous than the second. Other special cases include the asymptotic distribution of various censored linear rank tests and the asymptotic normality of the case-cohort estimator [Self and Prentice (1988)].

A useful robustness theory can be developed by considering the properties of estimators on contiguous alternatives to the Cox model. This is done in Section 5. In particular, we discuss the problem of constructing an estimator with minimum variance subject to a bound on the bias on local neighborhoods of the Cox model. Bednarski (1991) has considered a similar approach to robust estimation in the Cox model and proposes a slightly different family of estimators.

Finally, a variety of weighted estimators are compared on data concerning the prognostic value of serum $\beta 2$ microglobulin for myelomatosis patients (Section 6).

2. Estimators of class *K*. Consider first the Cox model: in the uncensored version, one observes a failure time, T^u, and covariates, Z. The model specifies that the conditional hazard of T^u given Z is $\lambda(t|Z) = \lambda_0(t)\exp(\beta'Z)$.

More generally, one observes the covariates Z; a random time T, which is the minimum of a failure time T^u and a censoring time T^c; and a 0–1 random variable, $\Delta = 1_{[T^u \le T^c]}$. If T^u and T^c are conditionally independent given Z and the conditional hazard of T^u is as above, then $X = (Z, T, \Delta) \sim P$ is said to come from the Cox model.

The state (sample) space for the Cox model is some subset of $\mathbb{R}^d \times \mathbb{R}_+ \times \{0, 1\}$. We wish to study the situation in which data of this sort do not necessarily come from the Cox model. To facilitate this, we will consider general probability measures on the Cox model's state space. Let $\mathscr{X}$ be an open subset of $\mathbb{R}^d \times \mathbb{R}_+ \times \{0, 1\}$ together with its Borel σ-field. Use Q to denote both a probability measure and the corresponding distribution function on $\mathscr{X}$. Thus if $X = (Z, T, \Delta)$ has distribution Q, then $Q(z, t, \delta) := Q(Z \le z,\ T \le t,\ \Delta = \delta)$. Suppose that $\{X_1, \dots, X_n\} \subset \mathscr{X}$: denote by $\mathbb{P}_n$ the measure placing mass $1/n$ at each of $X_1, \dots, X_n$.

Given a model $\mathscr{P}$ with an identifiable parameter $\beta \in \mathbb{R}^d$, one can regard β as a function from $\mathscr{P}$ to $\mathbb{R}^d$. Estimators of β can often be thought of as extensions of this function from $\mathscr{P}$ to a set $\mathscr{Q}$ containing both $\mathscr{P}$ and all possible empirical distributions $\mathbb{P}_n$. One then uses $\hat{\beta}_n := \beta(\mathbb{P}_n)$ as the estimator. Suppose now that W: $\mathbb{R}^d \times \mathscr{Q} \to \mathbb{R}^d$ is such that $W(\beta(P), P) = 0$ for all $P \in \mathscr{P}$, then it is reasonable to define $\hat{\beta}_n$ by $W(\hat{\beta}_n, \mathbb{P}_n) = 0$. Asymptotic properties of such generalized M-estimators can be studied by applying a one-step Taylor expansion to the implicitly defined functional β. Under conditions (A1)–(A5) of Appendix A, Bickel, Klaassen, Ritov and Wellner (1993) prove such a result.

An M-estimator is defined to be the argument that maximizes $\sum_{i=1}^n l(X_i\colon \beta)$ for some criterion function l of an observation X and parameter β. (When l is the log likelihood the corresponding estimator is just the mle.) If l is differentiable the maximizer $\hat{\beta}$ will be a solution to the equation $\sum_{i=1}^n l_\beta(X_i; \beta) = 0$, where l_β is the partial derivative of l. Replacing l_β by some score function ψ gives a generalized M-estimator. The estimators in this paper go one step further replacing the particular form of estimating equation, $\sum_{i=1}^n \psi(X_i\colon \beta) = 0$ by a general $W(\beta, \mathbb{P}_n) = 0$.

We will consider functions W: $\mathbb{R}^d \times \mathscr{Q} \to \mathbb{R}^d$ such that, whenever P is a member of the Cox model with parameter β_0, $W(\beta_0, P) = 0$. Let β: $\mathscr{Q} \to \mathbb{R}^d$ be implicitly defined by $W(\beta(Q), Q) = 0$ for all $Q \in \mathscr{Q}$. Corresponding to W, we define the estimator to be $\beta(\mathbb{P}_n)$. Restricting attention to functions W that are monotone in β will both ensure that $\beta(Q)$ is well defined and simplify the proof of consistency and asymptotic normality.

NOTATION. K will denote a measurable function from $\mathbb{R}^d \times \mathbb{R}_+ \times \mathscr{Q}$ to $\mathbb{R}^d$, that is, $K(z, t, Q) \in \mathbb{R}^d$. $Y(t)$ will be a 0–1 valued stochastic process. Define

$$S_1(t, \beta, Q) := E_Q[\exp(\beta' Z) Y(t)]$$

and

$$S_K(t, \beta, Q) := E_Q[K(Z, t, Q) \exp(\beta' Z) Y(t)].$$

For now, let $Y(t) = 1_{[T \ge t]}$; later we will consider more general $Y(t)$ to study the case-cohort estimator. Denote the marginal subdistributions of Z and T with $\Delta = 1$ by

$$Q^{(1)}(z) := Q(Z \le z, \Delta = 1) \quad \text{and} \quad Q^{(2)}(t) := Q(T \le t, \Delta = 1)$$

and let $Q^{(u)}(z,t) := Q(Z \le z, T \le t, \Delta = 1)$ and $Q^{(12)}(z,t) := Q(Z \le z, T \le t)$.

Assume that P is the true underlying probability measure. We use the following abbreviated notation:

$$\begin{aligned}
\mathbb{K}_n(z,t) &:= K(z,t,\mathbb{P}_n),\\
S_K(t,\beta) &:= S_K(t,\beta,P),\\
\mathbb{S}_{nK}(t,\beta) &:= \frac{1}{n}\sum_{j=1}^{n} K(Z_j,t,P)\exp(\beta' Z_j)Y_j(t),\\
\mathbb{S}_{n\mathbb{K}_n}(t,\beta) &:= \frac{1}{n}\sum_{j=1}^{n} K(Z_j,t,\mathbb{P}_n)\exp(\beta' Z_j)Y_j(t),\\
S_{\mathbb{K}_n}(t,\beta) &:= E_P[K(Z,t,\mathbb{P}_n)\exp(\beta' Z)Y(t)|\mathbb{P}_n].
\end{aligned}$$

Similar notation will be used with KZ replacing K. We will also, at times, drop the argument β when it is fixed at some value β_0 throughout a calculation.

Finally, let $\Delta_\tau := \Delta 1_{[T \le \tau]}$ and $N(t) := 1_{[T \le t, \Delta = 1]}$.

In order to motivate the new estimators, we first reexamine the mple, which is calculated by solving the score equation

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_i\left\{Z_i - \frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}}(T_i,\beta)\right\} = 0. \tag{2.1}$$

An asymptotically equivalent version of the score equation is given by

$$E\left[\Delta\left(Z - \frac{S_Z}{S_1}(T,\beta)\right)\right] = 0. \tag{2.2}$$

It can be shown algebraically [cf. Sasieni (1992), Lemma 2c] that, when the Cox model holds with parameter β_0,

$$\frac{S_K(t,\beta_0)}{S_1(t,\beta_0)} = E[K(Z,t)|T = t, \Delta = 1]. \tag{2.3}$$

Thus (2.2) holds at $\beta = \beta_0$.

In view of (2.3), one may try to estimate β by solving

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_i\left\{K(Z_i,T_i) - \frac{\sum_{j=1}^{n}K(Z_j,T_i)1_{[T_j \ge T_i]}\exp(\beta' Z_j)}{\sum_{j=1}^{n}1_{[T_j \ge T_i]}\exp(\beta' Z_j)}\right\} = 0, \tag{2.4}$$

where K is some d-dimensional function of Z and t. Such estimators were first proposed by Ritov and Wellner (1988) and have been studied by Sasieni (1989). To obtain a wide variety of estimators, we consider functions K that are data-dependent.

The estimators, $\hat{\beta}_{K,\tau}$, of class K, are defined by

$$W_{K,\tau}(\hat{\beta}_{K,\tau}, \mathbb{P}_n) = 0,$$

for some d-dimensional function K of z, t and Q where

$$\begin{aligned} W_{K,\tau}(\beta, Q) &:= \int_0^\tau \int K(z,t,Q)\, dQ^{(u)}(z,t) - \int_0^\tau \frac{S_K}{S_1}(t,\beta,Q)\, dQ^{(2)}(t) \\ &= E_Q\, \Delta_\tau \left[K(Z,T,Q) - \frac{S_K}{S_1}(T,\beta,Q) \right]. \end{aligned} \tag{2.5}$$

REMARK 1. By permitting K to depend on the measure Q, one can study functions K that are determined by the sample. In that case W is a function of Q both explicitly and through K.

REMARK 2. In practice, one would like to be able to use $\tau = \infty$, but problems occur at the end of the time interval. For this reason many authors define estimators on a finite time interval $[0, \tau]$. Here we provide rigorous proofs for estimators based on a finite time interval and, at the request of a referee, indicate how the results may be extended to cover estimators that use all the data.

EXAMPLES OF K.

EXAMPLE 1. $K(Z,t,Q) = S_Q(t)^\rho(1 - S_Q(t))^\gamma Z$, where $S_Q(t) = \prod_{[0,t)}(1 - dQ^{(2)}(u)/Q\{T \geq u\})$ so that $S_{\mathbb{P}_n}$ is the product limit (Kaplan–Meier) estimator of the survival function of T^u. The case with $(\rho, \gamma) = (1, 0)$ is of special interest since it corresponds to the Peto–Peto generalization of the Wilcoxon statistic. For general (ρ, γ), these estimators correspond to the $G_{(\rho,\gamma)}$ statistics of Fleming and Harrington (1991).

EXAMPLE 2. For univariate Z, consider $K(Z, t, Q) = \min(p_{1-\alpha}(Q), \max(p_\alpha(Q), Z))$, where $p_\alpha(Q)$ is such that $Q\{Z \leq p_\alpha(Q)\} = \alpha$, that is, the 100α percentile of the Q-distribution of Z. "Huberizing" the covariate is designed to decrease the influence of high leverage Z's. Notice, however, that S_1 and S_K still depend on $\exp(\beta' Z)$.

EXAMPLE 3. A variant on Example 2 has $K(Z,t,Q) = \min(p_{1-\alpha}(Q,t), \max(p_\alpha(Q,t), Z))$, where $p_\alpha(Q,t)$ is such that $Q\{Z \leq p_\alpha(Q,t) | T \geq t\} = \alpha$, that is, the 100α percentile of Z among those at risk at time t. Although K in Example 2 is simpler, this choice is natural in that the leverage of an individual failing at some time t depends on the values of the covariates of the other individuals at risk at time t.

EXAMPLE 4. As an alternative to Example 3, one may prefer a strictly monotone transformation of the covariate; the rank transformation offers a certain amount of robustness. $K(z,t,Q) = r(z|t,Q) = Q(Z \le z|T \ge t)$ so that $r(z|t,\mathbb{P}_n) = \sum_{i=1}^n 1_{[Z_i \le z,\ T_i \ge t]}/\sum_{i=1}^n 1_{[T_i \ge t]}$. Estimators based on this K are analogous to the tests proposed by Jones and Crowley (1989, 1990).

CONDITIONS.

(E1) *Moments.*

(i) $E[\exp(\beta' Z)] < \infty$ for all $\beta \in \mathbb{R}^d$.
(ii) There exist functions k: $\mathbb{R}^d \to \mathbb{R}_+$ and w: $\mathbb{R}_+ \to \mathbb{R}_+$, such that $|K(z,t)| \le k(z)w(t)$, with $E[k^m(Z)] < \infty$ for all $m \ge 1$ and $\|w\|_0^\tau < \infty$.

(E2) "*Strong monotonicity*" *of* $\alpha' K$ *as a function of* $\alpha' Z$. Set

$$t^* := \begin{cases} \inf\{t\colon \alpha' Z 1_{[T \ge t]} = 1_{[T \ge t]} \text{ a.s., for any } \alpha \ne 0\}, \\ \infty, \quad \text{if this set is empty.} \end{cases}$$

(i) $Q\{T \le t^*\} > 0$.
(ii) For each $\alpha \in \mathbb{R}^d$ and for all t in a set of Q-probability 1,

$$\alpha'(K(z_2,t) - K(z_1,t))\alpha'(z_2 - z_1) \ge 0.$$

[The domain of $K(\cdot,t)$ is the set of z's that are in the Q-support of (Z,T) intersected with the set $\{T \ge t\}$.]
(iii) The set A^* of t on which $\alpha'(K(z_2,t) - K(z_1,t))\alpha'(z_2 - z_1) > 0$ whenever $\alpha'(z_2 - z_1) \ne 0$ is such that $Q\{T \in A^* \cap [0, t^* \wedge \tau]\} > 0$.

(E3) "*Nondegeneracy*" *of the support of* (Z,T,Δ). For each direction $\alpha \ne 0$, there exists a set A with $Q\{T \in A,\ \Delta_\tau = 1\} > 0$ (A may depend on α) such that

$$\operatorname{ess\,sup}\{\alpha' K(Z,t)\colon T \ge t\} > \operatorname{ess\,inf}\{\alpha' K(Z,t)\colon T = t,\ \Delta = 1\},$$

for all $t \in A$.

REMARK 3. (E3) fails to hold if and only if there is a pair (α, ϕ), $\alpha \ne 0 \in \mathbb{R}^d$ and ϕ: $\mathbb{R} \to \mathbb{R}$ decreasing, such that for each t $(\le \tau)$ in all but a set of $Q^{(2)}$-probability 0,

$$\alpha' K(Z,t)\, dN(t) = \phi(t)\, dN(t) \quad \text{a.s.-}Q$$

and

$$\alpha' K(Z,t) Y(t) \le \phi(t) \quad \text{a.s.-}Q.$$

The following special cases seem to be worth stating explicitly.

LEMMA 2.1 (Univariate Z). *When Z is real valued* ($d = 1$), *conditions* (E2) *and* (E3) *simplify.*

(E2u) *With* t^* *defined as in* (E2):

(i) $Q\{T \le t^*\} > 0$.
(ii) *$K(z,t)$ is monotone in z for each fixed t.*
(iii) *It is strictly monotone on some set A^* with* $Q\{T \in A^* \cap [0, t^* \wedge \tau]\} > 0$.

(E3u) *There does not exist a monotone function* ϕ *such that for* $Q^{(2)}$*-almost all* $t < \tau$,

$$K(Z,t)\,dN(t) = \phi(t)\,dN(t) \quad a.s.\text{-}Q$$

and

$$K(Z,t)Y(t) \le \phi(t) \quad a.s.\text{-}Q$$

(*with the inequality reversed if* ϕ *is increasing.* ϕ *here is increasing if* α *in Remark* 3 *equals* -1).

For multivariate Z ($d > 1$), condition (E2) appears to be very restrictive. We do, however, have the following result.

LEMMA 2.2 [$K(Z,t) = Zw(t)$]. *Suppose* $K(z,t) = w(t)z$ *for some* $w\colon \mathbb{R}_+ \to \mathbb{R}$. *Then* (E2) *is satisfied provided there is no linear combination* $\alpha'Z$, $\alpha \neq 0$, *such that*

$$\alpha' Z w(t) 1_{[T \ge t]} = w(t) 1_{[T \ge t]} \quad a.s.\text{-}Q$$

holds for all t in a set of probability 1. (*It is enough if* $\alpha'Z$ *is not a.s. constant and if there exists* t_0 *with* $Q\{T \le t_0\} > 0$, *such that* $w(t) \neq 0$ *on* $[0, t_0]$.)

COX ESTIMATOR. $K(Z,t) \equiv Z$.

(C1) $E[\exp(\beta' Z)] < \infty$ for all $\beta \in \mathbb{R}^d$.

(C2) There is no pair (α, ϕ), $\alpha \neq 0 \in \mathbb{R}^d$ and $\phi\colon\mathbb{R} \to \mathbb{R}$ decreasing, such that for $Q^{(2)}$-almost all $t < \tau$,

$$\alpha' Z\, dN(t) = \phi(t)\, dN(t) \quad \text{a.s.-}Q$$

and

$$\alpha' Z Y(t) \le \phi(t) \quad \text{a.s.-}Q.$$

LEMMA 2.3. *When* $K(Z,t) \equiv Z$, *conditions* (E1)–(E3) *are implied by* (C1) *and* (C2).

PROOF. (C1) is identical to (E1)(i). Under (C1), by the theory of Laplace transforms, (E1)(ii) holds with $w(t) \equiv 1$ and $k(z) = |z|$. The positive semidefiniteness of (E2) holds without further assumption, and the quadratic form will be nonzero provided there is no linear combination ($\alpha \neq 0$) of the covariates that is $Q^{(1)}$-a.s. constant (i.e., $\Delta_\tau \alpha' Z = \Delta_\tau$ a.s.-Q), and this is implied by (C2) (with $\phi \equiv 1$). Finally, (E3) is implied by (C2)—see Remark 3. □

3. Existence and uniqueness of estimators of class *K*. The estimators studied here are defined implicitly as the roots of certain equations. Before discussing the asymptotic properties of such estimators, one must first establish that they exist and are well defined. In this section results concerning the existence and uniqueness of solutions to the estimating equation $W_{K,\tau}(\beta, Q) = 0$ are presented. Throughout, the function K and the probability measure Q

will be fixed and for this reason the dependence of various quantities upon Q may be dropped from the notation.

The question is: when does $W(\beta(P), P) = 0$ uniquely define a "parameter"? We are interested in the answer both for the population and for a sample. The situation for a finite sample is slightly easier than in the population because expectations are then simply sums and one does not have to worry about their existence. We present a theorem with sufficient conditions on the function K and the measure Q, for W to uniquely define a parameter, and then discuss the special cases when $K(Z, t) = Z$ (the Cox estimator) and when $Q = \mathbb{P}_n$, the empirical probability measure based upon a finite sample. When both $K(Z, t) = Z$ and $Q = \mathbb{P}_n$, these sufficient conditions are identical to the necessary and sufficient conditions given by Jacobsen (1989).

In the Andersen and Gill (1982) generalization of the Cox model, one observes three processes: the counting process of actual failures, the "at risk" or "under observation" process and the covariate process. It seems reasonable that the methods used here could be extended to allow coprocesses $Z_i(\cdot)$ and a general "at risk indicator function" $Y_i(\cdot)$. Certainly the estimators are consistent provided $Z(\cdot)$ and $Y(\cdot)$ are left continuous with right-hand limits. A problem arises in that, unless $Y(\cdot)$ takes a special form, the estimators will be "consistent" for some parameter other than β_0 even when the Cox model holds, that is, the estimators will be asymptotically biased. In the proof of asymptotic normality, we use

$$\|\mathbb{S}_{n1}(\cdot, \beta) - S_1(\cdot, \beta)\|_0^\tau = O_p(n^{-1/2})$$

and this requires further smoothness on $Y(\cdot)$ and $Z(\cdot)$. Rather than become involved in these uniform probability results, we shall restrict attention to time-independent covariates Z_i, and $Y_i(\cdot)$ of the form

$$Y_i(\cdot) = 1_{[T_i \geq \cdot]} B_i,$$

where B_i is a $\{0, 1\}$ random variable that is independent of $X_i := (Z_i, T_i, \Delta_i)$. It is assumed that $(X_1, B_1), (X_2, B_2), \ldots$ are iid and that $\mathrm{Prob}(B = 1) = \alpha > 0$. With minor modifications the proofs of the main theorems can be adapted to the Cox estimator, $K(Z, t) = Z$, even with time-dependent covariates. In that case the moment conditions (E1) would have to be strengthened slightly, but the results certainly hold for $Z(\cdot)$ bounded. This is done for the case-cohort estimator (Corollary 4.2).

THEOREM 3.1 (Existence and uniqueness). *Under conditions* (E1)–(E3), $W_{K,\tau}(\beta, Q) = 0$ *has a unique root* $\beta_{K,\tau}(Q) \in \mathbb{R}^d$.

COROLLARY 3.1. *The Cox functional exists and is unique for any probability measure*, Q, *on* $\mathscr{X}$ *satisfying* (C1) *and* (C2).

PROOF. See Lemma 2.3. □

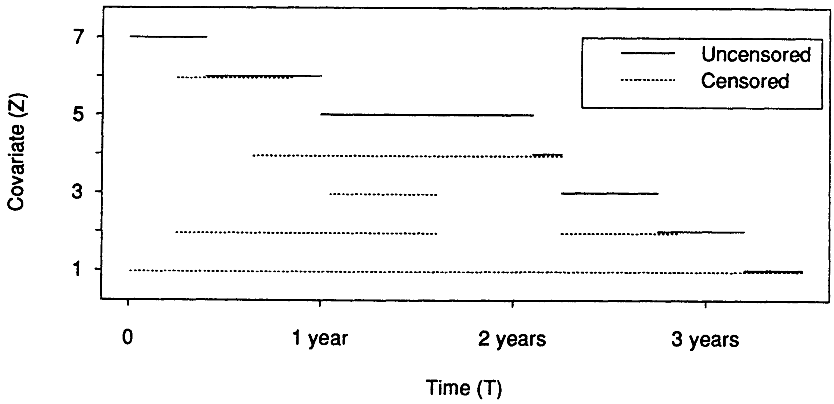

FIG. 1. *Support of* (T, Z): *illustrating when condition* $(C2)$ *fails.*

Thus the Cox functional exists and is unique provided two conditions hold. Condition (C1) states that the Laplace transform (or moment generating function) of Z is finite for all β. If one already knew that the partial likelihood score, W, had a root, then it would be enough to assume that the Laplace transform is finite in some open ball about the root. But in order to show that a root exists we use this stronger condition.

Suppose (C2) fails, so that $\Delta_\tau \alpha' Z = \Delta_\tau \phi(T)$ a.s. and let $(Z_i, T_i) = (z_i, t_i)$ for $i = 1, 2$, then $\alpha' z_1 < \alpha' z_2$ implies that $t_2 < t_1$ a.s. This means that the hazard for an individual with covariate z_2 is, at time t_2, infinitely greater than the hazard for an individual with covariate z_1, for, with probability 1, the individual with z_1 will not fail until $t_1 > t_2$.

Figure 1 illustrates the situation for a discrete univariate Z. The figure plots the support of (T, Z) in the (t, z)-plane. When $\Delta = 1$, the supports of T for each value of Z are nonoverlapping, and if $z_1 < z_2$, then the support of T given $Z = z_2$ is to the left of the support given $Z = z_1$, that is,

$$\sup\{\text{support}(T|Z = z_2, \Delta = 1)\} \leq \inf\{\text{support}(T|Z = z_1, \Delta = 1)\}.$$

This condition is very mild.

COROLLARY 3.2 (Finite sample). *When* $Q = \mathbb{P}_n$, *a probability measure that places mass* $1/n$ *at each of the n observations* (Z_i, T_i, Δ_i) *in* $\mathbb{R}^d \times \mathbb{R}_+ \times \{0, 1\}$, *the estimating equation* $W_{K,\tau}$ *has a unique* 0 *whenever* (E2) *and* (E3n) *hold.*

(E3n) *There is no linear combination* $\alpha' K(Z, t)$, $\alpha \neq 0$, *such that, at each failure time* t_i, *the value of the linear combination for the individual that fails is greater than or equal to the value for all other individuals at risk at* t_i.

PROOF. (E1) is satisfied without further assumptions, since a finite sum of finite terms must be finite. (E3n) is simply a restatement of (E3) for the case $Q = \mathbb{P}_n$. $\square$

REMARK. Combining Corollaries 3.1 and 3.2, one sees that the conditions of Theorem 3.1 reduce to the necessary and sufficient conditions that Jacobsen (1989) gives for the existence and uniqueness of the Cox estimate for a finite sample. Jacobsen shows that the logarithm of the partial likelihood is concave always and strictly concave if and only if

$$\operatorname{span}\{z_j - z_i \colon i \in \{\text{individuals that fail}\}, j \in \{\text{individuals at risk at } T_i\}\} = \mathbb{R}^d.$$

Further, the Cox estimate exists and is unique if and only if there is no $\alpha \neq 0$ such that for all i that fail and for all $j \neq i$ that are at risk at T_i,

$$\alpha'(z_j - z_i) \geq 0.$$

4. Consistency and asymptotic normality. The following theorem gives both consistency and asymptotic normality for estimators based on a finite time interval. Extensions covering estimators that use all the data are discussed following the proof of the theorem in Section 7.

In general, K may depend on the probability measure and it is necessary to introduce conditions that govern the convergence of $\mathbb{K}_n \equiv K(\cdot,\cdot,\mathbb{P}_n)$ to $K \equiv K(\cdot,\cdot,P)$. Basically one would like $\mathbb{K}_n$ to be uniformly asymptotically linear. Some relaxation of the uniformity is possible, but the weaker the conditions the more involved the proof. The following seems to be a fair compromise between generality and complexity.

CONDITION K1. [Recall that $X = (Z, T, \Delta)$.] Either (I) or (II) holds.

(I)

(i) There is a function $\psi_K \equiv \psi_K(x; z, t)$ so that

$$\|\mathbb{K}_n - K - \dot{\mathbb{K}}_n\|_\infty = o_p(n^{-1/2}),$$

where

$$\dot{\mathbb{K}}_n(z,t) = \int \psi_K(x; z, t)\, d(\mathbb{P}_n - P)(x) \quad \text{and} \quad \|f\|_\infty := \sup_{t\in[0,\tau]} \sup_{z\in\mathcal{Z}} |f(z,t)|.$$

(ii) $\|\dot{\mathbb{K}}_n\|_\infty = o_p(1)$ [and hence $\|\mathbb{K}_n - K\|_\infty = o_p(1)$].

(iii) Let $\psi_K^*(x,z) = \sup_{t\in[0,\tau]}|\psi_K(x;z,t)|$. There exists an $\varepsilon > 0$ such that

$$E\left[\psi_K^*(X,Z)^{2+\varepsilon}\right] < \infty \quad \text{and} \quad E\left[\psi_K^*(X_1,Z_2)^{2+\varepsilon}\right] < \infty,$$

where X_1 and Z_2 are independent.

(II) $\mathbb{K}_n(Z,t) = Zw_n(t)$.

(i) There exist functions $w(t)$ and $\psi_w(x,t)$ such that

$$\|w_n - w - \dot{w}_n\|_0^\tau = o_p(n^{-1/2}),$$

where $\dot{w}_n(t) = \int \psi_w(x,t)\, d(\mathbb{P}_n - P)(x)$.

(ii) $\|\dot{w}_n\|_0^\tau = o_p(1)$.

(iii) Let $\psi_w^*(x) = \sup_{t\in[0,\tau]}|\psi_w(x,t)|$. There exists an $\varepsilon > 0$ such that $E[\psi_w^*(X)^{2+\varepsilon}] < \infty$.

Let $\dot{\mathbb{K}}_n(z,t) = z\dot{w}_n(t)$, $\psi_K(x;z,t) = z\psi_w(x,t)$ and $\psi_K^*(x,z) = z\psi_w^*(x)$.

THEOREM 4.1 (Asymptotic normality). *Suppose that* $X_1, \dots, X_n$ *are iid and that conditions* (E1)–(E3) *and* (K1) *hold. If, in addition,* $\inf_{t \in [0,\tau]} P(Y(t) = 1) > 0$ *and* $K(t) := K(Z, t, P)$ *is a random element of* $D[0,\tau]^d$, *then the generalized M-estimator* $\beta_{K,\tau}(\mathbb{P}_n) \equiv \hat{\beta}_K$ *of* $\beta_{K,\tau}(P) \equiv \beta_0$ *corresponding to*

$$W_{K,\tau}(\beta, P) = \int_0^\tau \int K(z,t,P)\, dP^{(u)}(z,t) - \int_0^\tau \frac{S_K}{S_1}(t,\beta,P)\, dP^{(2)}(t)$$

is asymptotically linear with influence function $\tilde{l}_{K,\tau}(X)$. *That is,*

$$\sqrt{n}\left(\hat{\beta}_K - \beta_0\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \tilde{l}_{K,\tau}(X_j) + o_p(1),$$

where

$$\begin{aligned}
\tilde{l}_{K,\tau}(X_j) = \dot{W}_{K,\tau}^{-1}(\beta_0, P)\Bigg\{ & \Delta_{\tau j}\left(K(Z_j,T_j) - \frac{S_K}{S_1}(T_j)\right) \\
& - \int_0^\tau \left(K(Z_j,t) - \frac{S_K}{S_1}(t)\right) e^{\beta_0' Z_j} 1_{[T_j \ge t]} \frac{dP^{(2)}(t)}{S_1(t)} \\
& + \int_0^\tau \int \psi_K(X_j; z,t)\, dP^{(u)}(z,t) \\
& - \int_0^\tau \frac{\int\int \psi_K(X_j; z,t)\, e^{\beta_0' z} 1_{(s \ge t)}\, dP^{(12)}(z,s)}{S_1(t)}\, dP^{(2)}(t) \Bigg\}
\end{aligned}$$

and

$$\dot{W}_{K,\tau}(\beta, P) := \int_0^\tau \left(\frac{S_{KZ}}{S_1} - \frac{S_K}{S_1}\frac{S_Z'}{S_1} \right)(t,\beta,P)\, dP^{(2)}(t).$$

Hence $\sqrt{n}(\hat{\beta}_K - \beta_0) \to_d N(0, E(\tilde{l}_{K,\tau}(X))^{\otimes 2})$ *as* $n \to \infty$.

For the Cox estimator (with time-independent covariates), one has the following corollary, a version of which was first proved (under more restrictive hypotheses) in Section 7.2 of Bickel, Klaassen, Ritov and Wellner (1993).

COROLLARY 4.1. *If the data are iid and conditions* (C1) *and* (C2) *hold, then the Cox estimator, based on an interval* $[0,\tau]$ [*with* $P(T \ge \tau) > 0$], *is asymptotically normal.*

PROOF. By Lemma 2.3, (C1) and (C2) imply (E1)–(E3). (K1) is vacuous for the Cox estimator. The result is thus an immediate consequence of the theorem. □

The theorem shows that the estimators in the class W_K are asymptotically linear and gives the influence function $\tilde{l}_{K,\tau}(X)$. The formula for the influence function may look horrendous, but it is not so different from the more familiar expression for the mple.

Define

$$M_j(t|Q) = N_j(t) - \int e^{\beta' Z_j} 1_{[T_j \geq t]} \frac{dQ^{(2)}(t)}{S_1(t, \beta(Q), Q)}$$

and let $\Psi_K(X) = \dot{W}_{K,\tau}(\beta_0, P)\tilde{l}_{K,\tau}(X)$. The part of Ψ_K not involving the gradient of K can be written as

$$\int_0^\tau \left(K(Z_j, t) - \frac{S_K}{S_1}(t) \right) M_j(dt|P).$$

The remaining terms are

$$E\left[\int_0^\tau \psi_K(X_j; Z_i, t) M_i(dt|P) \,|X_j \right].$$

(Here X_i is independent of X_j.) If the Cox model holds, M_j is a martingale: in general it is not. When it is a martingale the conditional expectation is identically 0 and the influence function does not depend on the gradient ψ_K.

EXAMPLE 1 (Continued). Suppose that $\mathbb{K}_n(z, t) = z\overline{\mathbb{P}}_n^{(2)}(t)$, where $n\overline{\mathbb{P}}_n^{(2)}(t)$ is the number of individuals still at risk at time t. This particular choice of K yields an estimator with an influence function that is bounded in t both on and off the Cox model. Here we merely point out that conditions (E2) and (K1) do indeed hold for this choice of K and hence, by Theorem 4.1, the estimator is asymptotically normal. We have

$$(\mathbb{K}_n - K)(z, t) = z\int_0^\infty 1_{[s \geq t]}\, d\big(\mathbb{P}_n^{(2)} - P^{(2)}\big)(s),$$

that is, $\psi_K(X, z, t) = z1_{[T \geq t]}$, where $X = (Z, T, \Delta)$. Condition (K1)(II) is easily seen to hold by Donsker's theorem. Lemma 2.2 simplifies condition (E2) for K's that are merely time-weighted versions of Z.

The estimator with weights proportional to the number of individuals at risk corresponds to Gehan's (1965) version of the Wilcoxon test statistic. It has the disadvantage that the asymptotic distribution of the statistic depends on the censoring distribution. In the presence of censoring one would rather use weights analogous to those proposed by Peto and Peto (1972) to generalize the Wilcoxon test. The simplest way to produce a similar weighting that is asymptotically independent of censoring is to use the Kaplan–Meier estimator of the survival function for all individuals (i.e., pooled over the covariates). Once again condition (K1) holds—asymptotic linearity of the Kaplan–Meier estimator has been shown by Breslow and Crowley (1974).

Rather than use a pooled estimator of the survival function, one might like to use the usual estimator of the baseline survival function (i.e., the product integral of the Breslow estimator of the cumulative hazard).

EXAMPLE 4 (Continued). As a second example illustrating the conditions on K, consider $K(z,t,Q) = Q(Z \le z | T \ge t)$ for a single covariate Z. Here $\mathbb{K}_n(z,t) = \sum_{i=1}^n 1_{[Z_i \le z,\ T_i \ge t]} / \sum_i 1_{[T_i \ge t]}$. For each t and any Q, $K(z,t,Q)$ is a monotone function of z. Hence (E2) holds by Lemma 2.1. It remains to check condition (K1). Denote by A and B the events $\{Z \le z,\ T \ge t\}$ and $\{T \ge t\}$, respectively. Then

$$\begin{aligned} K(z,t,P) - \mathbb{K}_n(z,t) &= \frac{\mathbb{P}_n A}{\mathbb{P}_n B} - \frac{PA}{PB} \\ &= \frac{(\mathbb{P}_n - P)A}{PB} - \frac{PA}{PB}\frac{(\mathbb{P}_n - P)B}{PB} \\ &\quad + \left(\frac{(\mathbb{P}_n - P)A}{PB} - \frac{PA}{PB}\frac{(\mathbb{P}_n - P)B}{PB} \right) \frac{(\mathbb{P}_n - P)B}{\mathbb{P}_n B}. \end{aligned}$$

The third term in the expansion is a remainder which will be uniformly negligible when $PB \ge \varepsilon > 0$ as is required by the condition $\inf_{[0,\tau]} P\{Y(t) = 1\} > 0$ in Theorem 4.1. Note that under this condition K is bounded and has bounded gradient

$$\psi_K(X; z, t) = \frac{1_{[Z \le z,\, T \ge t]}}{P\{T \ge t\}} - P(Z \le z | T \ge t) \frac{1_{[T \ge t]}}{P\{T \ge t\}}.$$

REMARK. Weighted log rank statistics are of the form $W_K(0, \mathbb{P}_n)$. For example, $K = Z$ gives the log rank statistic and $K = \hat{S}(t)Z$ gives the Peto–Peto–Prentice censored data Wilcoxon statistic, where $\hat{S}(t)$ is the Kaplan–Meier estimator of the marginal (pooled) survival function. $K(z,t) = P(Z \le z | T \ge t)$ (as in Example 4) gives the Jones and Crowley robust log rank regression statistic. In the proof of Theorem 4.1, we show that $\sqrt{n}\, W_K(\beta_0, \mathbb{P}_n)$ is asymptotically linear. There is no part of the proof that uses the special nature of β_0 [that $W_K(\beta_0, P) = 0$] and the argument works for any fixed β. (In fact, the proof simplifies somewhat for $\beta = 0$.) Thus, as a corollary to the proof of Theorem 4.1, we have

$$\sqrt{n}\,(W_K(0, \mathbb{P}_n) - W_K(0, P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_K(X_i) + o_p(1).$$

Case-cohort design. Recall that the indicator $Y(t)$ may be more general than $1_{[T \ge t]}$. The reason for working in this much generality is to derive results for the "case-cohort design." This design was proposed by Prentice (1986) for use in large epidemiological or disease prevention studies in which the proportion of individuals, who will develop the disease end point (i.e., fail) within the follow-up period, is small.

The case-cohort design involves collecting "raw covariate" data on all subjects, but processing these data for only a small subsample. This subsample will consist of all the cases (observed failures) together with a cohort identified at the beginning of the study.

Let $Y_i(t) = 1_{[T_i \geq t]} B_i$, where B_i is a random variable taking values, 1, with probability $\alpha > 0$, and 0 otherwise, which is independent of $X_i = (Z_i, T_i, \Delta_i)$.

COROLLARY 4.2 (Case-cohort design). *Suppose that the data are iid from a member of the Cox model with time-dependent covariates and parameter* β_0 *and that* $P(T \geq \tau) > 0$ *and the* $Z_i(\cdot)$ *are almost surely bounded elements of* $D[0, \tau]^d$ *such that*

(C2*) *There is no pair* (α, ϕ), $\alpha \neq 0 \in \mathbb{R}^d$ *and* $\phi: \mathbb{R} \to \mathbb{R}$ *decreasing, such that, for* $Q^{(2)}$*-almost all* $t < \tau$,

$$\alpha' Z(t)\, dN(t) = \phi(t)\, dN(t) \quad a.s.\text{-}Q$$

and

$$\alpha' Z(t) Y(t) \leq \phi(t) \quad a.s.\text{-}Q.$$

Then the case-cohort estimator $\hat{\beta}$ *of* β_0 *is consistent and asymptotically normally distributed. That is, the solution,* $\hat{\beta}$, *of*

$$\sum_{j=1}^{n} \Delta_{\tau j} \left\{ Z_j(T_j) - \frac{\sum_{i=1}^{n} B_i 1_{[T_i \geq T_j]} Z_i(T_j)\, e^{\beta' Z_i(T_j)}}{\sum_{i=1}^{n} B_i 1_{[T_i \geq T_j]}\, e^{\beta' Z_i(T_j)}} \right\} = 0$$

satisfies

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \to_d N(0, \Sigma_\alpha) \quad as\ n \to \infty,$$

where $\Sigma_\alpha = \Sigma^{-1}(\Sigma + R_\alpha)\Sigma^{-1}$, $\Sigma := E \operatorname{Var}(\Delta_\tau Z(T) | T, \Delta)$ (*the information for the Cox estimator*),

$$R_\alpha := 2(1-\alpha)\alpha^{-1} \int_0^\tau \int_0^t E\big[(Z(t) - e(t))(Z(u) - e(u))' \times 1_{[T \geq t]} e^{\beta_0'(Z(t) + Z(u))}\big]\, d\Lambda_0(u)\, d\Lambda_0(t)$$

and $e(t) := E[Z(t) | T = t, \Delta = 1] = S_Z / S_1(t, \beta_0)$.

Note that this result agrees with (5.2) of Self and Prentice (1988). [Their Δ is equivalent to our R_α and their $Y(t)$ is our $1_{[T \geq t]}$.] We have removed their tightness condition (7), at the expense of requiring the covariate process to be external, so that the conditional hazard given the whole process $Z = \{Z(t): t \geq 0\}$ satisfies $\lambda(t|Z) = \lambda_0(t) \exp(\beta' Z(t))$.

Consistency of the covariance estimator. Discussion is limited to estimating the asymptotic covariance of estimators $\hat{\beta}_K$ for which K is not a function of P, that is, $\mathbb{K}_n \equiv K$. Estimation in the more general case is analogous under appropriate restrictions on the class of K. The estimator proposed here converges both on and off the Cox model to the appropriate covariance matrix. Thus it may be preferable to the usual estimator of covariance of the mple which is appropriate only when the model holds. The asymptotic variances of the variance estimators have not been considered, but intuitively the estimator discussed in this section will be less efficient than the usual one. [See Lin and

Wei (1989) for use of this estimator.] Let

$$\Psi_K(X_i,\beta;Q) = \Delta_{\tau i}K(Z_i,T_i) - \Delta_{\tau i}\frac{S_K}{S_1}(T_i,\beta;Q)$$
$$-\int_0^\tau \left\{K(Z_i,t) - \frac{S_K}{S_1}(t,\beta;Q)\right\}\frac{e^{\beta' Z_i}Y_i(t)}{S_1(t,\beta;Q)}\,dQ^{(2)}(t)$$

and $\hat{\mathscr{D}}_K := \hat{\Sigma}_{nKZ}^{-1}\int \Psi_K(x,\hat{\beta}_K,\mathbb{P}_n)^{\otimes 2}\,d\mathbb{P}_n(x)(\hat{\Sigma}_{nKZ}^{-1})'$, where $\hat{\Sigma}_{nKZ} = \dot{W}_{K,\tau}(\hat{\beta}_K,\mathbb{P}_n)$.

PROPOSITION 4.1. *Suppose that $\mathbb{K}_n \equiv K$. Under the conditions of Theorem 4.1,*

$$\hat{\mathscr{D}}_K \to_p E\left[\tilde{l}_{K,\tau}(x)^{\otimes 2}\right]$$
$$= \Sigma_{KZ}^{-1}E\left[\Psi_K(X,\beta_K(P);P)^{\otimes 2}\right](\Sigma_{KZ}^{-1})'.$$

PROOF. See Sasieni (1989), Proposition 9.2.2.

5. Estimating β under contiguous alternatives to the Cox model. In this section we use Le Cam's third lemma to compare various estimators of the regression parameter β in the proportional hazards model under contiguous alternatives. It is assumed that censoring is conditionally independent of failure and that the covariates are time-independent.

On the Cox model, the influence functions of estimators of class K are given by

$$\Psi_K = I_K^{-1}\left\{\Delta(K(Z,T) - e_K(T)) - \int_0^T (K(Z,t) - e_K(t))\exp(\beta_0' Z)\,d\Lambda_0(t)\right\},$$

where

$$e_K(t) = S_K/S_1(t,\beta_0) = E[K(Z,t)|T = t,\ \Delta = 1]$$

and

$$I_K = \dot{W}(\beta_0,P) = E\,\mathrm{Cov}(\Delta Z, \Delta K(Z,T)|T,\Delta).$$

Contiguity theory tells us that if the log likelihood ratio ($\log L_n$) is asymptotically linear with direction $\phi(X)$ and if $\hat{\beta}_n$ is an asymptotically linear estimator with influence function Ψ, then, under the contiguous alternative,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \to_d N(\mu, V),$$

where $\mu = E[\phi(X)\Psi(X)]$ and V is the asymptotic variance of $\hat{\beta}_n$ under the null (i.e., on the model).

For example, suppose that under

$$P_n\colon X_{n1},\dots,X_{nn} \text{ are iid with density } f,$$

and under

$$Q_n\colon X_{n1},\dots,X_{nn} \text{ are iid with density } f_n$$

such that

$$\left\|\sqrt{n}\left(f_n^{1/2}-f^{1/2}\right)-\tfrac{1}{2}\phi f^{1/2}\right\|_2 \to 0 \quad \text{as } n\to\infty,$$

where $\|\cdot\|_2$ is the L_2-norm (with respect to the dominating measure). Then

$$\log L_n = n^{-1/2}\sum_{i=1}^{n}\phi(X_i) - \frac{1}{2}E_F\phi(X) + o_{P_n}(1).$$

EXAMPLE 5. If the conditional hazard under Q_n is

$$\lambda_n(t|Z) = \lambda_0(t)\exp(\beta'Z + n^{-1/2}\nu(Z,t)),$$

then $\phi(Z,T,\Delta) = \int\nu(Z,t)\,dM(t)$, where $M(t) = N(t) - \int_0^t 1_{[T\ge u]}\lambda(u|Z)\,du$, the counting process martingale

$$E[\phi\Psi_K] = I_K^{-1}E\left[\int\nu(Z,t)\,dM(t)\int(K-e_K)\,dM(t)\right]$$

$$= I_K^{-1}E[\operatorname{Cov}(\Delta\nu(Z,T),\Delta K(Z,T)|T,\Delta)] \quad \text{by Sasieni (1992), Lemma 1.}$$

Notice that when

$$E[\operatorname{Cov}(\Delta\nu(Z,T),\Delta K(Z,T)|T,\Delta)] = 0$$

the asymptotic shift under the contiguous alternatives Q_n is 0.

EXAMPLE 6. If $\lambda_n(t|Z) = (1-n^{-1/2})\lambda(t|Z) + n^{-1/2}h(t,Z)$, where $\lambda(t|Z) = \lambda_0(t)e^{\beta'Z}$, then the direction $\phi = (h-\lambda)/\lambda - (H-\Lambda)$, where $H(t,Z) = \int_0^t h(s,Z)\,ds$. That is,

$$\phi(T,Z,\Delta) = \Delta\left(\frac{h(T,Z)-\lambda(T|Z)}{\lambda(T|Z)}\right) - \int_0^T\left(\frac{h(t,Z)-\lambda(t|Z)}{\lambda(t|Z)}\right)\lambda(t|Z)\,dt$$

$$= \int\frac{h-\lambda}{\lambda}dM.$$

So the asymptotic bias is

$$E[\phi\Psi_K] = I_K^{-1}E\operatorname{Cov}\left(\Delta\frac{h-\lambda}{\lambda},\Delta K|T,\Delta\right).$$

EXAMPLE 7 (ε-contamination). Let $f(z,t,\delta)$ denote a density from the Cox model, and $g(z,t,\delta)$ an arbitrary density on $\mathbb{R}^d\times[0,\infty)\times\{0,1\}$, and let the density $f_n = (1-n^{-1/2})f + n^{-1/2}g$. Let $h = g/\overline{G}$ be the hazard corresponding to g, and H the cumulative hazard. Then the log likelihood ratio is asymptotically linear with

$$\phi = \frac{h^\delta\exp(-H) - \lambda^\delta\exp(-\Lambda)}{\lambda^\delta\exp(-\Lambda)} = \left(\frac{h}{\lambda}\right)^\delta\exp(-(H-\Lambda)) - 1.$$

Thus the bias

$$E[\phi\Psi_K] = E\left[\left(\frac{h}{\lambda}\right)^{\delta} \exp(-(H-\Lambda)) \int (K - e_K)\, dM\right] = E_G[\Psi_K].$$

(Of course, it is always true that the shift under "contamination" alternatives is equal to the expected value of the influence function of the estimator under the contaminating distribution.)

Notice that the asymptotic bias on a local contamination neighborhood of the Cox model will be bounded if and only if the influence function is bounded. The influence function of the mple is unbounded both in the covariate (if the covariate space is not compact) and in time (if the estimator uses all the data). This was first noted by Samuels (1978) and explains the results of Bednarski (1989). In terms of Example 7, Bednarski considers a sequence of contaminating densities g_n and shows that, if the marginal distribution of the covariate in the contaminating distributions has support that grows like $\log n$ and if the conditional distribution of the failure time given the covariate in the contaminating distribution does not follow the Cox model, then the bias in estimation of β can be of any magnitude. His arguments may be extended to consider contamination neighborhoods smaller than $\sqrt{n}$ and show that provided the contaminating covariate values are large enough arbitrary large bias can be caused by extremely small amounts of contamination.

Let $X = (Z, T, \Delta)$. If we were dealing with a one-dimensional parametric model for β (e.g., if $d = 1$ and the baseline hazard is known), then standard theory [Bickel (1981)] gives the influence function of an estimator with minimum variance on the model subject to a bound on the bias on an infinitesimal contamination neighborhood to be

$$\psi(X) = \left[\alpha_0 + \alpha_i \dot{l}_\beta(X)\right]_{m_1}^{m_2},$$

where $\dot{l}_\beta(X)$ is the score for β, α_0 and $\alpha_1 > 0$ are constants (that may depend on β) chosen such that $E\psi(X) = 0$ and $[u]_a^b = \min(\max(a, u), b)$. (The bias is bounded between m_1 and m_2 where $-\infty \le m_1 \le 0 \le m_2 \le \infty$.) In the Cox model, however, even when we consider only a single covariate, there is the unknown baseline hazard to contend with. As is well known from the theory of semiparametric models [e.g., Bickel, Klaassen, Ritov and Wellner (1993)] the influence function of any estimator of β must be orthogonal to the tangent space for the baseline hazard function. Solving the variational problem for ψ in the presence of the nuisance function yields

$$\psi(X) = \left[\alpha_0 + \alpha_1\left(\dot{l}_\beta(X) - \dot{l}_\lambda a^*(X)\right)\right]_{m_1}^{m_2}, \tag{5.1}$$

where $\dot{l}_\lambda$ is the score operator for the baseline hazard and a^* is some function of T (with finite second moment) chosen such that $E\psi(X) = 0$ and $\psi(X)$ is orthogonal to $\dot{\mathscr{P}}_\lambda = \{\dot{l}_\lambda a\colon a \in L_2(P_T)\}$, the tangent space for λ_0. Sasieni (1992) shows that

$$\dot{l}_\beta(X) = \int Z\,dM(t) \quad \text{and} \quad \dot{l}_\lambda a(X) = \int a(t)\,dM(t),$$

and that, on the model, the efficient score for β is

$$l_\beta^*(X) = \int \{Z - E[Z|T = t, \Delta = 1]\}\,dM(t).$$

It does not appear to be a simple matter to find α_0, α_1 and $a^*(\cdot)$ in (5.1) so that $E\psi(X) = 0$ and $E[\psi(X)\dot{l}_\lambda a(X)] = 0$ for all $a \in L_2(P_T)$. One possible solution is to fix $a^*(t) = E[Z|T = 1, \Delta = 1]$ and to find α_0 and α_1 such that

$$\psi(X) = \left[\alpha_0 + \alpha_1 l_\beta^*(X)\right]_{m_1}^{m_2} \tag{5.2}$$

has mean 0. The interpretation that we would like to give to this solution is that it solves the variational problem in the least favorable parametric submodel. That is in the model $\mathscr{P}_0 = \{P_{\beta,\lambda(\beta)}: \beta \in \mathbb{R}\}$, where $P_{\beta,\lambda}$ is a probability measure from the Cox model with parameters β and λ, and $\lambda(\beta)$ is the least favorable one-dimensional parametrization of the baseline hazard function. Unfortunately, we have been unable to exhibit such a curve $\lambda(\beta)$. What we do know is that the least favorable direction is given by

$$\frac{d\lambda(t)}{d\beta} = E[Z|T = t, \Delta = 1]\lambda(t)$$

(note that the conditional expectation will depend on β, too). Hence an influence function of the form (5.2) will solve the variational problem for the locally least favorable submodel. Calculating α_0 and α_1 given m_1 and m_2 is still not straightforward. The constants will depend on β and the marginal distribution of Z, and, in the presence of censoring, on the censoring distribution (relative to the baseline hazard) via $dC(t|z)/dG(t)$, where $C(t|z)$ is the conditional censoring distribution function and $G(t) = 1 - \exp(-\Lambda_0(t))$ is the baseline failure time distribution.

Suppose that we have found α_0 and α_1 to solve the parametric problem. Since the influence function of any estimator must be orthogonal to the tangent space for the nuisance part of the model, it is still necessary to project the influence function from the parametric submodel so that it will be orthogonal to $\dot{\mathscr{P}}_\lambda$ and, in the presence of censoring, so that the conditional expectation of the influence function given the censoring time is 0.

Consider the simpler problem in which the contaminating distribution has compact Z-support. In that case, since we are mostly concerned with influential times, it seems reasonable to consider the class of maximum weighted partial likelihood estimators. Weight functions of the form $w(t) = 1$ if $m_1 \le \Lambda_0(t) \le m_2$, $w(t) = (1 - m_2)/(1 - \Lambda_0(t))$ if $\Lambda_0(t) > m_2$, $w(t) = (1 - m_1)/(1 - \Lambda_0(t))$ if $\Lambda_0(T) < m_1$, with $0 \le m_1 < 1 < m_2$, approximate the ideal of leaving noninfluential points unchanged and setting the influence function at other points to be maximum permitted value. This is analogous to "Huberizing" the standard residuals [rather than the influence residuals (or delta-β's)] in regular linear regression. In practice, one would want to center the covariates so that the baseline hazard is the hazard of an "average individual" and then choose $m_1 \sim 0$ and $m_2 > 2$. [Note that the survival function corresponding to the baseline hazard is equal to $\exp(-\Lambda_0)$ and that $e^{-3} \sim 0.05$.]

6. Illustration. Serum $\beta 2$ microglobulin (s-β2m) has been shown to be extremely important in predicting the survival of patients with myelomatosis. It has also been noted [e.g., Cuzick, Cooper and MacLennan (1985)] that s-β2m has less prognostic value in long-term follow-up than in the period immediately after diagnosis. The data come from the Medical Research Council's fifth trial on myelomatosis. The s-β2m levels of 553 patients at the time of diagnosis are related to their subsequent survival. The survival times of 123 patients are censored. Survival is recorded in months and for simplicity we have separated ties randomly. The range of log(s-β2m) is $(0.8, 4.4)$ (10th and 90th percentile 1.3 and 3.2, respectively).

Application of the Cox model to all the data gives the mple of the coefficient of log(s-β2m) to be 0.59 with estimated standard error 0.07. If the model is fit only to the subsequent survival of those individuals who survive at least 18 months ($n = 346$, 224 uncensored), then the estimates become 0.23 ± 0.10. Whereas if one artificially censors all survival times at 30 and 18 months, the coefficients are 0.78 ± 0.08 and 0.94 ± 0.09, respectively. Thus, as noted by previous authors, the effect of s-β2m on the hazard decreases with time since measurement, but it does have a large and highly significant influence on early survival. (With $\beta = 0.75$ the relative risk is 15 and 4 over the full and interdecile range, respectively.)

Table 1 summarizes the results of estimating the regression coefficient using a variety of weight functions. The estimates together with an estimate of their standard error based on Proposition 4.1, and where appropriate, the Cox estimate of their standard error are presented. In estimating the standard error we have ignored the increase in variability that one would expect (when the Cox model does not hold) from the randomness of the weight functions. That is, the standard errors are conditional on the actual (data dependent) weights used.

Notice, in particular, that use of Wilcoxon-type weights and artificial censoring gives estimates that are larger, and that the standard error of the Wilcoxon-type estimator compares favorably with that of artificially censored ones. Notice also that use of $K(z,t) = k(z)w(t)$ with $k(z) = \min(\max(1.3, z), 3.2)$ gives slightly smaller estimates than the corresponding weights with $k(z) = z$, but that the standard errors are considerably smaller. The asymptotic relative efficiency of the Wilcoxon estimator on the Cox model is discussed in Sasieni (1993).

Consider how these estimators might be used in practice. Suppose, for instance, one merely estimated the regression coefficient using the mple and the Wilcoxon estimator. Observing the difference in the standard error estimate for the mple depending on whether or not one assumes the Cox model

TABLE 1
Estimated coefficient and standard error of $\log(s\text{-}\beta 2m)$.
$w_1(t)$ *is the number at risk at time t,* $w_2(t)$ *is the Kaplan–Meier estimate of (pooled) survival at t and* $k(z) = \min(\max(z, 1.3), 3.2)$

$K(Z, t)$	Coefficient	Standard error	Cox standard error
Z	0.59	0.080	0.067
$Z1_{[t>12]}$	0.29	0.098	0.092
$Z1_{[t>18]}$	0.23	0.100	0.101
$Z1_{[t\le 30]}$	0.78	0.081	0.080
$Z1_{[t\le 18]}$	0.94	0.082	0.092
$Zw_1(t)$	0.74	0.078	—
$Zw_2(t)$	0.73	0.078	—
$Zw_2^2(t)$	0.84	0.079	—
$Zw_2(t)(1 - w_2(t))$	0.47	0.085	—
$k(Z)$	0.57	0.075	—
$k(Z)w_1(t)$	0.72	0.068	—
$k(Z)w_2^2(t)$	0.82	0.066	—

holds (0.80 versus 0.67) already indicates that the model may not hold. The difference between the mple and the Wilcoxon estimator is further indication that the model may not hold and suggests that the effect of s-β2m on survival is less strong in long-term survivors. (One could perform a formal test by looking at the joint distribution of the two estimators under the null hypothesis that the model holds. Using Theorem 4.1 this is quite straightforward.) Further investigation shows that s-β2m has a strong influence on short-term survival, but that its prognostic value decreases rapidly so that the relative risk associated with a unit increment in s-β2m is 2.6 during the first 18 months, but only 1.3 thereafter.

One possible explanation of a decreasing prognostic value is that there is an unobserved factor (often called a frailty). We have considered s-β2m levels at presentation, but data are also available on s-β2m levels at follow-up visits. When available, subsequent s-β2m levels again provide strong prognostic value for the next 18 months. This is important for understanding the progression of myeloma, for it suggests that the disease may still be reversible even after high levels of s-β2m are encountered.

In evaluating a new treatment, however, one will not want to use follow-up measurements because adjustment for such time-dependent variables may obscure the effect of treatment. (The treatment could be completely aliased if it not only reduces mortality but also reduces s-β2m levels.) For these reasons it is important to evaluate the effect of the presentation s-β2m level on long-term survival.

7. Proofs. The following lemmas will be used in the proofs of the main results.

LEMMA 7.1. *Suppose that* (E2) *holds. Then*

$$\beta' K(z^*(\beta, t), t) \geq \text{ess sup}\{\beta' K(Z, t): T \geq t\},$$

for all directions β *and for each* t *in all but a set of* Q*-measure* 0, *where* $z^*(\beta, t)$ *is such that* $\beta' z^*(\beta, t) = \text{ess sup}\{\beta' Z\colon T \geq t\}$.

PROOF. By (E2), $\beta'[K(z^*(\beta, t), t) - K(z, t)][z^*(\beta, t) - z]'\beta \geq 0$ a.s. But, by definition of z^*, $\beta'(z^*(\beta, t) - z) \geq 0$. □

LEMMA 7.2. *The moment conditions* (E1) *are sufficient for* $W_{K,\tau}(\beta, Q)$ *to be well defined for all* β. *Further*, $W_{K,\tau}(\beta, Q)$ *is continuously differentiable in* β, *and*

$$\Sigma_{KZ}(\beta, \tau) = -\frac{\partial}{\partial\beta} W_{K,\tau}(\beta, Q) = E_Q\left[\Delta_\tau\left(\frac{S_{KZ}}{S_1}(T, \beta) - \frac{S_K S_Z'}{S_1^2}(T, \beta)\right)\right].$$

PROOF. By the theory of Laplace transforms [see, e.g., Barndorff-Nielsen (1978)], (E1)(i) implies that $S_1(t, \beta) := E[Y(t)\exp(\beta' Z)]$ is, for each t, infinitely differentiable at each β, and that the derivatives can be evaluated by differentiating inside the expectation. Similarly (E1)(ii) implies that the same is true of $S_K(t, \beta)$. Thus, in particular,

$$\frac{\partial}{\partial\beta} S_K(t, \beta) = S_{KZ}(t, \beta)$$

and $|S_{KZ}(t, \beta)| < \infty$, for each t. Hence

$$\frac{\partial}{\partial\beta}\left(\frac{S_K}{S_1}(t, \beta)\right) = \frac{S_{KZ}}{S_1}(t, \beta) - \frac{S_K S_Z'}{S_1^2}(t, \beta) \tag{7.1}$$

and if the right-hand side of (7.1) is dominated in an open ball about each β by an integrable function of t, then

$$\frac{\partial}{\partial\beta} W_{K,\tau}(\beta, Q) = -E\left[\Delta_\tau\left(\frac{S_{KZ}}{S_1}(T, \beta) - \frac{S_K S_Z'}{S_1^2}(T, \beta)\right)\right]. \tag{7.2}$$

We now show that (E1) yields the required dominating function.

$$\left|\frac{S_{KZ}}{S_1}(t, \beta) - \frac{S_K S_Z'}{S_1^2}(t, \beta)\right| \leq \left|\frac{S_{KZ}}{S_1}(t, \beta)\right| + \left|\frac{S_K}{S_1}(t, \beta)\right|\left|\frac{S_Z}{S_1}(t, \beta)\right|.$$

Applying Hölder's inequality with $q > 1$ and $1/q + 1/r = 1$, gives

$$S_1(t, \beta) := E[Y(t)\exp(\beta' Z)] \geq \frac{(E[Y(t)])^{\eta+1}}{(E[\exp(-\beta' Z/\eta)])^{\eta}} \quad \text{for all } \eta > 0, \tag{7.3}$$

$$\begin{aligned} |S_{KZ}(t, \beta)| &:= |E[Y(t)K(Z, t)Z'\exp(\beta' Z)]| \\ &\leq E[Y(t)k(Z)w(t)|Z|\exp(\beta' Z)] \quad \text{by (E1)(ii)} \\ &\leq w(t)(E[Y(t)])^{1/q}(E[|k(Z)|^r |Z|^r \exp(r\beta' Z)])^{1/r}. \end{aligned} \tag{7.4}$$

[To see (7.3), put $U = Y(t)\exp(\beta'Z/(\eta+1))$, $V = \exp(-\beta'Z/(\eta+1))$ and $q = \eta + 1$ in Hölder's inequality.] Thus with $\overline{Q}_T(t-) := E[1_{[T \geq t]}] := \overline{Q}_{T-}(t) = \alpha^{-1}E[Y(t)]$ (recall the definition of Y and α in Section 3):

$$\left|\frac{S_{KZ}}{S_1}(t,\beta)\right| \leq w(t)\left(\alpha\overline{Q}_{T-}(t)\right)^{1/q-(\eta+1)} \times\left(E\left[k(Z)^r|Z|^r e^{r\beta'Z}\right]\right)^{1/r}\left(E\left[e^{-\beta'Z/\eta}\right]\right)^{\eta}. \tag{7.5}$$

Now by taking a compact (nontrivial) neighborhood $\mathscr{B}$ of β:

$$\sup_{\mathscr{B}} E[\exp(-\beta'Z/\eta)] = E[\exp(-\beta'_*Z/\eta)] \quad \text{for some } \beta_* \in \mathscr{B} \\ < \infty \quad \text{by (E1)(i)}, \tag{7.6}$$

$$\begin{aligned} &\sup_{\mathscr{B}} E\left[k(Z)^r|Z|^r \exp(r\beta'Z)\right] \\ &\quad = E\left[k(Z)^r|Z|^r \exp(r\beta'_*Z)\right] \quad \text{for some } \beta_* \in \mathscr{B} \\ &\quad \leq \left(E\left[k(Z)^{2r}\right]E\left[|Z|^{2r}\exp(2r\beta'_*Z)\right]\right)^{1/2} \\ &\quad < \infty \quad \text{by (E1)}, \end{aligned} \tag{7.7}$$

since the expectations on the left-hand side of (7.6) and (7.7) are continuous functions of β. Applying similar arguments to S_K and S_Z gives

$$\left|\frac{S_{KZ}}{S_1}(t,\beta) - \frac{S_K S'_Z}{S_1^2}(t,\beta)\right| \leq Aw(t)\overline{Q}_T(t-)^{-\gamma} \quad \text{for all } \beta \in \mathscr{B}, \tag{7.8}$$

where A is a constant and $0 < \gamma < 1$ by appropriate choice of q and η in (7.5) (in particular, $1/q > \eta$. By choosing q close to 1 and η close to 0, γ will be arbitrarily close to 0; the resulting A will then be very large, but that is of no consequence.) Hence, since w is bounded, it is enough to show that $E[\Delta_\tau \overline{Q}_T^{-\gamma}(T-)] < \infty$:

$$\begin{aligned} E\left[\Delta_\tau \overline{Q}_T^{-\gamma}(T-)\right] &\leq \int_0^\infty \overline{Q}_T^{-\gamma}(t-)\,dQ_T(t) \\ &= -\int_1^0 \left[\overline{Q}_{T-}(\overline{Q}_T^{-1}(x))\right]^{-\gamma} dx \\ &\leq \int_0^1 x^{-\gamma}\,dx = \frac{1}{1-\gamma} < \infty, \end{aligned}$$

since for any survival function $\overline{F}$ (right continuous) $\overline{F}_-(\overline{F}^{-1}(t)) \geq t$ and $1 > \gamma > 0$. (The first equality is from the change of variable theorem.) Thus (7.2) holds. □

LEMMA 7.3. *Suppose that* (E1) *and* (E2) *hold. Then* $\partial W_{K,\tau}(\beta, Q)/\partial\beta$ *is negative definite, and hence* $W_{K,\tau}(\beta, Q)$ *is strictly monotone.*

PROOF. Let $Q^{(12)}$ denote the marginal distribution of (Z, T, B). For each t and β, define $\nu = \nu_{t,\beta}$ to be a probability measure such that

$$\frac{d\nu}{dQ^{(12)}}(z, s, B) = \frac{\exp(\beta' z) 1_{[s \geq t]} B}{S_1(t, \beta)}.$$

Let E_ν denote expectation with respect to ν. Then, for all $\alpha \neq 0$,

$$\begin{aligned}(7.9)\qquad &\alpha'\left(\frac{S_{KZ}}{S_1}(t,\beta) - \frac{S_K S_Z'}{S_1^2}(t,\beta)\right)\alpha \\ &\quad = \alpha'(E_\nu[K(Z,t)Z'] - E_\nu[K(Z,t)]E_\nu[Z'])\alpha \\ &\quad = \tfrac{1}{2}E_\nu[\alpha'(K(Z_2,t) - K(Z_1,t))(Z_2 - Z_1)'\alpha],\end{aligned}$$

where Z_1 and Z_2 are independent and are distributed as Z. Equation (7.9) is nonnegative for Q-almost all t, and for all probability measures ν, by (E2). Further, by the strict inequality required in (E2)(iii) and the nondegeneracy in (E3),

$$\frac{\partial}{\partial\beta} W_{K,\tau}(\beta, Q) = -E\left[\Delta_\tau\left(\frac{S_{KZ}}{S_1}(T,\beta) - \frac{S_K S_Z'}{S_1^2}(T,\beta)\right)\right]$$

is negative definite for all β. □

PROOF OF THEOREM 3.1 (Existence and uniqueness). The proof will use the following two results from Ortega and Rheinboldt (1970).

6.3.4 (Existence). Let C be an open, bounded set in $\mathbb{R}^d$ and assume that F: $\bar{C} \subset \mathbb{R}^d \to \mathbb{R}^d$ is continuous and satisfies $(x - x_0)'F(x) \leq 0$ for some $x_0 \in C$ and all $x \in \partial C$ (the boundary of C). Then $F(x) = 0$ has a solution in $\bar{C}$.

5.4.4 (Uniqueness). If F: $C \subset \mathbb{R}^d \to \mathbb{R}^d$ is continuously differentiable on an open convex set $C_0 \subset C$ and the derivative, $\dot{F}(x)$, is positive definite for all $x \in C_0$, then F is one to one on C_0.

By Lemma 7.2, $W_{K,\tau}(\beta, Q)$ is well defined for all β and is continuously differentiable in β. By Lemma 7.3, the derivative of the estimating equation is positive definite. Thus, by the uniqueness result 5.4.4 [Ortega and Rheinboldt (1970)], W is one to one.

Consider a fixed $\beta \in C_1 := \{\beta \in \mathbb{R}^d\colon |\beta| = 1\}$. We show that $n\beta' W_{K,\tau}(n\beta) < 0$ for n sufficiently large. Further, since $W_{K,\tau}$ is a continuous monotone function of β and since $\{\beta\colon |\beta| = 1\}$ is compact, there is some M, such that $\beta' W_{K,\tau}(\beta) < 0$ for all β with $|\beta| = M$. For each β, let

$$f(t,\beta,n) := \frac{S_{\beta' K}}{S_1}(t, n\beta) = \beta' \frac{S_K}{S_1}(t, n\beta) = \frac{E[\beta' K(Z,t)\exp(n\beta' Z)1_{[T \geq t]}]}{E[\exp(n\beta' Z)1_{[T \geq t]}]}.$$

By (E2) and the argument proving Lemma 7.3, $f(t, \beta, n)$ is, for each t, an increasing function of n, and $f(t,\beta,n) \to \beta' K(z^*(\beta,t),t)$, where $z^*(\beta,t)$ is such that

$$\begin{aligned} \beta' z^*(\beta,t) &= \operatorname{ess\,sup}\{\beta' Z\colon T \ge t\} \\ &\ge \operatorname{ess\,sup}\{\beta' K(Z,t)\colon T \ge t\} := \phi(t,\beta) \quad \text{by Lemma 7.1.} \end{aligned}$$

It has also been shown above that $|W_{K,\tau}(\beta)| < \infty$ for all β. Thus, by monotone convergence, (S_K/S_1 is monotone in β)

$$\begin{aligned} \beta' W_{K,\tau}(n\beta) &= E[\Delta_\tau \beta' K(Z,T)] - E[\Delta_\tau f(T,\beta,n)] \\ &\to E[\Delta_\tau \beta' K(Z,T)] - E\Big[\Delta_\tau \lim_{n\to\infty} f(T,\beta,n)\Big] \\ &\le E[\Delta_\tau \beta' K(Z,T)] - E[\Delta_\tau \phi(T,\beta')] \qquad (7.10) \\ &= E[\Delta_\tau (E[\beta' K(Z,T)|T,\Delta] - \phi(T,\beta))] \\ &< 0 \quad \text{by (E3)}, \end{aligned}$$

for all directions β. Define $D_M =: \{\beta\colon |\beta| < M\}$. Since $W_{K,\tau}(\cdot,Q)$ is continuous in β, there is an M, sufficiently large, such that

$$\beta' W_{K,\tau}(\beta) < 0 \quad \text{for } \beta \in \partial D_M.$$

For suppose there is no such M. Let $K_n := \{\beta \in C_1\colon \beta' W_{K,\tau}(n\beta) \ge 0\}$. If $r \ge s$, then $K_r \subseteq K_s$, by the monotonicity of $W_{K,\tau}$ (Lemma 7.3). [Since $\partial W_{K,\tau}/\partial\beta$ is negative definite, it follows that, for each $\beta \in C_1$, $\beta' W_{K,\tau}(n\beta)$ is a decreasing function of n.] K_n is a compact set, since $W_{K,\tau}$ is a continuous function; by supposition K_n is nonempty for all n. Thus $K_\infty := \bigcap_n K_n$ is a nonempty set. But this contradicts (7.10).

Thus $W_{K,\tau}(\beta,Q) = 0$ has a solution in $\bar{D}_M$, for some finite M, by the existence result 6.3.4 [Ortega and Rheinboldt (1970)]. This completes the proof. □

LEMMA 7.4.

(A) *Suppose that* (E1) *holds. Then for each* $\alpha \in \mathbb{R}^d$ *there exists a compact neighborhood* $\mathscr{B}_\alpha$ *(which contains an open ball) about* α *such that* S_1, S_Z, S_K, S_{KZ} *and* S_{K^2} *are bounded on* $[0,\tau] \times \mathscr{B}_\alpha$.

(B) *If, in addition,* $\inf_{t\in[0,\tau]} P(Y(t) = 1) > 0$, *then* S_1 *is bounded away from* 0 *on* $[0,\tau] \times \mathscr{B}_\alpha$.

PROOF.

$$\begin{aligned} |S_{KZ}(t,\beta)| &= |E[K(Z,t)Z' \exp(\beta' Z) Y(t)]| \\ &\le E[k(Z) w(t) |Z| \exp(\beta' Z)] \quad \text{by (E1)(ii)} \\ &\le w(t)\Big(E[k^2(Z)]E\big[|Z|^2 \exp(2\beta' Z)\big]\Big)^{1/2} \quad \text{by Cauchy-Schwarz.} \end{aligned}$$

Hence, by (E1), $\|S_{KZ}(\cdot,\beta)\|_0^\tau < \infty$.

Now, by the theory of Laplace transforms, (E1)(i) implies that $E[|Z|^m \exp(\beta' Z)] < \infty$ for $m = 0, 1, 2$ for all β, and that they are continuous functions of β. Thus, by the dominated convergence theorem, for $m = 0, 1, 2$,

$$\sup_{\beta \in \mathscr{B}_\alpha} E\big[|Z|^m \exp(\beta' Z)\big] = E\big[|Z|^m \exp(\beta'_* Z)\big] < \infty,$$

for some β_* (depending on m only) in $\mathscr{B}_\alpha$ (since $\mathscr{B}_\alpha$ is compact). Hence S_{KZ} is bounded on $[0, \tau] \times \mathscr{B}_\alpha$. Similarly, so are S_1, S_Z, S_K and S_{K^2}.

By Hölder's inequality [as in the proof of Lemma 7.2, (7.3)],

$$S_1(t, \beta) \geq \big(E[Y(t)]\big)^\eta \big(E[\exp(-\beta' Z/(\eta - 1))]\big)^{-(\eta-1)} \quad \text{for all } \eta > 1.$$

By the continuity of $E[\exp(\beta' Z)]$, there exists, for any compact $\mathscr{B}_\alpha \subset \mathbb{R}^d$, a $\beta_* \in \mathscr{B}_\alpha$ such that $\inf_{\beta \in \mathscr{B}_\alpha} E[\exp(-\beta' Z/(\eta - 1))] = E[\exp(-\beta'_* Z/(\eta - 1))] > 0$ by (E1)(i). Also, $\inf_{t \in [0,\tau]} E[Y(t)] = \inf_{t \in [0,\tau]} P(Y(t) = 1) > 0$ by supposition. Hence S_1 is bounded away from 0 on $[0, \tau] \times \mathscr{B}_\alpha$. □

LEMMA 7.5. *Suppose that $K(t) := K(Z, t, P)$ is a random element of $D[0, \tau]^d$. In the iid case* (E1) *implies that there exists a (full dimensional) neighborhood $\mathscr{B}_0$ of β_0 such that for any P in the Cox model with parameter β_0; for $J = Z$, K, KZ, K^2 and "1" (i.e., $\mathbb{S}_{n1}$),*

$$\sup_{t \in [0,\tau]} \sup_{\beta \in \mathscr{B}_0} |\mathbb{S}_{nJ}(t, \beta) - S_J(t, \beta)| \to_p 0,$$

for some limiting function $S_J(t, \beta)$. Further, the limiting functions $S_J(t, \cdot)$ are continuous in $\beta \in \mathscr{B}_0$, uniformly in $t \in [0, \tau]$.

PROOF. Use the uniform law of large numbers for $D[0, 1]$ random variables [Rao (1963)] and the generalization of this result given by Andersen and Gill (1982), Theorem 3.1. [See also the proof of Andersen and Gill (1982), Theorem 4.1.] By the uniform law of large numbers, it suffices to show that there exists a compact (full dimensional) neighborhood $\mathscr{B}_0$ of β_0 such that for $J = Z$, K, KZ, K^2 and 1,

$$E\left[\sup_{t \in [0,\tau],\, \beta \in \mathscr{B}_0} Y(t) |J(Z, t)| \exp(\beta' Z)\right] < \infty.$$

By (E1), for all $\beta \in \mathbb{R}^d$, $E[\sup_{t \in [0,\tau]} Y(t) \exp(\beta' Z)] \leq E[\exp(\beta' Z)] < \infty$, and

$$E\left[\sup_{t \in [0,\tau]} Y(t) |K(Z, t)| \exp(\beta' Z)\right] \leq \|w\|_0^\tau \big(E[k^2(Z)] E[\exp(2\beta' Z)]\big)^{1/2} < \infty.$$

Without loss of generality, take $\mathscr{B}_0$ to be the closed ball of radius α, say, about β_0. Now

$$|z| \leq \sum_{i=1}^{k} |z_i| = u' z,$$

where z_i is the ith coordinate of z and u is a vector of 1's and -1's of length

d. Let U denote the 2^d different vectors u. Then

$$\begin{aligned} E\left[\sup_{t\in[0,\tau],\,\beta\in\mathscr{B}_0} Y(t)\exp(\beta'Z)\right] &\le E\left[\sup_{\beta\in\mathscr{B}_0}\{\exp(\beta'Z)\}\right] \\ &\le E\left[\exp(\beta_0'Z+\alpha|Z|)\right] \\ &\le E\left[\sum_{u\in U}\exp((\beta_0+\alpha u)'Z)\right] \\ &= \sum_{u\in U} E\left[\exp((\beta_0+\alpha u)'Z)\right] \\ &< \infty \quad \text{by (E1)(i)}. \end{aligned}$$

Similarly, for $J = Z, K, K^2$ and KZ, $E[\sup_{t\in[0,\tau],\,\beta\in\mathscr{B}_0} Y(t)|J(Z,t)|\exp(\beta'Z)] < \infty$. Hence $\mathbb{S}_{nJ}$ converges uniformly to S_J.

By dominated convergence, S_1, S_Z, S_K, S_{KZ} and S_{K^2} are continuous functions of $\beta\in\mathscr{B}_0$ for each $t\in[0,\tau]$, uniformly in $t\in[0,\tau]$. □

Proof of Theorem 4.1 (Asymptotic normality). The proof follows the recipe of Theorem 7.2.4 of Bickel, Klaassen, Ritov and Wellner (1993), as given in Theorem A.1. In particular, we shall verify conditions (A1)–(A5) of that theorem.

Conditions (A1)–(A3) follows from Theorem 3.1 and its proof [see, in particular, Lemma 7.2 for (A3)]. It remains to verify (A4′) (by Lemma A.2) and (A5).

To see (A4′) write

$$\begin{aligned} &\left|\left(\dot W(\beta,\mathbb{P}_n)-\dot W(\beta_0,P)\right)_{ij}\right| \\ &\quad\le \left|\left(\dot W(\beta,\mathbb{P}_n)-\dot W(\beta,P)\right)_{ij}\right| + \left|\left(\dot W(\beta,P)-\dot W(\beta_0,P)\right)_{ij}\right|. \end{aligned}$$

We show that the first term above is $o_p(1)$, uniformly in β in some neighborhood of β_0, and that the second term tends to 0 as $\beta\to\beta_0$:

$$\begin{aligned} |\dot W(\beta,\mathbb{P}_n)-\dot W(\beta,P)| &= \left|\int_0^\tau\left(\frac{\mathbb{S}_{n\mathbb{K}_nZ}}{\mathbb{S}_{n1}}-\frac{\mathbb{S}_{n\mathbb{K}_n}\mathbb{S}'_{nZ}}{\mathbb{S}_{n1}^2}\right)(t,\beta)\,d\mathbb{P}_n^{(2)}(t)\right. \\ &\qquad\left.-\int_0^\tau\left(\frac{S_{KZ}}{S_1}-\frac{S_KS_Z'}{S_1^2}\right)(t,\beta)\,dP^{(2)}(t)\right| \\ &\le\left|\int_0^\tau\left(\frac{S_{KZ}}{S_1}-\frac{S_KS_Z'}{S_1^2}\right)(t,\beta)\,d\left(\mathbb{P}_n^{(2)}-P^{(2)}\right)(t)\right| \\ &\quad+\int_0^\tau\left|\left(\left(\frac{\mathbb{S}_{n\mathbb{K}_nZ}}{\mathbb{S}_{n1}}-\frac{\mathbb{S}_{n\mathbb{K}_n}\mathbb{S}'_{nZ}}{\mathbb{S}_{n1}^2}\right)\right.\right. \\ &\qquad\left.\left.-\left(\frac{S_{KZ}}{S_1}-\frac{S_KS_Z'}{S_1^2}\right)\right)(t,\beta)\right|\,d\mathbb{P}_n^{(2)}(t). \end{aligned}$$

The first term is $o_p(1)$ by the weak law of large numbers since, by Lemma 7.4,

$$\left|\int_0^\tau\left(\frac{S_{KZ}}{S_1}-\frac{S_K S_Z'}{S_1^2}\right)(t,\beta)\,dP^{(2)}(t)\right|<\infty.$$

The second term is dealt with in two parts. Let $\mathscr{B}_0$ denote a compact neighborhood of β_0 (as in Lemma 7.4):

$$\begin{aligned}
&\sup_{\beta\in\mathscr{B}_0}\int_0^\tau\left|\frac{\mathbb{S}_{n\mathbb{K}_nZ}}{\mathbb{S}_{n1}}-\frac{S_{KZ}}{S_1}\right|(t,\beta)\,d\mathbb{P}_n^{(2)}\\
&\quad\le\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{n(\mathbb{K}_n-K)Z}}{\mathbb{S}_{n1}}(t,\beta)\right\|_0^\tau+\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nKZ}}{\mathbb{S}_{n1}}-\frac{S_{KZ}}{S_1}\right\|_0^\tau\\
&\quad\le\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}}\right\|_0^\tau\|\mathbb{K}_n(z,t)-K(z,t)\|_\infty\\
&\qquad+\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nKZ}}{\mathbb{S}_{n1}}\right\|_0^\tau\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{n1}-S_1}{S_1}\right\|_0^\tau+\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nKZ}-S_{KZ}}{S_1}\right\|_0^\tau\\
&\quad=o_p(1),
\end{aligned}$$

by condition (K1) and Lemmas 7.4 and 7.5;

$$\begin{aligned}
&\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{n\mathbb{K}_n}\mathbb{S}_{nZ}'}{\mathbb{S}_{n1}^2}-\frac{S_KS_Z'}{S_1^2}\right\|_0^\tau\\
&\qquad=\sup_{\beta\in\mathscr{B}_0}\left\|\frac{S_K}{S_1}\left(\frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}}-\frac{S_Z}{S_1}\right)'+\left(\frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}}-\frac{S_K}{S_1}\right)\frac{\mathbb{S}_{nZ}'}{\mathbb{S}_{n1}}\right\|_0^\tau\\
&\qquad\le\sup_{\beta\in\mathscr{B}_0}\left\|\frac{S_K}{S_1}\right\|_0^\tau\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}}-\frac{S_Z}{S_1}\right\|_0^\tau\\
&\qquad\quad+\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}}-\frac{S_K}{S_1}\right\|_0^\tau\sup_{\beta\in\mathscr{B}_0}\left\|\frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}}\right\|_0^\tau\\
&\qquad=o_p(1)
\end{aligned}$$

by the boundedness of $\sup_{[0,\tau]\times\mathscr{B}_0}|S_K/S_1|$ and $\sup_{[0,\tau]\times\mathscr{B}_0}|\mathbb{S}_{nZ}/\mathbb{S}_{n1}|$ for sufficiently large n (Lemma 7.4) and by an argument identical to that used to show that $\sup_{[0,\tau]\times\mathscr{B}_0}|\mathbb{S}_{n\mathbb{K}_nZ}/\mathbb{S}_{n1}-S_{KZ}/S_1|=o_p(1)$. Hence $\sup_{\beta\in\mathscr{B}_0}|\dot{W}(\beta,\mathbb{P}_n)-\dot{W}(\beta,P)|=o_p(1)$. It is thus sufficient to show that $\dot{W}(\beta,P)$ is continuous in β at β_0, and this follows from the continuity of the integrand uniformly in $t\in[0,\tau]$ [Lemma 7.4(B) and 7.5]. Hence (A4′) is satisfied.

Next we verify (A5). The form of the influence curve can be motivated by formally differentiating W with respect to the measure P. The rigorous proof below is quite long. The argument β, which is fixed at β_0, will often be dropped in the rest of this proof, for example, $\mathbb{S}_{n\mathbb{K}_n}(t) := \mathbb{S}_{n\mathbb{K}_n}(t, \beta_0)$.

$$W(\beta_0, \mathbb{P}_n)$$

$$:= \int^{\tau}\int \mathbb{K}_n \, d\mathbb{P}_n^{(u)} - \int_0^{\tau} \frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}} d\mathbb{P}_n^{(2)}(t)$$

$$= W(\beta_0, P) + \int_0^{\tau}\int K \, d\big(\mathbb{P}_n^{(u)} - P^{(u)}\big) + \int_0^{\tau}\int \dot{\mathbb{K}}_n \, dP^{(u)}$$

$$- \int_0^{\tau} \frac{S_K}{S_1} d\big(\mathbb{P}_n^{(2)} - P^{(2)}\big)$$

$$(*) \qquad - \int_0^{\tau}\left(\frac{\mathbb{S}_{nK}}{S_1} - \frac{S_K}{S_1} + \frac{S_{\dot{\mathbb{K}}_n}}{S_1} - \frac{S_K}{S_1}\left[\frac{\mathbb{S}_{n1} - S_1}{S_1}\right]\right) dP^{(2)}$$

$$+ \int_0^{\tau}\int \dot{\mathbb{K}}_n \, d\big(\mathbb{P}_n^{(u)} - P^{(u)}\big) - \int_0^{\tau}\int \big(\mathbb{K}_n - K - \dot{\mathbb{K}}_n\big) \, d\mathbb{P}_n^{(u)}$$

$$- \int_0^{\tau}\left(\frac{\mathbb{S}_{nK}}{S_1} - \frac{S_K}{S_1} + \frac{S_{\dot{\mathbb{K}}_n}}{S_1} - \frac{S_K}{S_1}\left[\frac{\mathbb{S}_{n1} - S_1}{S_1}\right]\right) d\big(\mathbb{P}_n^{(2)} - P^{(2)}\big)$$

$$- \int_0^{\tau}\left\{\frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}} - \frac{S_K}{S_1} - \left(\frac{\mathbb{S}_{nK}}{S_1} - \frac{S_K}{S_1} + \frac{S_{\dot{\mathbb{K}}_n}}{S_1} - \frac{S_K}{S_1}\left[\frac{\mathbb{S}_{n1} - S_1}{S_1}\right]\right)\right\} d\mathbb{P}_n^{(2)}.$$

The identity is easily seen to hold by cancelling terms that have been added and subtracted to the right-hand side of $(*)$. Recall that $W(\beta_0, P) = 0$ [assumption (A1)], the next four terms are all linear and the final four, which we shall refer to as A, B, C and D, respectively, will be shown to be $o_p(n^{-1/2})$.

A: This term may be written as a V-statistic and will be dealt with later.

B: Under (K1)(I), $|B| \le \|\mathbb{K}_n - K - \dot{\mathbb{K}}_n\|_\infty = o_p(n^{-1/2})$, whereas under (K1)(II), $|B| \le \|w_n - w - \dot{w}_n\|_0^{\tau} \int |z| dP^{(1)}(z) = o_p(n^{-1/2})$.

C: Another V-statistic—see below.

D: Let $\|\cdot\|_0^{\tau}$ denote the supremum norm over $t \in [0, \tau]$. Under (K1)(I),

$$|D| \le \left\|\frac{\mathbb{S}_{n1} - S_1}{S_1}\right\|_0^{\tau} \left\|\frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}} - \frac{S_K}{S_1}\right\|_0^{\tau}$$

$$+ \left\|\frac{\mathbb{S}_{n1}}{S_1}\right\|_0^{\tau} \left\|\sup_{z \in \mathscr{Z}} |\mathbb{K}_n(z, \cdot) - K(z, \cdot) - \dot{\mathbb{K}}_n(z, \cdot)|\right\|_0^{\tau}$$

$$+ \left|\int_0^{\tau}\left(\frac{\mathbb{S}_{n\dot{\mathbb{K}}_n}}{S_1} - \frac{S_{\dot{\mathbb{K}}_n}}{S_1}\right)(t, \beta_0) \, d\mathbb{P}_n^{(2)}\right|,$$

since $\|\mathbb{S}_{nK}\|_0^{\tau} \le \|\mathbb{S}_{n1}\|_0^{\tau}\|K\|_\infty$ and $\|\mathbb{S}_{n\mathbb{K}_n}\|_0^{\tau} \le \|\mathbb{S}_{n1}\|_0^{\tau}\|\mathbb{K}_n\|_\infty$ and so on. Now $\|(\mathbb{S}_{n1} - S_1)/S_1\|_0^{\tau} = O_p(n^{-1/2})$ by Lemma A.3 of Tsiatis (1981) [or directly using the central limit theorem of Pollard (1982)], and as was shown in verifying (A4′),

$$\left\| \frac{\mathbb{S}_{n\mathbb{K}_n}}{\mathbb{S}_{n1}} - \frac{S_K}{S_1} \right\|_0^\tau = o_p(1).$$

Also $\|\mathbb{S}_{n1}/S_1\|_0^\tau$ is finite and $\|\mathbb{K}_n - K - \dot{\mathbb{K}}_n\|_\infty = o_p(n^{-1/2})$. Hence the first two terms are $o_p(n^{-1/2})$. That leaves

$$\left| \int_0^\tau \left(\frac{\mathbb{S}_{n\dot{\mathbb{K}}_n}}{S_1} - \frac{S_{\dot{\mathbb{K}}_n}}{S_1} \right) d\mathbb{P}_n^{(2)} \right|,$$

which is shown to be $o_p(n^{-1/2})$ in Lemma 7.6 following this proof.

If instead (K1)(II) holds, replace the second term on the right of the expansion of $|D|$ by $\|\mathbb{S}_{nZ}/\mathbb{S}_{n1}\|_0^\tau \|w_n - w - \dot{w}_n\|_0^\tau = o_p(n^{-1/2})$.

Returning to term A,

$$\int_0^\tau \int \dot{\mathbb{K}}_n \, d(\mathbb{P}_n^{(u)} - P^{(u)}) = \int_0^\tau \int \int \psi_K(x; z, t) \, d(\mathbb{P}_n - P)(x) \, d(\mathbb{P}_n^{(u)} - P^{(u)})(z, t).$$

This is a V-statistic with (centered) kernel $v_A(X_1, X_2) = \Delta_{\tau 1} \psi_K(X_2; Z_1, T_1)$ and hence, by (K1) and Appendix B, $A = o_p(n^{-1/2})$.

Term C will be dealt with in pieces:

$$\begin{aligned}
&\int_0^\tau \left(\frac{\mathbb{S}_{nK}}{S_1} - \frac{S_K}{S_1} + \frac{\mathbb{S}_{\dot{\mathbb{K}}_n}}{S_1} - \frac{S_K}{S_1}\left[\frac{\mathbb{S}_{n1} - S_1}{S_1} \right] \right) d(\mathbb{P}_n^{(2)} - P^{(2)}) \\
&\quad = \int_0^\tau \left(\frac{\mathbb{S}_{nK} - S_K}{S_1} - \frac{S_K}{S_1}\left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \right) d(\mathbb{P}_n^{(2)} - P^{(2)}) \\
&\qquad + \int_0^\tau \frac{S_{\dot{\mathbb{K}}_n}}{S_1} d(\mathbb{P}_n^{(2)} - P^{(2)}) \\
&\quad = E + F, \quad \text{say.}
\end{aligned}$$

$$F = \int_0^\tau \int v_F(x_1, x_2) \, d(\mathbb{P}_n - P)(x_1) \, d(\mathbb{P}_n^{(2)} - P^{(2)})(t_2),$$

where

$$v_F(x_1, x_2) = E[\psi_K(x_1; Z, t_2) \exp(\beta_0' Z) Y(t_2)] / S_1(t_2),$$

another V-statistic. Recall that $\psi^*(x, z) = \sup_{t \in [0,\tau]} |\psi_K(x; z, t)|$, so, by Cauchy-Schwarz,

$$|v_F(x_1, x_2)| \le \left\| \frac{1}{S_1} \right\|_0^\tau \left(E\left[\psi_K^*(x_1, Z)^2 \right] E[\exp(2\beta_0' Z)] \right)^{1/2},$$

for all $t_2 \in [0, \tau]$. Thus, by (E1) and Lemma 7.4, $|v_F(x_1, x_2)|^2 = O(E[\psi_K^*(x_1, Z)^2])$ and $F = O_p(n^{-1/2})$ by (K1). Similarly,

$$E = \int_0^\tau \int v_E(x_1, x_2) \, d(\mathbb{P}_n^{(12)} - P^{(12)})(x_1) \, d(\mathbb{P}_n^{(2)} - P^{(2)})(t_2),$$

where

$$v_E(x_1, x_2) = \left[K(z_1, t_2) - \frac{S_K}{S_1}(t_2)\right] \frac{\exp(\beta_0' z_1)}{S_1(t_2)} 1_{[t_1 \geq t_2]} B_1 \delta_2.$$

[Recall that $Y(t) = 1_{[T \geq t]} B$ and use the convention that integration with respect to $P(x)$ gives the expectation over B, too.] Now

$$\begin{aligned} v_E(x_1, t_2)^2 &= \left[K(z_1, t_2) - \frac{S_K}{S_1}(t_2)\right]^2 \frac{\exp(2\beta_0' z_1)}{S_1^2(t_2)} 1_{[t_1 \geq t_2]} B_1 \delta_2 \\ &\leq \left(k(z_1) \|w\|_0^\tau + \left\| \frac{S_K}{S_1} \right\|_0^\tau \right)^2 \exp(2\beta_0' z_1) \left\| \frac{1}{S_1^2} \right\|_0^\tau \\ &= M k^2(z_1) \exp(2\beta_0' z_1), \end{aligned}$$

where M is a constant, by Lemma 7.4 and condition (E1). Thus, by (E1),

$$E v_E(X_i, T_j)^2 \leq M E\left[k^2(Z_i) \exp(2\beta_0' Z_i)\right] < \infty.$$

Hence, by the theory of V-statistics, Corollary B.1, $E = o_p(n^{-1/2})$.

We have shown that the remainder terms A through D in $(*)$ are indeed $o_p(n^{-1/2})$, and that the other terms may all be written as the sum of iid random variables. Thus, by the central limit theorem, we have established that

$$\sqrt{n}\, W_K(\beta_0, \mathbb{P}_n) = \sqrt{n} \int \phi(x, P)\, d\mathbb{P}_n(x) + o_p(1)$$

and that $E[\phi(X, P)] = 0$.

It remains to prove that $\phi \in L_2(P)$.

Using the inequality $(\sum_{i=1}^m a_i)^2 \leq m \sum_{i=1}^m a_i^2$, it is enough to consider the square of each of the nonconstant terms in ϕ. There are six terms all of which can be shown to have finite second moment. The following argument for the fifth term is the most involved:

$$\begin{aligned} &E\left[\left(\int_0^\tau \frac{E\left[\psi_K(X_1; Z_2, t) e^{\beta_0' Z_2} Y_2(t) | X_1\right]}{S_1(t)} dP^{(2)}(t)\right)^2\right] \\ &\quad \leq E\left[\int_0^\tau \frac{E\left[\psi_K^*(X_1, z_2)^2 e^{2\beta_0' Z_2} | X_1\right]}{S_1^2(t)} dP^{(2)}(t)\right] \quad \text{by Jensen (twice)} \\ &\quad \leq \left\| \frac{1}{S_1^2} \right\|_0^\tau \left(E\left[\psi_K^*(X_1, z_2)^{2r}\right]\right)^{1/r} \left(E\left[e^{2s\beta_0' Z}\right]\right)^{1/2} \quad \text{for any } r > 1,\ \frac{1}{r} + \frac{1}{s} = 1 \\ &\quad < \infty \quad \text{by Lemma 7.4 and conditions (E1) and (K1).} \end{aligned}$$

This completes the proof. □

Extending the proof to cover estimators that use all the data. It is possible to extend the proof of Theorem 4.1 to cover estimators that use all the observed data. Rather than adding to the complexity of this paper, we merely sketch the additional steps. The basic idea is to extend the arguments of the existing proofs to cover the case of a slowly increasing end point $\tau_n := \inf\{t: E[e^{\beta' Z} 1_{[T \geq t]}] \leq n^{-\alpha}\}$ for some $\alpha \in (0, 1)$, and to show that the contribution to the statistics from τ_n to ∞ is negligible.

The proofs of Lemmas 7.1–7.3 and Theorem 3.1 all hold without modification. Conditions (A1)–(A3) still hold for the estimating equations with $\tau = \infty$. Thus the only difficulty in proving asymptotic normality is to show that (A4′) and (A5) still hold.

To show (A4′), one needs to look at terms like

$$\sup_{\beta \in \mathscr{B}} \left\| \frac{\mathbb{S}_{nKZ} - S_{KZ}}{S_1} \right\|_0^{\tau_n}.$$

Essentially, one can use Bennett's (1962) exponential inequality as extended by Alexander (1984). Theorem 1 of Lai and Ying (1988) is an application similar to our needs. Care must be taken if one wants to work in the generality of unbounded Z and data-dependent K. The basic result gives

$$\sup_{\beta \in \mathscr{B}} \left\| \frac{\mathbb{S}_{nKZ} - S_{KZ}}{S_1} \right\|_0^{\tau_n} = n^{-1/2+\alpha/2+\theta} \quad \text{for all } \theta > 0.$$

From the proof of Lemma 7.2 it is clear that

$$\sup_{\beta \in \mathscr{B}} \left| \int_{\tau_n}^{\infty} \left(\frac{S_{KZ}}{S_1} - \frac{S_K S_Z'}{S_1^2} \right)(t, \beta)\, dP^{(2)}(t) \right| \to 0 \quad \text{as } n \to \infty$$

and the convergence holds in probability for the empirical version.

Bennett's exponential inequality is again the key to showing that (A5) holds on $[0, \tau_n]$. Notice that the V-statistic argument holds even when one has a sequence of kernels whose second moments are increasing. Greater care is required to show that the contribution from (τ_n, ∞) is negligible, since negligible now means $o_p(n^{-1/2})$. In (A5), however, we are working with a fixed β_0 and (at least when the covariates are bounded and K is not data-dependent) Lemma 2 of Ying (1991) can be used directly to obtain the desired result. □

Since $Y_i(t) = 1_{[T_i \geq t]} B_i$, let $P^{(B)}$ denote the probability measure for (Z, T, B), that is,

$$P^{(B)}(z, s, 1) = P^{(12)}(z, s)\alpha \quad \text{and} \quad P^{(B)}(z, s, 0) = P^{(12)}(z, s)(1 - \alpha).$$

Similarly, $\mathbb{P}_n^{(B)}$ is the empirical measure corresponding to $P^{(B)}$.

LEMMA 7.6. *Under conditions* (E1)–(E3) *and* (K1), *for* τ *such that* $\inf_{t \in [0, \tau]} P(Y(t) = 1) > 0$,

$$\left| \int_0^{\tau} \left(\frac{\mathbb{S}_{n\mathbb{K}_n} - S_{\mathbb{K}_n}}{S_1} \right)(t, \beta_0)\, d\mathbb{P}_n^{(2)}(t) \right| = o_p(n^{-1/2}).$$

PROOF.

$$\int_0^\tau \left(\frac{\mathbb{S}_{n\dot{\mathbb{K}}_n} - S_{\dot{\mathbb{K}}_n}}{S_1} \right)(t)\, d\mathbb{P}_n^{(2)}(t)$$

$$= \int_0^\tau \left(\frac{\mathbb{S}_{n\dot{\mathbb{K}}_n} - S_{\dot{\mathbb{K}}_n}}{S_1} \right) dP^{(2)} + \int_0^\tau \left(\frac{\mathbb{S}_{n\dot{\mathbb{K}}_n} - S_{\dot{\mathbb{K}}_n}}{S_1} \right) d\big(\mathbb{P}_n^{(2)} - P^{(2)}\big).$$

Now the first term is a V-statistic, with kernel

$$v(X_1, X_2) = \int_0^\tau \frac{1_{(T_1 \ge t)} B_1}{S_1(t)} e^{\beta_0' Z_1} \psi_K(X_2; Z_1, t)\, dP^{(2)}(t),$$

for it is equal to

$$\int_0^\tau \int \frac{1_{(s \ge t)} B}{S_1(t)} e^{\beta_0' z} \dot{\mathbb{K}}_n(z,t)\, d\big(\mathbb{P}_n^{(B)} - P^{(B)}\big)(z,s,B)\, dP^{(2)}(t)$$

$$= \int_0^\tau \int\int \frac{1_{(s \ge t)} B}{S_1(t)} e^{\beta_0' z} \psi_K(x; z, t)\, d(P_n - P)(x)$$

$$\times d\big(\mathbb{P}_n^{(B)} - P^{(B)}\big)(z,s,B)\, dP^{(2)}(t)$$

$$= \int\int \left\{ \int_0^\tau \frac{1_{(s \ge t)} B}{S_1(t)} e^{\beta_0' z} \psi_K(x; z, t)\, dP^{(2)}(t) \right\}$$

$$\times d(\mathbb{P}_n - P)(x)\, d\big(\mathbb{P}_n^{(B)} - P^{(B)}\big)(z,x,B)$$

by Fubini. Thus, by (K1)(iii),

$$v^2(X_1, X_2) \le \left(\left\| \frac{1}{S_1} \right\|_0^2 \right)^2 \big(e^{\beta_0' Z_1} \psi_K^*(X_2, Z_1) \big)^2.$$

By Lemma 7.4, Hölder's inequality (K1) and (E1)(i),

$$Ev^2(X, X) < \infty \quad \text{and} \quad Ev^2(X_1, X_2) < \infty \quad \text{for } X_1, X_2 \text{ independent}.$$

Thus the hypotheses of Corollary B.1 for the kernel of the V-statistic are satisfied and

$$\left| \int_0^\tau \frac{\mathbb{S}_{n\dot{\mathbb{K}}_n} - S_{\dot{\mathbb{K}}_n}}{S_1} dP^{(2)} \right| = O_p(n^{-2}).$$

The remaining term is another V-statistic, with centered kernel

$$\omega(x_1, x_2, x_3) := \frac{1_{(t_1 \ge t_3)} B_1\, e^{\beta_0' z_1} \psi_K(x_2; z_1, t_3)}{S_1(t_3)}.$$

Once again apply Corollary B.1. The hypotheses are easily checked since the kernel is bounded in its third argument. $|\omega(x_1, x_2, x_3)| \le C e^{\beta_0 z_1} \psi_K^*(x_2, z_1)$ where $C = \|1/S_1\|_0^\tau$. Hence, using the inequality $|2E[AB]| \le EA^2 + EB^2$, it is enough to show that

$$E\left[e^{2\beta_0' Z_1}(\psi_K^*(X_2, Z_1))^2\right] < \infty \quad \text{and} \quad E\left[e^{2\beta_0' Z}(\psi_K^*(X, Z))^2\right] < \infty.$$

These follow from Hölder's inequality and conditions (E1) and (K1). Hence the V-statistic is $O_p(n^{-3})$.

This completes the proof of the lemma. □

PROOF OF COROLLARY 4.2 (Case-cohort). When the covariates are time-independent, Theorem 4.1 is applicable and it is necessary only to show that on the Cox model the parameter defined by the case-cohort estimating equation is the Cox model parameter β_0, and then to evaluate the asymptotic variance.

We know that $\hat{\beta}$ will be consistent for β_0 on the model if

$$\frac{E\left[Y(t)e^{\beta_0' Z(t)}Z(t)\right]}{E\left[Y(t)e^{\beta_0' Z(t)}\right]} = E[Z(t)|T = t, \Delta = 1] := e(t), \quad \text{say.}$$

Now

$$E\left[Y(t)e^{\beta_0' Z(t)}Z(t)\right] = \alpha E\left[1_{[T \geq t]}e^{\beta_0' Z(t)}Z(t)\right]$$

and

$$E\left[Y(t)\, e^{\beta_0' Z(t)}\right] = \alpha E\left[1_{[T \geq t]}e^{\beta_0' Z(t)}\right],$$

by independence of B from (Z, T, Δ). Thus [cf. Sasieni (1992), Lemma 2c]

$$\frac{E\left[Y(t)\, e^{\beta_0' Z(t)}Z(t)\right]}{E\left[Y(t)\, e^{\beta_0' Z(t)}\right]} = \frac{E\left[1_{[T \geq t]}e^{\beta_0' Z(t)}Z(t)\right]}{E\left[1_{[T \geq t]}e^{\beta_0' Z(t)}\right]} = E[Z(t)|T = t, \Delta = 1].$$

Hence, by Theorem 4.1, $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically distributed as a normal random variable with mean 0 and variance given by $E(\tilde{l}(X))^{\otimes 2}$.

For the case-cohort estimator,

$$\begin{aligned}\tilde{l}(X_j) = \Sigma^{-1}\Bigg\{&\Delta_{\tau j}Z_j(T_j) - \Delta_{\tau j}\frac{S_Z}{S_1}(T_j) \\ &- \int_0^\tau \frac{Z_j(t)\, e^{\beta_0' Z_j(t)}1_{[T_j \geq t]}B_j}{S_1(t)}\, dP^{(2)}(t) \\ &+ \int_0^\tau \frac{S_Z}{S_1}(t)\frac{e^{\beta_0' Z_j(t)}1_{[T_j \geq t]}B_j}{S_1(t)}\, dP^{(2)}(t)\Bigg\},\end{aligned}$$

where

$$\Sigma = \int_0^\tau \left(\frac{S_{Z^2}}{S_1} - \left(\frac{S_Z}{S_1}\right)^{\otimes 2}\right)(t)\, dP^{(2)}(t) = E\,\mathrm{Var}(\Delta_\tau Z(T)|T, \Delta),$$

since by the independence of the cohort $S_Z/S_1(t) = e(t)$.

Now,

$$
\begin{aligned}
&E\big(\Sigma\tilde{l}(X)\big)^{\otimes 2} \\
&\quad = E\Big[\Delta_\tau(Z(T) - e(T))^{\otimes 2}\Big] \\
&\qquad + E\left[\left(\int_0^\tau (Z(t) - e(t)) \frac{e^{\beta_0' Z(t)} 1_{[T \geq t]} B}{S_1(t)} dP^{(2)}(t)\right)^{\otimes 2}\right] \\
&\qquad - 2E\left[\Delta_\tau(Z(T) - e(T)) \int_0^\tau (Z(t) - e(t))' \frac{e^{\beta_0' Z(t)} 1_{[T \geq t]} B}{S_1(t)} dP^{(2)}(t)\right] \\
&\quad = \Sigma + R_\alpha, \quad \text{say,}
\end{aligned}
$$

where

$$
\begin{aligned}
R_\alpha = \alpha E\Bigg[\int_0^\tau \int_0^\tau &(Z(t) - e(t))(Z(u) - e(u))' \\
&\times \frac{e^{\beta_0'(Z(t)+Z(u))} 1_{[T \geq t \vee u]}}{S_1(t) S_1(u)} dP^{(2)}(u)\, dP^{(2)}(t)\Bigg] \\
- 2\alpha E\Bigg[\int_0^\tau &\Delta_\tau(Z(T) - e(T))(Z(t) - e(t))' e^{\beta_0' Z(t)} 1_{[T \geq t]} \frac{dP^{(2)}(t)}{S_1(t)}\Bigg].
\end{aligned}
$$

Now, a little algebra gives [Sasieni (1992), Lemma 2b]

$$
\frac{dP^{(2)}(t)}{d\Lambda_0(t)} = \frac{S_1(t)}{\alpha},
$$

and, for any function ϕ, $E[\Delta_\tau \phi(T) | Z] = \int_0^\tau \phi(t)\, e^{\beta_0' Z(t)} E[1_{[T \geq t]} | Z]\, d\Lambda_0(t)$. Thus

$$
\begin{aligned}
R_\alpha = \alpha^{-1} E\Bigg[\int_0^\tau \int_0^\tau &(Z(t) - e(t))(Z(u) - e(u))' e^{\beta_0'(Z(t)+Z(u))} \\
&\times 1_{[T \geq t \vee u]}\, d\Lambda_0(u)\, d\Lambda_0(t)\Bigg] \\
- 2E\Bigg[\int_0^\tau \int_0^\tau &(Z(t) - e(t))(Z(u) - e(u))' e^{\beta_0'(Z(t)+Z(u))} \\
&\times 1_{[T \geq t]} 1_{[t \geq u]}\, d\Lambda_0(u)\, d\Lambda_0(t)\Bigg] \\
= 2(\alpha^{-1} - 1) E\Bigg[\int_0^\tau \int_0^\tau &(Z(t) - e(t))(Z(u) - e(u))' e^{\beta_0'(Z(t)+Z(u))} \\
&\times 1_{[T \geq t]}\, d\Lambda_0(u)\, d\Lambda_0(t)\Bigg].
\end{aligned}
$$

It remains to extend the proof of Theorem 4.1 to permit bounded time-dependent covariates when $K(Z, t) = Z(t)$. Lemmas 7.1–7.6 extend easily. The key step in the proof of the theorem that presents difficulty is showing that $|D| = o_p(n^{-1/2})$ since without some smoothness constraint on the covariate paths $\|\mathbb{S}_{n1} - S_1\| \neq O_p(n^{-1/2})$. But for the case-cohort estimator $\mathbb{K}_n = K$ and $\dot{\mathbb{K}}_n = 0$. Thus

$$|D| = \left| \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left(\frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}} - \frac{S_Z}{S_1} \right) d\mathbb{P}_n^{(2)} \right|$$

$$= \left| \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left\{ \left(\frac{\mathbb{S}_{nZ} - S_Z}{S_1} \right) - \frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}} \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \right\} d\mathbb{P}_n^{(2)} \right|$$

$$\leq \left| \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left(\frac{\mathbb{S}_{nZ} - S_Z}{S_1} \right) d\mathbb{P}_n^{(2)} \right| + \left\| \frac{\mathbb{S}_{nZ}}{\mathbb{S}_{n1}} \right\|_0^\tau \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right)^2 d\mathbb{P}_n^{(2)}.$$

Proceed as in the proof of Lemma 7.6:

$$\int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left(\frac{\mathbb{S}_{nZ} - S_Z}{S_1} \right) d\mathbb{P}_n^{(2)} = \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left(\frac{\mathbb{S}_{nZ} - S_Z}{S_1} \right) d\left(\mathbb{P}_n^{(2)} - P^{(2)}\right)$$
$$+ \int_0^\tau \left(\frac{\mathbb{S}_{n1} - S_1}{S_1} \right) \left(\frac{\mathbb{S}_{nZ} - S_Z}{S_1} \right) dP^{(2)}.$$

The right-hand side is the sum of two V-statistics with bounded kernels. The first kernel

$$k_1(X_1, X_2, T_3) = \frac{\exp(\beta_0' Z_1(T_3)) 1_{[T_1 \geq T_3]} Z_2(T_3) \exp(\beta_0' Z_2(T_3)) 1_{[T_2 \geq T_3]}}{S_1(T_3)^2},$$

and the second $k_2(X_1, X_2) = \int_0^\tau k_1(X_1, X_2, t)\, dP^{(2)}(t)$, by Fubini.

Arguing in this way shows that $|D| = o_P(n^{-1/2})$. □

APPENDIX A: GENERALIZED M-ESTIMATORS

Given a model $\mathscr{P}$ with an identifiable parameter β, regard β as a function from $\mathscr{P}$ to $\mathbb{R}^m$. Consider W: $\mathbb{R}^m \times \mathscr{Q} \to \mathbb{R}^m$ such that $W(\beta(P), P) = 0$ for all $P \in \mathscr{P}$. A reasonable estimator, $\beta(\mathbb{P}_n)$, may be defined by $W(\beta(\mathbb{P}_n), \mathbb{P}_n) = 0$. Asymptotic properties of such an estimator may be studied by applying a one-step Taylor expansion to the implicitly defined functional β.

Under conditions (A1)–(A5) below, Bickel, Klaassen, Ritov and Wellner (1993), Theorem 7.2.4, prove such a result. Their proof relies heavily on earlier results of Brown (1985) and Ritov (1987).

(A1) There exists β: $\mathscr{Q} \to \mathbb{R}^m$ such that $\beta(P)$ satisfies $W(\beta(P), P) = 0$ for all $P \in \mathscr{P}$.

(A2) With probability tending to 1, $W(\cdot, \mathbb{P}_n)$ satisfies:

(i) $W(\cdot, \mathbb{P}_n)$ is monotone [i.e., for all $u \in \mathbb{R}^n$, $t \in N$, the map $\lambda \to u'W(t + \lambda u, \mathbb{P}_n)$ is increasing].

(ii) $W(\beta(\mathbb{P}_n), \mathbb{P}_n) = 0$ for a unique $\beta(\mathbb{P}_n) \in N$, where N is some open convex subset of $\mathbb{R}^m$.

(A3) $W(\cdot, P) = (W_1(\cdot, P), \dots, W_m(\cdot, P))'$ is differentiable, and

$$\dot{W}(P) := \left[\frac{\partial W_i}{\partial \beta_j}(\beta(P), P)\right]_{m \times m}$$

is nonsingular.

(A4) Set $\beta_0 := \beta(P)$. For each fixed $\gamma \in \mathbb{R}^m$,

$$\sqrt{n}\left(W(\beta_0 + n^{-1/2}\gamma, \mathbb{P}_n) - W(\beta_0, \mathbb{P}_n)\right) = \dot{W}(P)\gamma + o_p(1).$$

(A5) $\sqrt{n}\, W(\beta_0, \mathbb{P}_n) = \sqrt{n} \int \psi(x, P)\, d\mathbb{P}_n + o_p(1)$, where

$$\psi \in L_2^0(P) := \left\{\psi(\cdot, P): \int \psi(x, P)\, dP(x) = 0, \int \psi\psi'(x, P)\, dP(x) < \infty\right\}.$$

THEOREM A.1 [Bickel, Klaassen, Ritov and Wellner (1993), 7.2.4]. *Under* (A1)–(A5) *the generalized M-estimator* $\hat{\beta}_n$ *implicitly defined by*

$$W(\hat{\beta}_n, \mathbb{P}_n) = 0$$

exists and is unique with probability converging to 1, *and when it exists, it is asymptotically linear with influence function* $-\dot{W}^{-1}(P)\psi(\cdot, P)$.

REMARK. If (A1)–(A5) hold for all $P \in \mathscr{P}$, then the result holds whenever $X_1, \dots, X_n$ are iid P for any $P \in \mathscr{P}$.

The following lemmas will be useful.

LEMMA A.1. (A4) *is implied by* (A4′):

(A4′) *For each* $\gamma \in \mathbb{R}^m$,

$$\sup\left\{|\dot{W}(\beta, \mathbb{P}_n) - \dot{W}(\beta_0, P)|: |\beta - \beta_0| \le n^{-1/2}\gamma\right\} \to_p 0,$$

where $W(\cdot, \mathbb{P}_n) = (W_1(\cdot, \mathbb{P}_n), \dots, W_m(\cdot, \mathbb{P}_n))'$ *and*

$$\dot{W}(\beta, \mathbb{P}_n) := \left[\frac{\partial W_i}{\partial \beta_j}(\beta, \mathbb{P}_n)\right]_{m \times m}.$$

PROOF. For each fixed n,

$$\begin{aligned}&\sqrt{n}\left(W(\beta_0 + n^{-1/2}\gamma, \mathbb{P}_n) - W(\beta_0, \mathbb{P}_n)\right)\\ &\quad = \dot{W}(\boldsymbol{\beta}^*, \mathbb{P}_n)\sqrt{n}\left(\beta_0 + n^{-1/2}\gamma - \beta_0\right) = \dot{W}(\boldsymbol{\beta}^*, \mathbb{P}_n)\gamma,\end{aligned}$$

where

$$\dot{W}(\boldsymbol{\beta}^*, \mathbb{P}_n) := \left[\frac{\partial W_i}{\partial \beta_j}(\beta_i^*, \mathbb{P}_n)\right]_{m\times m}$$

and each β_i^* lies on the line segment between β_0 and $\beta_0 + n^{-1/2}\gamma$.
Now, by (A4′), $\dot{W}(\boldsymbol{\beta}^*, \mathbb{P}_n) = \dot{W}(\beta_0, P) + o_p(1)$. Hence (A4) holds. □

LEMMA A.2. (A2)(i) *is satisfied if*:

(A2)(i′) *With probability tending to* 1, $\dot{W}(\beta, \mathbb{P}_n)$ *is positive semidefinite for all* β.

PROOF.

$$\begin{aligned} u'W(t+\lambda u, \mathbb{P}_n) - u'W(t+\lambda^0 u, \mathbb{P}_n) &= u'\dot{W}(t+\lambda^* u, \mathbb{P}_n)(\lambda - \lambda^0)u \\ &= u'\dot{W}(\beta^*, \mathbb{P}_n)u(\lambda - \lambda^0) \quad \text{for some } \beta^*. \end{aligned}$$

Hence $u'W(t + \lambda u, \mathbb{P}_n)$ is increasing (nondecreasing) in λ if $\dot{W}(\beta, \mathbb{P}_n)$ is positive (semi-) definite for all β. □

APPENDIX B: V-STATISTICS

The following is a brief summary of V-statistics for the proof of Theorem 4.1. Proofs of the results stated here and further details may be found in Serfling (1980), Chapter 5 and subsection 6.3.2.

DEFINITION 1. A V-statistic, V_n, is a functional of the data $X_i, \dots, X_n$, which can be written in the form

$$V_n = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m+1}^{n} \tilde{v}_n(X_{i_1}, \dots, X_{i_m}).$$

$\tilde{v}_n(\cdot)$ is called the kernel of the V-statistic.

Suppose that $\int \cdots \int \tilde{v}_n(x_1, \dots, x_m) \prod_{i=1}^m dP^{\alpha_i}(x_i)$ exists for all values of $x_1, \dots, x_m$ and for all 2^m vectors, α, of 0's and 1's. Then, if $X_1, \dots, X_n$ are iid P, one may rewrite the V-statistic in terms of a centered kernel, v_n:

$$V_n = \int \cdots \int v_n(x_1, \dots, x_m) \prod_{i=1}^{m} d(\mathbb{P}_n - P)(x_i).$$

LEMMA B.1 [Based on Lemma 6.3.2.B of Serfling (1980)]. *Suppose that* $E_P[v_n^2(X_{i_1}, \dots, X_{i_m})] = O(n^\alpha)$ *for all* $1 \le i_1, \dots, i_m \le m$ *and some* $\alpha \ge 0$. *Then*

$$\text{(B.1)} \quad E_P\left[\left(\int \cdots \int v_n(x_1, \dots, x_m) \prod_{i=1}^{m} d(\mathbb{P}_n - P)(x_i)\right)^2\right] = O(n^{-m+\alpha}).$$

COROLLARY B.1. *Suppose that* $E_P[v_n^2(X_{i_1},\dots,X_{i_m})] = O(n^\alpha)$ *for all* $1 \le i_1,\dots,i_m \le m$ *and some* $\alpha \ge 0$. *Then*

$$\int \cdots \int v_n(x_1,\dots,x_m) \prod_{i=1}^{m} d(\mathbb{P}_n - P)(x_i) = O_p(n^{(\alpha-m)/2}). \tag{B.2}$$

PROOF. (B.2) follows immediately from (B.1) by applying Chebyshev's inequality. □

REMARK. The lemma is a trivial extension of Serfling's result to allow a sequence of v_n. Such an extension is only used in order to study estimators not restricted to a finite time interval $[0,\tau]$ for $\tau < \infty$.

Acknowledgments. Much of this research was done while the author was a Ph.D. student at the University of Washington, under the supervision of Professor Jon Wellner.

REFERENCES

ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067. [Correction (1987) *Ann. Probab.* **15** 428–430.]

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1121.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

BEDNARSKI, T. (1989). On the sensitivity of Cox's estimator. *Statist. Decision* **7** 215–228.

BEDNARSKI, T. (1991). Robust estimation in Cox's regression model. Preprint, Inst. Mathematics, Polish Academy of Sciences.

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BENNETT, G. (1962). Probability inequalities for sums of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

BICKEL, P. J. (1981). Quelques aspects de la statistique robuste. *École d'Été de Probabilités de St. Flour. Lecture Notes in Math.* **876** 2–68. Springer, Berlin.

BICKEL, P. J. (1984). Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12** 1349–1368.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

BROWN, B. M. (1985). Multiparameter linearization theorems. *J. Roy. Statist. Soc. Ser. B* **47** 323–331.

CUZICK, J., COOPER, E. H. and MACLENNAN, I. C. M. (1985). The prognostic value of serum β2 microglobulin compared with other presentation features in myelomatosis. *British Journal of Cancer* **52** 1–6.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika* **52** 203–223.

JACOBSEN, M. (1989). Existence and unicity of MLE's in discrete exponential family distributions. *Scand. J. Statist.* **16** 335–349.
JONES, M. P. and CROWLEY, J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics* **45** 157–170.
JONES, M. P. and CROWLEY, J. (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Ann. Statist.* **18** 1203–1220.
LAI, T. L. and YING, Z. (1988). Stochastic integrals of empirical-type processes with application 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. and WEI, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84** 1074–1078.
ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic, New York.
PETO, R. and PETO, J. (1972). Asymptotic efficient rank invariant test procedures. *J. Roy. Statist. Soc. Ser. A* **135** 185–206.
POLLARD, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235–248.
PRENTICE, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73** 1–11.
RAO, R. R. (1963). The law of large numbers for $D[0,1]$-valued random variables. *Theory Probab. Appl.* **8** 70–74.
RITOV, Y. (1987). Tightness of monotone random fields. *J. Roy. Statist. Soc. Ser. B* **49** 331–333.
RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales, and the Cox model. *Contemp. Math.* **80** 191–219. Amer. Math. Soc., Providence, RI.
SAMUELS, S. (1978). Robustness of survival estimates. Ph.D. Thesis, Dept. Biostatistics, Univ. Washington.
SASIENI, P. (1993). Maximum weighted partial likelihood estimates for the Cox model. *J. Amer. Statist. Assoc.* **88** 144–152.
SASIENI, P. D. (1989). Beyond the Cox model: extensions of the model and alternative estimates. Ph.D. Thesis, Dept. Biostatistics, Univ. Washington.
SASIENI, P. D. (1992). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54** 617–635.
SELF, S. G. and PRENTICE, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16** 64–81.
SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
STRUTHERS, C. A. and KALBFLEISCH, J. D. (1986). Misspecified proportional hazards models. *Biometrika* **73** 363–369.
TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.
YING, Z. (1991). A large sample study of rank estimation for censored regression data. Preprint, Dept. Statistics, Univ. Illinois.

DEPARTMENT OF MATHEMATICS, STATISTICS
AND EPIDEMIOLOGY
IMPERIAL CANCER RESEARCH FUND
P.O. BOX 123
LINCOLN'S INN FIELDS
LONDON WC2A 3PX
UNITED KINGDOM