

Asymptotic Theory for Weighted Least Squares Estimators in Aalen's Additive Risk Model

IAN W. MCKEAGUE

Abstract. Let $h(t|Z_i)$ be the conditional hazard function for the survival time of an individual i given the p -dimensional covariate process $Z_i(t)$. We study inference for Aalen's additive risk model $h(t|Z_i) = Z_i(t)'\alpha(t)$, where α is a p -vector of unknown hazard functions. The theory of counting processes is used to obtain weak convergence results for weighted least squares estimators of the hazard functions and the cumulative hazard functions based on continuous data. Results for weighted least squares estimators based on grouped data are also described.

1. Introduction.

The proportional hazards regression model of Cox (1972) for the analysis of censored survival data has had considerable influence on the theory and practice of biostatistics. In recent years this has led to the study of a wide variety of hazard function based regression models which generalize Cox's model in some way. For a comprehensive list of references to such work see the paper of Ritov and Wellner (1987, in these proceedings).

Let $h(t|Z_i)$ denote the conditional hazard function for the survival time T_i of an individual i given the covariate process $Z_i(t) = (Z_{i1}(t), \dots, Z_{ip}(t))'$, $t \geq 0$. The most general model for $h(t|Z_i)$ that seems to be amenable to statistical analysis is

$$h(t|Z_i) = \alpha(t, Z_i(t)), \quad (1.1)$$

where α is a completely general (unknown) function of time and the state of the covariate process. Inference for this model has been studied by Beran (1981) and Dabrowska (1987a, 1987b) in the case of time-independent covariate, and by McKeague and Utikal (1987) in the case of time-dependent covariate process. Although this model is attractive from a theoretical standpoint, in that it can encompass goodness of fit tests for any particular model, its large sample size requirements make it difficult to apply in practice.

Research supported by the Army Research Office under Grant No. DAAL03-86-K-0094
1980 *Mathematics subject classifications*. Primary: 62G05, 62M99; Secondary: 62P10

© 1988 American Mathematical Society
0271-4132/88 \$1.00 + \$.25 per page

If we assume that $\alpha(t, 0) = 0$ and ignore all terms higher than first order in Taylor's expansion of $\alpha(t, z)$ about $z = 0$, then the model (1.1) reduces to Aalen's (1980) additive risk model:

$$h(t|Z_i) = \sum_{j=1}^p \alpha_j(t) Z_{ij}(t), \quad (1.2)$$

where $\alpha_1, \dots, \alpha_p$ are unknown functions of time. Aalen's model, which we study in this paper, is capable of providing information concerning the temporal influence of each covariate not possible to obtain using Cox's model, yet it does not require the extremely large sample size needed for fitting the general model (1.1). McKeague (1986) studied estimation for $\alpha = (\alpha_1, \dots, \alpha_p)'$ using the method of sieves. Recently, Huffer and McKeague (1987) proposed various weighted least squares estimators for α and its integrated counterpart $A(\cdot) = \int_{t_0}^{\cdot} \alpha(s) ds$. The purpose of the present paper is to establish weak convergence results for such weighted least squares estimators in the case of continuous data. Weak convergence results for grouped data based weighted least squares estimators are given in McKeague (1988). In the grouped data case only the total number of uncensored survival times falling in successive time intervals and the corresponding total times at risk, for all levels of the covariates, are assumed to be available, whereas in the continuous data case treated here the exact values of the uncensored survival times T_1, \dots, T_n are assumed to be known. Not surprisingly, better results can be obtained in the continuous data case.

In Section 2 we describe the counting process formulation of Aalen's model. The weighted least squares estimators based on continuous data are defined in Section 3.1 and compared with their grouped data analogues in Section 3.2. Proofs of the main weak convergence results, stated in Section 3.1, are given in Section 4.

2. Aalen's model in the counting process framework.

Suppose that the observable portion of the i th individual's lifetime T_i is given by $\tilde{T}_i = \min(T_i, C_i)$, where C_i is conditionally independent of T_i given the covariate process Z_i . Also suppose that T_i and C_i are absolutely continuous. The observations consist of i.i.d. triples $(\tilde{T}_i, \delta_i, Y_i)$, $i = 1, \dots, n$, where $\delta_i = I(T_i \leq C_i)$ and Y_i is the process $Y_i(t) = Z_i(t)I(\tilde{T}_i \geq t)$. Now let $N_i(t)$ denote the indicator of an uncensored failure for individual i prior to time t :

$$N_i(t) = I(\tilde{T}_i \leq t, \delta_i = 1),$$

and suppose that each covariate process Z_i is left-continuous with right-limits. Under Aalen's model (1.2) the counting process N_i has intensity

$$\lambda_i(t) = \sum_{j=1}^p \alpha_j(t) Y_{ij}(t) \quad (2.1)$$

with respect to the right-continuous filtration $\mathcal{F}_t = \sigma(N_i(s), Y_{ij}(s-), 0 \leq s \leq t, i \geq 1, j = 1, \dots, p)$. Also, no two of the counting processes N_1, \dots, N_n jump simultaneously.

More generally, let $N(t) = (N_1(t), \dots, N_n(t))'$, $t \in [0, 1]$ be a multivariate counting process with respect to a right-continuous filtration (\mathcal{F}_t) , i.e. N is adapted to the filtration and has components N_i which are right-continuous step functions, zero at time zero, with jumps of size +1 such that no two components jump simultaneously. Let Λ be the compensator of N , so that $N = \Lambda + M$, where $M = (M_1, \dots, M_n)'$ and M_1, \dots, M_n are local martingales. Suppose that Λ is absolutely continuous (a.s.): $\Lambda(t) = \int_0^t \lambda(s) ds$, where $\lambda = (\lambda_1, \dots, \lambda_n)'$ and $\lambda_1, \dots, \lambda_n$ are nonnegative predictable processes (intensity processes). The counting process version of Aalen's model is given by

$$\lambda(t) = Y(t)\alpha(t), \quad (2.2)$$

where $\alpha = (\alpha_1, \dots, \alpha_p)'$ is a vector of unknown nonrandom integrable functions and $Y(t) = (Y_{ij}(t))$ is an $n \times p$ matrix of covariate processes assumed to be predictable and locally bounded. For fixed t_0 , $0 \leq t_0 \leq 1$, denote $A(t) = \int_{t_0}^t \alpha(s) ds$, where $t_0 \leq t \leq 1$. The statistical problem is to estimate α and A .

3. The weighted least squares estimators.

3.1. The continuous data case.

Suppose that the entire sample paths of the process N and Y are observed over $[0, 1]$. Aalen (1980) proposed estimators \hat{A} of A of the form

$$\hat{A}(t) = \int_{t_0}^t Y^-(s) dN(s), \quad (3.1)$$

where $Y^-(s)$ is a predictable generalized inverse of $Y(s)$. In the case $p = 1$ with

$$(Y^-(s))_{1i} = \left(\sum_{k=1}^n Y_{k1}(s) \right)^{-1}, \quad i = 1, \dots, n$$

(where $1/0 \equiv 0$) \hat{A} is the Nelson-Aalen estimator for which a general asymptotic theory was derived by Aalen (1978). For $p > 1$ Aalen suggested using $Y^-(s) = (Y'(s)Y(s))^{-1}Y'(s)$, where here and in the sequel, for any square matrix (or scalar) D , D^{-1} denotes the inverse of D if D is invertible, the zero matrix otherwise. Aalen observed that this choice of Y^- can be motivated by a formal least squares principle and that the resulting estimator

$$\hat{A}(t) = \int_{t_0}^t (Y'(s)Y(s))^{-1}Y'(s) dN(s), \quad (3.2)$$

referred to as *Aalen's least squares estimator*, probably gives reasonable but not optimal estimates of A . Recently Huffer and McKeague (1987) suggested using the following generalized inverse of $Y(s)$:

$$Y^-(s) = (Y'(s)\hat{W}(s)Y(s))^{-1}Y'(s)\hat{W}(s), \quad (3.3)$$

where $\hat{W}(t)$ is the $n \times n$ diagonal matrix with i th diagonal entry $\hat{W}_i(t) = (\hat{\lambda}_i(t))^{-1}$ and

$$\hat{\lambda}_i(t) = \sum_{j=1}^p \hat{\alpha}_j(t) Y_{ij}(t), \quad (3.4)$$

where $\hat{\alpha}_j$ is a predictable estimator of α_j . The estimator $\hat{\alpha}_j$ is taken to be the j th component of the *smoothed least squares estimator*

$$\hat{\alpha}(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) d\hat{A}(s), \quad (3.5)$$

where K is a left-continuous bounded kernel function having integral 1, support $[0, 1]$ and $b_n > 0$ is a bandwidth parameter. The choice of generalized inverse (3.3) defines what we call the *weighted least squares estimator*

$$\tilde{A}(t) = \int_{t_0}^t (Y'(s)\hat{W}(s)Y(s))^{-1} Y'(s)\hat{W}(s) dN(s). \quad (3.6)$$

Observe that in the case of a single covariate the weighted least squares estimator coincides with the Nelson-Aalen estimator and no estimate of the weights $\hat{W}_i(t)$ is needed.

A heuristic explanation for the choice of weight matrix $\hat{W}(t)$ is as follows. By conditioning on the past \mathcal{F}_t we may interpret the stochastic differential equation $dN(t) = Y(t)\alpha(t)dt + dM(t)$, increment by increment, as a standard linear regression model with heteroscedastic errors $dM_i(t), i = 1, \dots, n$. We should choose the weight matrix to be proportional to the inverse of the error covariance matrix $\text{Cov}(dM(t)|\mathcal{F}_t) =$ the $n \times n$ diagonal matrix with i th diagonal entry $\lambda_i(t)dt$. However $\lambda_i(t)$ depends on the unknown $\alpha(t)$. Estimating $\lambda_i(t)$ by (3.4), where the estimator $\hat{\alpha}(t)$ only depends on the past (since the kernel K has support $[0, 1]$), leads to $\hat{W}(t)$.

In order to obtain an estimator of α itself we can smooth \tilde{A} to obtain a *smoothed weighted least squares estimator*

$$\tilde{\alpha}(t) = \frac{1}{\tilde{b}_n} \int_{t_0}^t \tilde{K}\left(\frac{t-s}{\tilde{b}_n}\right) d\tilde{A}(s), \quad t_0 < t < 1, \quad (3.7)$$

where \tilde{K} is a bounded kernel function having integral 1, support $[-1, 1]$ and $\tilde{b}_n > 0$ is a bandwidth parameter. In the case of a single covariate the smoothed weighted least squares estimator coincides with the kernel estimator introduced by Ramlau-Hansen (1983).

Let $D[t_0, 1]^p$ denote the product of p copies of the Skorohod space $D[t_0, 1]$ and endow it with the Skorohod product topology. Also denote

$$\tilde{L}_{jk}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t) Y_{ik}(t),$$

$$\tilde{R}_{jke}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t) Y_{ik}(t) Y_{ie}(t),$$

$$\tilde{V}_{jk}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t) Y_{ik}(t) \lambda_i^{-1}(t).$$

Our first result, which gives the asymptotic distribution of Aalen's least squares estimator (3.2), employs the following conditions:

- (A1) (Asymptotic stability). For $j, k, \ell = 1, \dots, p$ there exist continuous functions L_{jk} and $R_{jk\ell}$ defined on $[0, 1]$ such that

$$\sup_{t \in [0, 1]} |\tilde{L}_{jk}(t) - L_{jk}(t)| \xrightarrow{P} 0,$$

$$\sup_{t \in [0, 1]} |\tilde{R}_{jk\ell}(t) - R_{jk\ell}(t)| \xrightarrow{P} 0.$$

- (A2) (Lindeberg condition). For each $j = 1, \dots, p$

$$n^{-\frac{1}{2}} \sup_{\substack{i=1, \dots, n \\ t \in [0, 1]}} |Y_{ij}(t)| \xrightarrow{P} 0.$$

- (A3) (Asymptotic nondegeneracy condition). The $p \times p$ matrix $L(t) = (L_{jk}(t))$ in (A1) is nonsingular for all $t \in [0, 1]$.

THEOREM 3.1. *Let $t_0 = 0$. Under conditions (A1)-(A3)*

$$\sqrt{n}(\hat{A} - A) \xrightarrow{D} m \text{ in } D[0, 1]^p$$

where m is a p -variate continuous Gaussian martingale with mean zero and covariance function

$$\text{Cov}(m_j(t), m_k(t)) = \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \int_0^t R_{uvw}(s) (L^{-1}(s))_{jv} (L^{-1}(s))_{kw} \alpha_u(s) ds.$$

In order to establish our weak convergence results for the weighted least squares estimator (3.6) and the smoothed weighted least squares estimator (3.7) we need the following additional conditions:

- (B1) (Asymptotic stability). For $j, k = 1, \dots, p$ there exist continuous functions V_{jk} defined on $[0, 1]$ such that

$$\sup_{t \in [0, 1]} |\tilde{V}_{jk}(t) - V_{jk}(t)| \xrightarrow{P} 0.$$

- (B2) (Bounded covariates). The processes $Y_{ij}, i \geq 1, j = 1, \dots, p$ are uniformly bounded.

- (B3) (Asymptotic nondegeneracy condition). The $p \times p$ matrix $V(t) = (V_{jk}(t))$ in (B1) is nonsingular for all $t \in [0, 1]$.

- (B4) (Intensity regularity condition). There exists $\delta > 0$ such that if $Y_{ij}(t) \neq 0$ for some $j = 1, \dots, p$ then $\lambda_i(t) \geq \delta$.

- (B5) The functions $\alpha_1, \dots, \alpha_p$ are continuous.

THEOREM 3.2. *Suppose that conditions (A1), (A3), (B1) - (B5) hold, $b_n \rightarrow 0$, $nb_n^2 \rightarrow \infty$ and the kernel function K has bounded variation. Let $0 < t_0 < 1$. Then*

$$\sqrt{n}(\tilde{A} - A) \xrightarrow{D} m' \text{ in } D[t_0, 1]^p$$

where m' is a p -variate continuous Gaussian martingale with mean zero and covariance function

$$\text{Cov}(m'_j(t), m'_k(t)) = \int_{t_0}^t (V^{-1}(s))_{jk} ds.$$

THEOREM 3.3. *Suppose that conditions (A1), (A3), (B1) - (B4) hold, $b_n \rightarrow 0$, $nb_n^2 \rightarrow \infty$, $n\tilde{b}_n \rightarrow \infty$, $n\tilde{b}_n^3 \rightarrow 0$ and $\alpha_1, \dots, \alpha_p$ have bounded derivatives in a neighbourhood of t , where $0 < t_0 < t < 1$. Then*

$$(n\tilde{b}_n)^{\frac{1}{2}} (\tilde{\alpha}(t) - \alpha(t))$$

converges in distribution to a p -dimensional normal distribution with mean zero and covariance matrix $V^{-1}(t) \int_{-1}^1 \tilde{K}^2(s) ds$.

In order to apply the weak convergence results, to obtain hypothesis tests and confidence bands for instance, it is necessary to estimate the covariances of the limiting Gaussian distributions. Using Lengart's inequality it can be shown that

$$n \sum_{i=1}^n \int_{t_0}^t (Y^-(s))_{ji} (Y^-(s))_{ki} dN_i(s)$$

is a uniformly consistent estimator of $\text{Cov}(m'_j(t), m'_k(t))$ over $[t_0, 1]$, where $Y^-(s)$ is given by (3.3). Lemma 4.3(c) provides a uniformly consistent estimator of the matrix function $V^{-1}(t)$ over $[t_0, 1]$.

Note that we have restricted weighted least squares estimation to the time interval $[t_0, 1]$ with $0 < t_0 < 1$. The reason for this is that the smoothed least squares estimator $\hat{\alpha}$ is not necessarily a uniformly consistent estimator of α over the whole of $[0, 1]$, but only $[t_0, 1]$, $0 < t_0 < 1$; see Lemma 4.2. Uniform consistency of $\hat{\alpha}$ is needed to ensure that the inaccuracy introduced by approximating the "true" weights $W_i(t) = (\lambda_i(t))^{-1}$ by $\hat{W}_i(t)$ becomes uniformly negligible as $n \rightarrow \infty$.

3.2. The grouped data case.

Suppose that the data are grouped into d_n time intervals $I_r^{(n)}$, $r = 1, \dots, d_n$ which partition $[0, 1]$ and depend on the sample size n . It is natural to assume that the hazard function $\alpha_1, \dots, \alpha_p$ are constant over each interval, giving rise to a sequence of Aalen models indexed by n . The hazard functions $\alpha_1^{(n)}, \dots, \alpha_p^{(n)}$ in the n th model are constrained to be the piecewise constant approximations to fixed underlying hazard functions $\alpha_1, \dots, \alpha_p$:

$$\alpha_j^{(n)}(t) = \frac{1}{\ell_r} \int_{I_r^{(n)}} \alpha_j(s) ds \text{ for } t \in I_r^{(n)},$$

where $\ell_r = \ell_r^{(n)}$ is the length of $I_r^{(n)}$.

Assume that the covariate processes Z_i are time-independent and the total time at risk and number of uncensored failures are known for each interval and covariate level. Least squares and weighted least squares estimators of $\alpha^{(n)} = (\alpha_1^{(n)}, \dots, \alpha_p^{(n)})'$ based on such data are given by

$$\hat{\alpha}(t) = \left(\int_{I_r^{(n)}} Y'(s)Y(s)ds \right)^{-1} \int_{I_r^{(n)}} Y'(s)dN(s), \quad t \in I_r^{(n)}$$

$$\hat{\alpha}(t) = \left(\int_{I_r^{(n)}} Y'(s)\hat{W}(s)Y(s)ds \right)^{-1} \int_{I_r^{(n)}} Y'(s)\hat{W}(s)dN(s), \quad t \in I_r^{(n)}$$

respectively, where $\hat{W}(t)$ is the $n \times n$ diagonal matrix having i th diagonal entry $\hat{W}_i(t) = (\hat{\lambda}_i(t))^{-1}$ and $\hat{\lambda}_i(t)$ is an estimate of the intensity

$$\lambda_i^{(n)}(t) = \sum_{j=1}^p \alpha_j^{(n)}(t)Y_{ij}^{(n)}(t).$$

It is reasonable to expect that if $\hat{\lambda}_i(t)$ is chosen appropriately and the mesh of the partition $I_1^{(n)} \dots I_{d_n}^{(n)}$ tends to zero at a suitable rate as $n \rightarrow \infty$, then the estimate $\tilde{A}(\cdot) = \int_{t_0}^{\cdot} \tilde{\alpha}(s)ds$ satisfies a functional central limit theorem analogous to Theorem 3.2.

The estimator $\hat{\lambda}_i(t)$ is taken to be

$$\hat{\lambda}_i(t) = \sum_{j=1}^p \alpha_j^{\dagger*}(t)Y_{ij}(t),$$

where $\alpha^{\dagger*} = (\alpha_1^{\dagger*}, \dots, \alpha_p^{\dagger*})'$ is the piecewise constant approximation to the smoothed least squares estimator

$$\alpha^{\dagger}(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) \hat{\alpha}(s)ds,$$

where K is a left-continuous kernel function having integral 1, support $(0, 1]$ and $b_n > 0$ is a bandwidth parameter. Suppose that the intervals are chosen so that $t_0, 0 < t_0 < 1$, is always a boundary point of one of them. Define $\bar{A}(\cdot) = \int_0^{\cdot} \alpha^{(n)}(s)ds$, the piecewise linear approximation to A .

We require the following conditions:

- (C1) (Asymptotic stability). For $j, k, \ell = 1, \dots, p$ there exist functions L_{jk} , $R_{jk\ell}$ and V_{jk} defined on $[0, 1]$ such that

$$\sup_{t \in [0, 1]} |\tilde{L}_{jk}(t) - L_{jk}(t)| = o_P(\sqrt{\min(\ell_1, \dots, \ell_d)})$$

$$\sup_{t \in [0, 1]} |\tilde{R}_{jk}(t) - R_{jk}(t)| \xrightarrow{L^1} 0$$

$$\sup_{t \in [0,1]} |\tilde{V}_{jk}(t) - V_{jk}(t)| = O_P(\sqrt{\min(\ell_1, \dots, \ell_d)}).$$

(C2) The functions $\alpha_j, L_{jk}, V_{jk}, j, k = 1, \dots, p$ are Lipschitz.

(C3) The bandwidth parameter b_n and the interval lengths ℓ_1, \dots, ℓ_d satisfy

$$b_n^2 n \min(\ell_1, \dots, \ell_d) \rightarrow \infty$$

$$\frac{b_n}{\sqrt{\min(\ell_1, \dots, \ell_d)}} \rightarrow 0$$

$$\frac{b_n}{\min(\ell_1, \dots, \ell_d)} \rightarrow \infty.$$

PROPOSITION 3.1. (*Grouped data case*). *Suppose that conditions (A3), (B2)-(B4), (C1)-(C3) hold and the kernel function K has bounded variation. Let $0 < t_0 < 1$. Then $\sqrt{n}(\tilde{A} - \bar{A})$ converges weakly in $D[t_0, 1]^p$ to the p -variate Gaussian martingale m' of Theorem 3.2.*

We refer to McKeague (1988) for a proof of this proposition. Note that the estimators $\hat{\alpha}, \tilde{\alpha}, \tilde{A}$ require the total time at risk in each interval at each covariate level to be available. When only interval count data are available, survival analytic techniques do not apply. In that case, contingency table techniques provide an alternative approach, see Koch, Johnson and Tolley (1972), Bishop, Fienberg and Holland (1975), Gilula (1986) and Kiefer (1987).

A grouped data based estimator of the covariance matrix of the limiting Gaussian martingale m' is given in McKeague (1988). Although it is possible to obtain a pointwise weak convergence result for the least squares estimator $\hat{\alpha}(t)$, we have not been able to do so for the weighted least squares estimator $\tilde{\alpha}(t)$ or any smoothed version of it, so a grouped data analogue of Theorem 3.3 is not yet available.

4. Proofs of Theorems 3.1-3.3.

The following lemma, analogous to Lemma 4.2 of McKeague (1988), is stated without proof.

LEMMA 4.1. *Under conditions (A1) and (A3)*

(a) $P(\tilde{L}(t) \text{ is invertible for all } t \in [0, 1]) \rightarrow 1,$

(b) $\sup_{t \in [0,1]} \|\tilde{L}^{-1}(t) - L^{-1}(t)\| \xrightarrow{P} 0,$ where $\|\cdot\|$ denotes operator norm.

PROOF OF THEOREM 3.1. From (3.2) and (2.2) we can write

$$\sqrt{n}(\hat{A}(t) - A(t)) = X^{(n)}(t) - \sqrt{n} \int_0^t J(s) dA(s), \quad (4.1)$$

where

$$X^{(n)}(t) = \frac{1}{\sqrt{n}} \int_0^t \tilde{L}^{-1}(s) Y'(s) dM(s),$$

$J(t) = I(\tilde{L}(t))$ is not invertible).

By Lemma 4.1 and conditions (A1), (A3), the last term on the r.h.s. of (4.1) converges uniformly to zero in probability. It remains to show that $X^{(n)} \xrightarrow{D} m$ in $D[0, 1]^p$. By Lemma 4.1 there exists a constant $C > 0$ such that

$$P(\hat{L}^{-1}(t) = U(t) \text{ for all } t \in [0, 1]) \rightarrow 1 \quad (4.2)$$

as $n \rightarrow \infty$, where $U(t) = (U_{jk}(t))$ is the $p \times p$ matrix with entries

$$U_{jk}(t) = \begin{cases} (\hat{L}^{-1}(t))_{jk} & \text{if } |(\hat{L}^{-1}(t))_{jk}| \leq C \\ C & \text{otherwise.} \end{cases}$$

Define the process

$$\tilde{X}^{(n)}(t) = \frac{1}{\sqrt{n}} \int_0^t U(s) Y'(s) dM(s)$$

and note that

$$\sup_{t \in [0, 1]} \|\tilde{X}^{(n)}(t) - X^{(n)}(t)\| \xrightarrow{P} 0. \quad (4.3)$$

The j th component of $\tilde{X}^{(n)}$ can be written

$$\tilde{X}_j^{(n)}(t) = \sum_{i=1}^n \int_0^t \tilde{H}_{ij}^{(n)}(s) dM_i(s),$$

where

$$\tilde{H}_{ij}^{(n)}(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^p U_{jk}(t) Y_{ik}(t).$$

The local martingales M_1, \dots, M_n are local square integrable martingales on the time interval $[0, 1]$ and their predictable quadratic variation processes are given by

$$\langle M_i, M_i \rangle(t) = \int_0^t \lambda_i(s) ds \quad \text{and} \quad \langle M_i, M_j \rangle(t) = 0, \quad i \neq j.$$

The process $\tilde{H}_{ij}^{(n)}$ is locally bounded and predictable so that $\tilde{X}_j^{(n)}$ is a local square integrable martingale. The predictable quadratic variation processes of $\tilde{X}_1^{(n)}, \dots, \tilde{X}_p^{(n)}$ are given by

$$\begin{aligned} \langle \tilde{X}_j, \tilde{X}_k \rangle(t) &= \int_0^t \sum_{i=1}^n \tilde{H}_{ij}^{(n)}(s) \tilde{H}_{ik}^{(n)}(s) \lambda_i(s) ds \\ &= \sum_{u=1}^p \sum_{v=1}^p \sum_{w=1}^p \int_0^t (\hat{L}^{-1}(s))_{jv} (\hat{L}^{-1}(s))_{kw} \tilde{R}_{uvw}(s) \alpha_u(s) ds \\ &\quad + o_P(1) \end{aligned}$$

(by (4.2))

$$\xrightarrow{P} \text{Cov}(m_j(t), m_k(t)) \text{ as } n \rightarrow \infty,$$

by the asymptotic stability condition (A1) and Lemma 4.1. The Lindeberg condition

$$\int_0^1 \sum_{i=1}^n \tilde{H}_{ij}^{(n)}(t)^2 \lambda_i(t) I(|\tilde{H}_{ij}^{(n)}(t)| > \varepsilon) dt \xrightarrow{P} 0$$

as $n \rightarrow \infty$, for each $\varepsilon > 0$, $j = 1, \dots, p$, is a consequence of

$$\sup_{\substack{i=1, \dots, n \\ t \in [0,1]}} |\tilde{H}_{ij}^{(n)}(t)| \leq Cpn^{\frac{1}{2}} \sup_{\substack{i=1, \dots, n \\ t \in [0,1]}} |Y_{ij}(t)|$$

$\xrightarrow{P} 0$ (by condition (A2)).

Thus, by Rebolledo's central limit theorem for local square integrable martingales in the form given by Andersen and Gill (1982, Theorem I.2), it follows that $\tilde{X}^{(n)} \xrightarrow{D} m$ in $D[0, 1]^p$. Combining this with (4.3) we obtain that $X^{(n)} \xrightarrow{D} m$ in $D[0, 1]^p$, which completes the proof.

LEMMA 4.2. *Suppose that conditions (A1)-(A3), (B5) hold, the kernel function K has bounded variation, $b_n \rightarrow 0$ and $nb_n^2 \rightarrow \infty$. Let $0 < t_0 < 1$. Then*

$$\sup_{t \in [t_0, 1]} |\hat{\alpha}_j(t) - \alpha_j(t)| \xrightarrow{P} 0.$$

PROOF. Use integration by parts and Theorem 3.1, cf. the proof of Theorem 2.2 of McKeague (1988).

Let $\hat{V}(t)$ denote the $p \times p$ matrix with entries $\hat{V}_{jk}(t) = \frac{1}{n} \sum_{i=1}^n Y_{ij}(t) Y_{ik}(t) \hat{W}_i(t)$.

LEMMA 4.3. *Suppose that the conditions of Theorem 3.2 hold. Then*

- (a) $P(\hat{V}(t) \text{ is invertible for all } t \in [t_0, 1]) \rightarrow 1$,
- (b) $\sup_{\substack{i=1, \dots, n \\ t \in [t_0, 1]}} |\hat{W}_i(t) - W_i(t)| \xrightarrow{P} 0$,
- (c) $\sup_{t \in [t_0, 1]} \|\hat{V}^{-1}(t) - V^{-1}(t)\| \xrightarrow{P} 0$.

PROOF. Use conditions (B1)-(B5), especially the bounded covariates condition (B2), and Lemma 4.2. The proof is similar to the proof of Lemma 4.3 of McKeague (1988).

PROOF OF THEOREM 3.2. We can write

$$\sqrt{n}(\tilde{A}(t) - A(t)) = X^{(n)}(t) - \sqrt{n} \int_{t_0}^t J(s) dA(s), \tag{4.4}$$

where

$$X_j^{(n)}(t) = \sum_{i=1}^n \int_{t_0}^t H_{ij}^{(n)}(s) dM_i(s),$$

$$H_{ij}^{(n)}(s) = \frac{1}{\sqrt{n}} G_{ij}^{(n)}(s),$$

$$G_{ij}^{(n)}(s) = \sum_{k=1}^p (\hat{V}^{-1}(s))_{jk} Y_{ik}(s) \hat{W}_i(s),$$

$$J(s) = I(\hat{V}(s) \text{ is not invertible}).$$

By Lemma 4.3(a) and conditions (B1), (B3), the last term on the r.h.s. of (4.4) converges uniformly to zero in probability. It remains to show that $X^{(n)} \xrightarrow{D} m'$ in $D[0, 1]^p$. By Lemma 4.3 and conditions (B2), (B4), (B5), there exists a constant $C > 0$ such that

$$P(G_{ij}^{(n)}(s) = \tilde{G}_{ij}^{(n)}(s) \text{ for all } i = 1, \dots, n, s \in [t_0, 1]) \rightarrow 1 \quad (4.5)$$

where

$$\tilde{G}_{ij}^{(n)}(s) = \begin{cases} G_{ij}^{(n)}(s) & \text{if } |G_{ij}^{(n)}(s)| \leq C \\ C & \text{otherwise.} \end{cases} \quad (4.6)$$

Define the processes

$$\tilde{H}_{ij}^{(n)}(s) = \frac{1}{\sqrt{n}} \tilde{G}_{ij}^{(n)}(s), \quad (4.7)$$

$$\tilde{X}_j^{(n)}(t) = \sum_{i=1}^n \int_{t_0}^t \tilde{H}_{ij}^{(n)}(s) dM_i(s). \quad (4.8)$$

Since $\tilde{H}_{ij}^{(n)}$ is predictable and bounded, $\tilde{X}_j^{(n)}$ is a local square integrable martingale. Also

$$\sup_{t \in [t_0, 1]} \|X^{(n)}(t) - \tilde{X}^{(n)}(t)\| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty, \quad (4.9)$$

where $\tilde{X}^{(n)} = (\tilde{X}_1^{(n)}, \dots, \tilde{X}_p^{(n)})'$. The predictable quadratic variation processes of $\tilde{X}_1^{(n)}, \dots, \tilde{X}_p^{(n)}$ are given by

$$\begin{aligned} \langle \tilde{X}_j^{(n)}, \tilde{X}_k^{(n)} \rangle (t) &= \int_{t_0}^t \sum_{i=1}^n \tilde{H}_{ij}^{(n)}(s) \tilde{H}_{ik}^{(n)}(s) \lambda_i(s) ds \\ &= \sum_{u=1}^p \sum_{v=1}^p \int_{t_0}^t (\hat{V}^{-1}(s))_{ju} (\hat{V}^{-1}(s))_{kv} \hat{V}_{uv}(s) ds + o_P(1), \end{aligned} \quad (4.10)$$

where

$$\hat{V}_{uv}(s) = \frac{1}{n} \sum_{i=1}^n Y_{iu}(s) Y_{iv}(s) \lambda_i(s) \hat{W}_i^2(s). \quad (4.11)$$

Thus, by Lemma 4.3 and condition (B2),

$$\begin{aligned} \langle \tilde{X}_j^{(n)}, \tilde{X}_k^{(n)} \rangle (t) &\xrightarrow{P} \sum_{u=1}^p \sum_{v=1}^p \int_{t_0}^t (V^{-1}(s))_{ju} (V^{-1}(s))_{kv} V_{uv}(s) ds \\ &= \int_{t_0}^t (V^{-1}(s))_{jk} ds. \end{aligned}$$

Next, since

$$\sup_{\substack{i=1, \dots, n \\ s \in [0,1]}} |\tilde{H}_{ij}^{(n)}(s)| \leq \frac{C}{\sqrt{n}},$$

the Lindeberg condition

$$\int_{t_0}^1 \sum_{i=1}^n \tilde{H}_{ij}^{(n)}(s)^2 \lambda_i(s) I(|\tilde{H}_{ij}^{(n)}(s)| > \varepsilon) d^P \rightarrow 0 \quad (4.12)$$

as $n \rightarrow \infty$ for each $\varepsilon > 0$, is satisfied. Thus, by Rebolledo's central limit theorem for local square integrable martingales, $\tilde{X}^{(n)} \xrightarrow{D} m'$ in $D[0,1]^p$, and by (4.9), $X^{(n)}$ has the same limit distribution. This completes the proof.

PROOF OF THEOREM 3.3. Define the smoothed version of α :

$$\alpha^*(t) = \frac{1}{\tilde{b}_n} \int_{t_0}^1 \tilde{K}\left(\frac{t-s}{\tilde{b}_n}\right) \alpha(s) ds.$$

Since $\alpha_1, \dots, \alpha_p$ are assumed to have bounded derivatives in a neighbourhood of t ,

$$(n\tilde{b}_n)^{\frac{1}{2}} \|\alpha(t) - \alpha^*(t)\| = O(n\tilde{b}_n^{\frac{3}{2}})^{\frac{1}{2}} \rightarrow 0.$$

Using (2.2), (3.6) and (3.7) and Lemma 4.3(a) we can write

$$(n\tilde{b}_n)^{\frac{1}{2}} (\tilde{\alpha}(t) - \alpha^*(t)) = X^{(n)} + o_P(1),$$

where $X^{(n)}$ is the p -dimensional random vector with j th component

$$X_j^{(n)} = \sum_{i=1}^n \int_{t_0}^1 H_{ij}^{(n)}(s) dM_i(s),$$

$$H_{ij}^{(n)}(s) = \frac{1}{\sqrt{n\tilde{b}_n}} G_{ij}^{(n)}(s),$$

$$G_{ij}^{(n)}(s) = \tilde{K}\left(\frac{t-s}{\tilde{b}_n}\right) \sum_{k=1}^p (\hat{V}^{-1}(s))_{jk} Y_{ik}(s) \hat{W}_i(s).$$

As in the proof of Theorem 3.2 we truncate $G_{ij}^{(n)}$ in order to apply Rebolledo's central limit theorem. This gives new processes $\tilde{G}_{ij}^{(n)}$ and $\tilde{H}_{ij}^{(n)}$ satisfying (4.5)-(4.7). The Lindeberg condition (4.12) is satisfied since

$$\sup_{\substack{i=1, \dots, n \\ s \in [0,1]}} |\tilde{H}_{ij}^{(n)}(s)| = O\left(\frac{1}{\sqrt{n\tilde{b}_n}}\right),$$

and $n\tilde{b}_n \rightarrow \infty$. Since $\alpha_1, \dots, \alpha_p$ are bounded in a neighbourhood I_t of t , by Lemma 4.3,

$$\sup_{s \in I_t} |\tilde{V}_{uv}(s) - V_{uv}(s)| \xrightarrow{P} 0, \quad (4.13)$$

where $\tilde{V}_{uv}(s)$ is defined by (4.11). Now consider

$$\begin{aligned} & \int_{t_0}^1 \sum_{i=1}^n \tilde{H}_{ij}^{(n)}(s) \tilde{H}_{ik}^{(n)}(s) \lambda_i(s) ds \\ &= \frac{1}{\tilde{b}_n} \sum_{u=1}^p \sum_{v=1}^p \int_{t_0}^1 \tilde{K}^2\left(\frac{t-s}{\tilde{b}_n}\right) (\hat{V}^{-1}(s))_{ju} (\hat{V}^{-1}(s))_{kv} \tilde{V}_{uv}(s) ds \\ & \quad + o_P(1) \\ &= \sum_{u=1}^p \sum_{v=1}^p \int_{-1}^1 \tilde{K}^2(x) (\hat{V}^{-1}(t - \tilde{b}_n x))_{ju} (\hat{V}^{-1}(t - \tilde{b}_n x))_{kv} \tilde{V}_{uv}(t - \tilde{b}_n x) dx \\ & \quad + o_P(1) \\ & \xrightarrow{P} \int_{-1}^1 \tilde{K}^2(x) dx (V^{-1}(t))_{jk} \end{aligned}$$

by (4.13) and Lemma 4.3(c). Application of Rebolledo's central limit theorem (see Liptser and Shirayev (1980, Remark 1)) to $X^{(n)}$ gives the result.

REFERENCES

- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
- Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics* **2**, 1-25. Springer-Verlag, New York.
- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.
- Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Tech. Report*, University of California, Berkeley.
- Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practices*, Cambridge, Mass., M.I.T. Press.
- Cox, D. R. (1972). Regression models with life tables (with discussion). *J. Roy. Statist. Soc.* **B34**, 187-220.
- Dabrowska, D. M. (1987a). Nonparametric regression with censored survival time data. To appear in *Scand. J. Statist.*
- Dabrowska, D. M. (1987b). Uniform consistency of nearest neighbour and kernel conditional Kaplan-Meier estimates. *Preprint*, University of California, Berkeley.
- Gilula, Z. (1986). On some similarities between contingency table methods and grouped survival analytic techniques. *J. Stat. Planning and Inference* **14**, 203-211.
- Huffer, F. W. and McKeague, I. W. (1987). Survival analysis using additive risk models. *Preprint*, Department of Statistics, Florida State University.
- Kiefer, N. M. (1987). Analysis of grouped duration data. In these *Proceedings*.
- Koch, G. G., Johnson, W. D. and Tolley, H. D. (1972). A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. *J. Americ. Statist. Assoc.* **67**, 783-796.
- Liptser, R. S. and Shirayev, A. N. (1980). A functional central limit theorem for semimartingales. *Theor. Prob. Appl.* **25**, 667-688.
- McKeague, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14**, 579-589.

- McKeague, I. W. and Utikal, K. J. (1987). Inference for a nonlinear semimartingale regression model. *Preprint*, Department of Statistics, Florida State University.
- McKeague, I. W. (1988). A counting process approach to the regression analysis of grouped survival data. *Stochastic Process. Appl.*, to appear.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.
- Ritov, Y. and Wellner, J. A. (1987). Censoring, martingales, and the Cox model. In these *Proceedings*.

Department of Statistics, Florida State University, Tallahassee, FL 32306-3033