TIME-DEPENDENT COEFFICIENTS IN A COX-TYPE REGRESSION MODEL

by

S. A. Murphy

and

P. K. Sen

Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

Abstract: Estimation of a time-varying coefficient in a Cox-type parametrization of the stochastic intensity of a point process is considered. A sieve estimation procedure (Grenander, 1981) is used to estimate the coefficient. A rate of convergence in probability for the sieve estimator is given and a functional CLT for the integrated sieve estimator is proved.

AMS subject classification: 60G55, 62M09, 62P10

Keywords: Point Process, Cox Regression, Method of Sieves

0. Introduction

Suppose an output or dependent counting process, N and an input or independent covariate process X is observed. A model relating N to X which is often used in survival analysis is the Cox Regression Model (Cox, 1972; Anderson and Gill, 1982). This model stipulates that the stochastic intensity of N is

$$\lambda_t(X) = e^{\beta X_t} \lambda_0(t).$$

In the above, the regression coefficient, β, is an unknown scalar and λ_0 is an unspecified deterministic function. Since β is constant in time, the above model implies that the regression relationship between N and X is stationary. Since this may not be the case, several authors have considered a time-varying regression coefficient (Brown, 1975; Taulbee, 1979; Stablein et al., 1981 and Zucker and Karr, 1988). Brown, Taulbee and Stablein et al. make simplifying assumptions on the form of β so as to maintain a finite dimensional parameter space. Zucker and Karr, using a penalized likelihood technique, allow β to be infinite dimensional (i.e., a function of time). Their analysis is developed within the survival analysis context; that is where N can have at most one jump. The method presented here, which also allows β to be infinite dimensional, utilizes the method of sieves (Grenander, 1981), and in particular, a very simple sieve, the histogram sieve. This choice of a sieve retains the simplicity of analysis present in methods involving only a finite dimensional parameterization of the regression coefficient β. In addition, the estimation method presented below is applicable not only in the survival analysis context, but also in the more general context where N is allowed multiple jumps. The histogram sieve was used by Friedman (1982) in the survival analysis context for the purpose of estimating λ_0. McKeague (1987) and Leskow (1987) also use the histogram sieve for estimation purposes in multiplicative intensity model of Aalen (1978).

Section 1 contains a description of the statistical model with a list of assumptions made in the following theorems. Weak consistency (with a rate of convergence) is proved in Section 2. Next in Section 3, a functional central limit theorem is given for the integrated regression coefficient. Section 4 presents a consistent estimator of the asymptotic variance process. and the last section contains the technical details.

1. Statistical Model

For each n, one observes an n-component multivariate counting process, $\underset{\sim}{N}^n = (N^n(1), \ldots N^n(n))$, over the time interval $[0,T]$. For example, $N^n(i)$ might count certain life events for individual i. $\underset{\sim}{N}^n$ is defined on a stochastic base $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}^n_t : t \in [0,T]\})$ with respect to which $\underset{\sim}{N}^{(n)}$ has stochastic intensity $\underset{\sim}{\lambda}^n = (\lambda^n(1), \ldots \lambda^n(n))$ where

$$\lambda^n_t(i) = \lambda_0(t)\, e^{\beta_0(t) X^n_t(i)}\, Y^n_t(i).$$

In the above, both β_0 and λ_0 are deterministic functions on $[0.T]$, $\underset{\sim}{X}^n = (X^n(1), \ldots X^n(n))$ is a vector of locally bounded, predictable stochastic processes, and $\underset{\sim}{Y}^n = (Y^n(1), \ldots Y^n(n))$ is a vector of predictable stochastic processes each taking values in $\{0,1\}$. In this paper, $\underset{\sim}{N}^n$ having stochastic intensity $\underset{\sim}{\lambda}^n$ implies that

$$M^n_t(i) = N^n_t(i) - \int_0^t \lambda^n_s(i)ds$$

is a local square integrable martingale with predictable variation,

$$\begin{aligned} \langle M^n(i), M^n(j)\rangle_t &= \int_0^t \lambda^n_s(i)ds \qquad \text{for } i=j \\ &= 0 \qquad \text{for } i \neq j. \end{aligned}$$

Since the focus of this paper is on β_0, inference for β_0 is based on the logarithm of Cox's partial likelihood (Cox, 1972),

$$\mathcal{L}_n(\beta) = \sum_{i=1}^n \int_0^T \ell n \left[\frac{e^{\beta(s) X^n_s(i)}}{\sum_{j=1}^n e^{\beta(s) X^n_s(j)} Y^n_s(j)} \right] dN^n_s(i) .$$

A direct maximization of $\mathcal{L}_n(\beta)$ for β will not produce a meaningful estimate. For example, let $\underset{\sim}{X}^n$ be time independent and each component of $\underset{\sim}{N}^n$ have at most one jump, then if $\text{Rank}(X_i^n) = n$ and the jump of $N^n(i)$ occurs at T_0, $\ell n\left[\frac{e^{\beta(T_0)X^n(i)}}{\sum_{j=1}^{n} e^{\beta(T_0)X^n(j)}}\right]$ can be made as large as desired simply by increasing $\beta(T_0)$ (Zucker and Karr, 1988). In this situation, the method of sieves (Grenander, 1981) is often useful. Essentially an increasing sequence of parameter spaces, say $\{\Theta_n, n\geq 1\}$, is given so that within each Θ_n there exists a maximum likelihood estimate, say $\hat{\beta}_n$, and $\bigcup_n \Theta_n$ is dense in Θ, where Θ is the parameter space of interest. The histogram sieve is used here,

$$\Theta_n = \{\beta : \beta(s) = \sum_{i=1}^{K_n} b_i\, I\{s \in I_i^n\} \text{ for } (b_1,\ldots,b_{K_n}) \in \mathbb{R}^{K_n}\}$$

The $(I_1^n,\ldots,I_{K_n}^n)$ are consecutive segments of $[0,T]$.

Defining, for each $s \in [0.T]$,

$$S_n^i(\beta,s) = \frac{1}{n}\sum_{j=1}^{n} e^{\beta(s)X_s^n(j)}(X_s^n(j))^i\, Y_s^n(j)\quad i=0,1,2,3,4,$$

consider the following assumptions:

A. (Asymptotic stability) There exist $S^i(\beta_0,s)$, $i=0,1,2$, such that

1) $\sup_{s\in[0,T]} |S_n^i(\beta_0,s) - S^i(\beta_0,s)| = o_p(1)$,

2) $n \int_0^T (S_n^i(\beta_0,s) - S^i(\beta_0,s))^2\, ds = o_p(1)$, and

3) there exist $\gamma > 0$ such that

$$\sup_{s\in[0,T]} \sup_{\substack{b\in\mathbb{R} \\ |b-\beta_0(s)|<\gamma}} \left|\frac{S_n^i(b,s)}{S_n^0(b,s)}\right| = O_p(1) \qquad i=1,2,3,4.$$

B. (Lindeberg Condition) For all $\epsilon > 0$,

1) $\max_{1\le j\le n} \int_0^T I\{s : |X_s(j)Y_s(j)| > \frac{\epsilon}{2}\sqrt{n}\}\, ds = o_p(1)$.

C. (Asymptotic Regularity)

1) There exist constants $U_1 > 0$, $U_2 > 0$ such that $\max\{\lambda_0(s), S^i(\beta_0,s), i = 0,1,2\} \le U_1$ and $S^0(\beta_0,s) \ge U_2$ a.e. Lebesque on $[0,T]$.

2) There exists a constant $L > 0$ such that for

$$V(\beta_0,s) = \frac{S^2(\beta_0,s)}{S^0(\beta_0,s)} - \left[\frac{S^1(\beta_0,s)}{S^0(\beta_0,s)}\right]^2,$$

$V(\beta_0,s)\, S(\beta_0,s)\, \lambda_0(s) > L$ a.e. Lebesque on $[0,T]$.

D. (Bias)

1) $\beta_0(s)$ is Lipshitz of order 1 on $[0,T]$.

2) $\beta_0(s)$ has bounded second derivative a.e. Lebesque on $[0,T]$.

3) $V(\beta_0,s)\, S^0(\beta_0,s)\, \lambda_0(s)$ is continuous in s on $[0,T]$.

4) $V(\beta_0,s)\, S^0(\beta_0,s)\, \lambda_0(s)$ is Lipshitz of order 1 on $[0,T]$.

In the following section, a member of Θ_n will be denoted either by its' functional form, $\beta(s) = \sum_{i=1}^{K_n} \beta_i I_i(s)$, or by its' vector form, $\beta = (\beta_1, \ldots \beta_{K_n})$. It should be clear from the context which form of β is pertinent. The lengths of the K_n intervals, $I_1^n, \ldots, I_{K_n}^n$, will be denoted by $\ell = (\ell_1^n, \ldots, \ell_{K_n}^n)$ with $\ell_{(1)}^n$, $\ell_{(K_n)}^n$, and $\|\ell^n\|$ being the minimum length, maximum length and the ℓ_2 norm, respectively. Other definitions are:

1) $E_n(\beta,s) = S_n^1(\beta,s)/S_n^0(\beta,s)$,

2) $V_n(\beta,s) = S_n^2(\beta,s)/S_n^0(\beta,s) - (E_n(\beta,s))^2$,

3) for $\beta \in \Theta_n$, $\|\beta\| = \sqrt{\sum_{i=1}^{K_n} \beta_i^2}$,

4) $\sigma_i^2 = \int_0^1 I_i^n(s)\ V(\beta_0,s)\ S^0(\beta_0,s)\ \lambda_0(s)\ ds$, $i=1,\ldots,K_n$, and

5) for $\beta_0^n(i) = \dfrac{\int_0^T I_i^n(s)\ \beta_0(s)\ V(\beta_0,s)\ S^0(\beta_0,s)\ \lambda_0(s)ds}{\sigma_i^2}$, $i=1,\ldots,K_n$,

$$\beta^n(u) = \sum_{i=1}^{K_n} \beta_0^n(i) I_i^n(u), \text{ for } u \text{ in } [0,T].$$

In the following, the superscripts and subscripts, n, are dropped. Only λ_0 and β_0 are constant with increasing n.

2. Consistency

One way to prove consistency of the maximum likelihood estimator is to expand the log-likelihood about the true parameter, say β_0, and then use a fixed point theorem as in Aitchison and Silvey (1958) or Billingsley (1968). However, in the problem considered here, β_0 is, in general, not a member of Θ_n for any finite n; hence in the following proof, the idea is to expand the log-likelihood about a point in Θ_n, say β_0^n, which is close to β_0, instead of expanding about β_0. This introduces a technical difficulty as the score function is no longer a martingale but a martingale plus a bias term. To the first order, this bias term can be eliminated by proper choice of β_0^n as is given in the previous section. Assumptions D and A2 are then useful in showing that the bias is asymptotically negligible.

Theorem 1. Assume

a) $\overline{\lim_n}\ n\|\ell\|^{10} \to 0$ (Bias $\to$ 0),

b) $\underline{\lim_n}\ n\|\ell\|^4 \to \infty$ (Variance converges), and

c) A, C, D1, $\overline{\lim_n} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$,

then for $\hat{\beta}$ maximizing $\mathcal{L}_n(\beta)$ in Θ_n,

$$\sqrt{\|\ell\|^4 n}\ \ \|\hat{\beta}-\beta_0^n\| = O_p(1) .$$

PROOF: Recalling that L is defined in assumption C2, let

$$\delta_n^2 = \|\ell\|^4 \sum_{i=1}^K \int_0^T I_i(s)\ \ell_i^{-2}\ V(\beta_0,s)\ S^0(\beta_0,s)\ \lambda_0(s)ds\ \frac{8}{L^2}.$$

If

$$\sup_{\substack{\beta\in\Theta_n \\ \|\beta-\beta_0^n\|=(\|\ell\|^4 n)^{-1/2}\delta_n}} \sum_{i=1}^{K} n^{-1}\ell_i^{-1} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta)(\beta_i-\beta_0(i)) < 0$$

with probability going to 1 (as $n \to \infty$, $\|\ell\| \to 0$), then by lemma 2 of Aitchison and Silvey (1958), $\exists\ \hat{\beta} \in \Theta_n$ such that $\frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta)|_{\beta=\hat{\beta}} = 0\ \forall i$ and $\|\hat{\beta}-\beta_0^n\| \le \delta_n(\|\ell\|^4 n)^{-1/2}$ on a set of probability going to 1. Since $\frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta))$ is nonpositive for each i, this proves the conclusion. Using a Taylor series about the vector $(\beta_0(1), \ldots \beta_0(K))$, gives,

$$\sum_{i=1}^{K} (n\ \ell_i)^{-1} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta)(\beta_i-\beta_0(i))$$

$$= \sum_{i=1}^{K} (n\ \ell_i)^{-1} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta_0^n)(\beta_i-\beta_0(i))$$

$$+ \sum_{i=1}^{K} (n\ \ell_i)^{-1} \left(\frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta_0^n)\right)(\beta_i-\beta_0(i))^2$$

$$+ \frac{1}{2} \sum_{i=1}^{K} (n\ \ell_i)^{-1} \left(\frac{\partial^3}{\partial\beta_i^3} \mathcal{L}_n(\beta^*)\right)(\beta_i-\beta_0(i))^3$$

where $\|\beta^*-\beta_0^n\| \le \|\beta-\beta_0^n\|$

$$\le \left(\sum_{i=1}^{K} (n\ \ell_i)^{-2} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta_0^n)^2\right)^{1/2} \|\beta-\beta_0^n\|$$

$$+ \sup_{1\le i\le K} \left|(n\ \ell_i)^{-1} \frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta_0^n) + \ell_i^{-1}\sigma_i^2\right| \|\beta-\beta_0^n\|^2 - L\|\beta-\beta_0^n\|^2$$

$$+ \frac{1}{2} \sup_{1\le i\le K} \sup_{\substack{\beta^*\in\Theta_n \\ \|\beta^*-\beta_0^n\|<\|\beta-\beta_0^n\|}} \left|(n\ \ell_i)^{-1} \frac{\partial^3}{\partial\beta_i^3} \mathcal{L}_n(\beta^*)\right| \|\beta-\beta_0^n\|^3$$

Consider $\beta \in \Theta_n$, where $\|\beta-\beta_0^n\|^2 = (\|\ell\|^4 n)^{-1} \delta_n^2$, then, by lemma 1,

$$\sum_{i=1}^{K} (n\, \ell_i)^{-1} \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta)(\beta_i-\beta_0(i))$$

$$\leq \|\beta-\beta_0^n\|^2 \left\{ \frac{L}{2} \sqrt{\frac{\delta_n^2 + O_p(n\|\ell\|^{10}) + O_p(\|\ell\|^4)}{\delta_n^2}} \right.$$

$$\left. + O_p((\sqrt{n}\, \|\ell\|^2)^{-1}) + o_p(1) - L + O_p(1)\|\beta-\beta_0^n\| \right\}$$

(since $(\|\ell\|^4 n)^{-\frac{1}{2}} \delta_n \downarrow 0$).

Since $\underline{\lim}_n\, \delta_n^2 > 0$,

$$\sum_{i=1}^{K} (n\, \ell_i)^{-1} \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta)\, (\beta_i-\beta_0(i))$$

$$\leq \|\beta-\beta_0^n\|^2 \left\{ \frac{L}{2} \sqrt{1 + O_p(n\|\ell\|^{10}) + O_p(\|\ell\|^4)} + O_p((\sqrt{n}\, \|\ell\|^2)^{-1}) \right.$$

$$\left. + o_p(1) - L + O_p(1)\|\beta-\beta_0^n\| \right\}.$$

It is obvious that, for $\epsilon > 0$, $\exists\, n_\epsilon$ such that for $n \geq n_\epsilon$

$$P\left[\sum_{i=1}^{K} (n\, \ell_i)^{-1} \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta)\, (\beta_i-\beta_0(i)) < 0 \quad \forall\, \beta \in \Theta_n \right.$$

$$\left. \text{such that } \|\beta-\beta_0^n\| = (\|\ell\|^2\, n)^{-\frac{1}{2}}\, \delta_n \right] > 1-\epsilon \qquad \square$$

Notes to Theorem 1.

1) Assuming D4 and $\overline{\lim}_n\, n\|\ell\|^6 < \infty$, results in $\int_0^T (\beta_0^n(s)-\beta_0(s))^2 ds =$

$O(\|\ell\|^4)$. Since $n\|\ell\|^4 \; \|\hat{\beta}-\beta_0^n\|^2 = O_p(1)$ implies that $n\|\ell\|^2$ $\int_0^T(\hat{\beta}(s)-\beta_0^n(s))^2 ds = O_p(1)$, one gets $(n\|\ell\|^2) \int_0^T (\hat{\beta}(s)-\beta_0(s))^2 ds = O_p(1)$.

2) It is natural to question whether the rate $\sqrt{\|\ell\|^4 n}$ from Theorem 1. can be improved. In general this will not be possible. To see this, let T=1, and $\ell_i = 1/K$ for each i (so $\|\ell\|^2=1/K$). It turns out that $\sqrt{n\sigma_i^2}\,(\hat{\beta}_i - \beta_0(i))$, i=1,...,K, behave asymptotically like N(0,1) random variables; this indicates that the approximate distribution of $\sum_{i=1}^K (\sqrt{n\sigma_i^2}\,(\hat{\beta}_i - \beta_0(i)))^2$ is chi-squared on K degrees of freedom. So one expects that $\sum_{i=1}^K (\sqrt{n\sigma_i^2}\,(\hat{\beta}_i - \beta_0(i)))^2/K \xrightarrow{P} 1$. This can be proven rigorously using lemmas 1 and 3. Since $\sigma_i^2 = O_p(1/K)$, this gives the rate $\sqrt{n/K^2}$, i.e., $\sqrt{n\|\ell\|^4}$. Other norms might allow for different rates. For example, using the above intuitive reasoning, it is expected that $\sqrt{\frac{n}{\ell n(K)}} \max_{1\le i\le K} \sigma_i \left| \hat{\beta}_i - \beta_0(i) \right| = O_p(1)$, yielding $\sqrt{\frac{n}{K\ell n(K)}} \max_{1\le i\le K} \left| \hat{\beta}_i - \beta_0(i) \right| = O_p(1)$.

3) To understand why the choice of β_0^n given above eliminates the bias to a first order, consider the following:

Maximizing $\mathcal{L}_n(\beta)$ is equivalent to maximizing,

$$\frac{1}{n} \sum_{i=1}^n \int_0^T (\beta(s) - \beta_0(s))X_s(i) - \ell n \left[\frac{S_n^0(\beta,s)}{S_n^0(\beta_0,s)}\right] dN_s(i)$$

for $\beta \in \Theta_n$. This is "asymptotically" like maximizing

$$\int_0^T \left\{(\beta(s) - \beta_0(s))\, S^1(\beta_0,s) - \ell n \left[\frac{S^0(\beta,s)}{S^0(\beta_0,s)}\right] S^0(\beta_0,s)\right\}\lambda_0(s)ds \tag{2.1}$$

$$\doteqdot -\int_0^T (\beta(s)-\beta_0(s))^2 \, V(\beta_0,s) \, S^0(\beta_0,s) \, \lambda_0(s)ds.$$

(under suitable conditions)

But the β maximizing the RHS of (2.1) is given by β_0^n. Therefore it is natural to expect that for the maximum partial likelihood estimator, $\hat{\beta}$, the convergence of $\int_0^t (\hat{\beta}(s)-\beta_0^n(s))^2 ds$ to 0 will be of a faster rate than for choices of $\beta \in \Theta_n$ other than β_0^n.

4) Further consideration of (2.1) lends substance to the use of the L_2 norm in proving consistency. Usually in the method of sieves, the Kullback-Leibler information (in this case, (2.1)) determines the norm in which the maximum likelihood estimator converges to β_0 (see Grenander (1981), Geman & Hwang (1982) and Karr (1987)). In the situation considered here, the L_2 norm approximates, to the first order, the Kullback-Leibler information.

3. Asymptotic Normality

In order to conduct inference about the regression coefficient function, β_0, it is useful to consider some sort of weak convergence result for $\hat{\beta}$. However, in this case and in other situations where the parameter of interest is a function (Karr, 1985; Leskow, 1988; Ramlau-Hansen, 1983) normalized versions of $\hat{\beta}(t)$ and $\hat{\beta}(s)$ have asymptotically independent normal distributions. Intuitively, this means that the limiting distribution of $\hat{\beta}$ is "white noise." This complicates inference using $\hat{\beta}$ taken as a function, as this excludes a functional central theorem. Karr (1985) circumvents this by giving a supremum type statistic which has an asymptotic extreme value distribution. Another possibility is to consider an integrated version of $\hat{\beta}$ as will be done below. McKeague (1987) also considers an integrated version and then proposes the use of a supremum type statistic based on the integrated estimator for inference purposes. One might also consider various weighted integrals of $\hat{\beta}$, i.e. $\int_0^T w_n(x)(\hat{\beta}_n(x)-\beta_0(x))\,dx$ as is done in Aalen (1978) and in Gill (1980). In a later paper, issues involving inference will be addressed.

In the following, the existence of a sequence of estimators $(\hat{\beta}_n \in \Theta_n)$ is assumed such that $\|\hat{\beta}_n-\beta_0^n\| = o_p(1)$, as $n \to \infty$.

Theorem 2. Assume,

a) $\overline{\lim_n}\, n\|\ell\|^8 = 0$ (Bias $\to 0$),

b) $\underline{\lim_n}\, n\|\ell\|^4 = \infty$ (Variance converges), and

c) A, B, C, D2, D4, $\overline{\lim_n} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$, then,

$$\sqrt{n}\int_0^{\bullet} \hat{\beta}_n(s) - \beta_0(s)ds \stackrel{W}{\Rightarrow} G$$

where G is a Gaussian martingale with $G_0 = 0$ a.s.,and

$$\langle G\rangle_t = \int_0^t (V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s))^{-1} ds.$$

PROOF:

Using assumptions D2 and D4, it is easily proved that

$$\sup_{t\in(0,T]} \sqrt{n}\int_0^t \beta_0^n(s) - \beta_0(s)ds = O(n^{1/2}\,\|\ell\|^4).$$

To show that $\sqrt{n}\int_0^{\circ} \hat{\beta}_n(s) - \beta_0^n(s)\,ds \stackrel{W}{\Rightarrow} G$ consider the following Taylor series:

$$0 = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\beta_i}\mathcal{L}_n(\hat{\beta}) = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\beta_i}\mathcal{L}_n(\beta_0^n)$$
$$+ \sqrt{n}\,(\hat{\beta}_i-\beta_0(i))\left[\frac{1}{n}\frac{\partial^2}{\partial\beta_i^2}\mathcal{L}_n(\beta_0^n) + \frac{1}{2n}\frac{\partial^3}{\partial\beta_i^3}\mathcal{L}_n(\beta^*)(\hat{\beta}_i-\beta_0(i))\right]$$

where $\|\beta^*-\hat{\beta}\| \le \|\beta_0^n-\hat{\beta}\|$.

Define $\hat{\sigma}_i^2 = \frac{1}{n}\frac{\partial^2}{\partial\beta_i^2}\mathcal{L}_n(\beta_0^n) + \frac{1}{2n}\frac{\partial^3}{\partial\beta_i^3}\mathcal{L}_n(\beta^*)(\hat{\beta}_i-\beta_0(i))$. Lemma 3 implies that $P[\min_{1\le i\le K} \ell_i^{-1}\,\hat{\sigma}_i^2 > \frac{L}{2}] \to 1$ so it is sufficient to consider $\sqrt{n}\int_0^t (\hat{\beta}_n(s) - \beta_0^n(s))ds$ on this set only. Therefore, solving for $\sqrt{n}\,(\hat{\beta}_i-\beta_0(i))$, multiplying by $I_i(s)$ and integrating from zero to t results in,

$$\sqrt{n}\int_0^t \hat{\beta}_n(s) - \beta_0^n(s)ds$$

$$(3.1)\qquad = \frac{1}{\sqrt{n}} \int_0^t \sum_{i=1}^K I_i(s) \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right] \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n) ds$$

$$+ \frac{1}{\sqrt{n}} \int_0^t \sum_{i=1}^K I_i(s) \frac{1}{\sigma_i^2} \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n) ds$$

To show that the first term on the RHS of (3.1) is $o_p(1)$ in sup norm consider,

$$\left[\frac{1}{\sqrt{n}} \int_0^T \sum_{i=1}^K \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right] \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n) | I_i(s) ds\right]^2$$

$$\le \frac{1}{n} \sum_{i=1}^K \ell_i^2 \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right]^2 \sum_{i=1}^K \left[\frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n)\right]^2$$

$$\le \frac{1}{n} \ell_{(1)}^{-2} \sum_{i=1}^K \ell_i^4 \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right]^2 n^2 \ell_{(K)}^2 \sum_{i=1}^K (n\, \ell_i)^{-2} \left[\frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n)\right]^2$$

$$= \sum_{i=1}^K \ell_i^4 \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right]^2 n \frac{\ell_{(K)}^2}{\ell_{(1)}^2} \left[O_p\left[\frac{1}{\|\ell\|^4 n}\right] + O_p(\|\ell\|^6) + O_p(\tfrac{1}{n})\right]$$

(by lemma 1)

$$= O_p(1) \|\ell\|^4 \left[O_p\left[\frac{1}{\|\ell\|^4 n}\right] + O_p(\|\ell\|^2) + O_p(\|\hat{\beta} - \beta_0^n\|^2)\right]$$

$$\cdot\, n \left[O_p\left[\frac{1}{\|\ell\|^4 n}\right] + O_p(\|\ell\|^6) + O_p(\tfrac{1}{n})\right]$$

$$= o_p(1) \quad \text{(by 1), 2), and lemma 3).}$$

As for the second term on the RHS of (3.1),

$$\frac{1}{\sqrt{n}} \int_0^t \sum_{i=1}^K I_i(s) \frac{1}{\sigma_i^2} \frac{\partial}{\partial \beta_i} \mathcal{L}_n(\beta_0^n) ds$$

$$= \frac{1}{\sqrt{n}} \int_0^t \sum_{i=1}^K I_i(s) \frac{1}{\sigma_i^2} \sum_{j=1}^n \int_0^T I_i(u) \left(X_u(j) - E_n(\beta_0^n, u)\right) dN_u(j) ds.$$

Let,

$$Z_t = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_0^t \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right] (X_s(j) - E_n(\beta_0^n, s))\, dN_s(j)$$

Using McKeague's (1987) lemma 4.1, one gets that if $Z \overset{w}{\Rightarrow} G$, then the second term on the RHS of (3.1) converges weakly to G.

Now,

$$Z_t = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_0^t \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right](X_s(j) - E_n(\beta_0^n, s))\, dM_s(j)$$

$$+ \sqrt{n} \int_0^t \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right](E_n(\beta_0, s) - E_n(\beta_0^n, s))S_n^0(\beta_0, s)\lambda_0(s)\, ds.$$

By lemma 4, the second term of Z_t is $o_p(1)$ in sup norm. As for the first term, the idea is to use the version of Rebolledo's central limit theorem in Anderson & Gill (1982). Call the first term of Z_t, Y_t. Since

$$\langle Y\rangle_t = \int_0^t \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right]^2 [S_n^2(\beta_0, s) + E_n^2(\beta_0^n, s)S_n^0(\beta_0, s) -$$

$$2\, E_n(\beta_0^n, s)S_n^1(\beta_0, s)]\lambda_0(s)\, ds,$$

and $\max_{1 \le i \le K} \sup_{s \in I_i} |\ell_i^{-1}\sigma_i^2 - V(\beta_0, s)S^0(\beta_0, s)\lambda_0(s)| \to 0$ (by the continuity of $V(\beta_0, s)S^0(\beta_0, s)\lambda_0(s)$ in s), one gets, using A1 and lemma 2, that

$$\langle Y_t\rangle \overset{P}{\to} \int_0^t [V(\beta_0, s)S^0(\beta_0, s)\lambda_0(s)]^{-1} ds.$$

A Lindeberg condition must be satisfied also; that is, show

$$\int_0^T \frac{1}{n} \sum_{j=1}^{n} (X_s(j) - E_n(\beta_0^n, s))^2 e^{\beta_0(s)X_s(j)} Y_s(j)\lambda_0(s) \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right]^2$$

$$* \; I\{s : |X_s(j) - E_n(\beta_0^n, s)| > \epsilon\sqrt{n} \left[\sum_{i=1}^{K} I_i(s) \frac{\ell_i}{\sigma_i^2}\right]^{-1}\} ds,$$

is $o_p(1)$ for each $\epsilon > 0$. Recall that $\min_i \ell_i^{-1}\sigma_i^2 \geq L$ so the Lindeberg condition will be satisfied if,

$$\int_0^T \frac{1}{n} \sum_{j=1}^{n} (X_s(j) - E_n(\beta_0^n, s))^2 e^{\beta_0(s)X_s(j)} Y_s(j)\lambda_0(s) \tag{3.2}$$

$$* \; I\{s : |X_s(j) - E_n(\beta_0^n, s)| > \epsilon\sqrt{n}\} ds$$

$$= o_p(1) \quad \forall \epsilon > 0.$$

The LHS of (3.2) is bounded above by,

$$4 \int_0^T \frac{1}{n} \sum_{j=1}^{n} X_s(j)^2 \, e^{\beta_0(s)X_s(j)} Y_s(j) \, \lambda_0(s) \, I\{s : |X_s(j)| > \frac{\epsilon}{2}\sqrt{n}\} \, ds$$

$$+ 4 \int_0^T E_n(\beta_0^n, s)^2 \, S_n^0(\beta_0, s) \, \lambda_0(s) \, I\{s : |E_n(\beta_0^n, s)| > \frac{\epsilon}{2}\sqrt{n}\} ds$$

$$\leq 4 \int_0^T \frac{1}{n} \sum_{j=1}^{n} X_s(j)^2 e^{\beta_0(s)X_s(j)} Y_s(j)\lambda_0(s) \; I\{s : |X_s(j)| \cdot Y_s(j) > \frac{\epsilon}{2}\sqrt{n}\} ds$$

$$+ o_p(1) \quad \text{by A1, C1, and lemma 2.}$$

So the LHS of 3.2 is

$$O_p(1) \cdot \max_{1 \leq j \leq n} \int_0^T I\{s : |X_s(j) \, Y_s(j)| > \frac{\epsilon}{2}\sqrt{n}\} ds + o_p(1)$$

$$= o_p(1) \quad \text{(by B and A1).}$$

4. A Consistent Estimator for the Asymptotic Variance Process

Theorem 4.1. Assume,

a) $n\|\ell\|^4 \to \infty$, and

b) A1, A3, C, D1, D3, $\overline{\lim} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$,

then,

$$\sup_{0\le t\le T} \left| \int_0^t \left[\sum_{i=1}^K I_i(s)(\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\hat{\beta}) \right]^{-1} ds - \int_0^t [V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)]^{-1} ds \right|$$

$$= o_p(1).$$

PROOF:

$$\sup_{0\le t\le T} \left| \int_0^t \left[\sum_{i=1}^K I_i(s)\ (\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\hat{\beta}) \right]^{-1} - [V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)]^{-1}\ ds \right|$$

$$(4.1) \qquad \le \sup_{0\le t\le T} \left| \int_0^t \sum_{i=1}^K I_i(s)\ (\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\hat{\beta}) - V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s) ds \right|$$

$$\cdot \sup_{0\le s\le T} [V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s) \cdot \sum_{i=1}^K I_i(s)(\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\hat{\beta})]^{-1}$$

Consider the first factor on the RHS of (4.1),

$$\sup_{0\le t\le T} \left| \int_0^t \sum_{i=1}^K I_i(s)\ (\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_1^2} \mathcal{L}_n(\hat{\beta}) - [V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)]^{-1} ds \right|$$

$$\le \sup_{0\le t\le T} \left| \int_0^t \sum_{i=1}^K I_i(s)\ (\ell_i n)^{-1} \left[\frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\hat{\beta}) - \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\beta_0^n) \right] ds \right|$$

$$+ \sup_{0\le t\le T} \left| \int_0^t \sum_{i=1}^K I_i(s)[(\ell_i n)^{-1} \frac{\partial^2}{\partial \beta_i^2} \mathcal{L}_n(\beta_0^n) - \ell_i^{-1}\sigma_i^2] ds \right|$$

$$+ \sup_{0\le t\le T} \left| \int_0^t \sum_{i=1}^K I_i(s)\ \ell_i^{-1}\sigma_i^2 - V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s) ds \right|$$

The second term above is $o_p(1)$ by lemma 1 and the third term is $o_p(1)$ by the continuity of $V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)$. As for the first term above,

$$\sup_{0\le t\le T} \left|\int_0^t \sum_{i=1}^K I_i(s)\ (\ell_i n)^{-1}\left[\frac{\partial^2}{\partial\beta_i^2}\mathscr{L}_n(\hat{\beta}) - \frac{\partial^2}{\partial\beta_i^2}\mathscr{L}_n(\beta_0^n)\right]ds\right|$$

$$\le \int_0^T |V_n(\hat{\beta},s) - V_n(\beta_0^n,s)|\ d\bar{N}_s(\cdot)$$

$$= \sum_{i=1}^K |\hat{\beta}_i - \beta_0(i)| \int_0^T I_i(s)\ \left|\frac{\partial}{\partial\beta_i}V_n(\beta^*,s)\right|\ d\bar{N}_s(\cdot)$$

where $\|\beta^*-\beta_0^n\| \le \|\hat{\beta}-\beta_0^n\|$

$= o_p(1)$ by A3, lemma 2 and the fact that $\|\hat{\beta}-\beta_0^n\| \underset{n\to\infty}{\overset{P}{\to}} 0$.

That the second factor in (4.1) is $O_p(1)$, can be proved by lemma 2, a), and A3. □

5 Appendix

Lemma 1. Assume,

a) $\overline{\lim_{n}} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$, and

b) A, C1, D1,

then,

1) $\sum_{i=1}^{K} ((n\ \ell_i)^{-1} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta_0^n))^2$

$$\leq 2(\|\ell\|^4 n)^{-1} [\|\ell\|^4 \int_0^T \sum_{i=1}^{K} I_i(s)\ \ell_i^{-2} V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)\ ds$$

$$+ o_p(1)] + O_p(\|\ell\|^6) + O_p(\tfrac{1}{n}),$$

2) $\max_{1\leq i\leq K} |(n\ \ell_i)^{-1} \frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta_0^n) + \ell_i^{-1}\ \sigma_i^2| = O_p((\sqrt{n}\ \|\ell\|^2)^{-1}) + o_p(1)$, and

3) $\max_{1\leq i\leq K} \sup_{\substack{\beta^*\in\Theta_n \\ \|\beta^*-\beta_0^n\|<.5\gamma}} |(n\ \ell_i)^{-1} \frac{\partial^3}{\partial\beta_i^3} \mathcal{L}_n(\beta^*)| = O_p((\sqrt{n}\ \|\ell\|^2)^{-1}) + O_p(1).$

PROOF:

1) $\sum_{i=1}^{K} ((n\ \ell_i)^{-1} \frac{\partial}{\partial\beta_i} \mathcal{L}_n(\beta_0^n))^2$

$$= \sum_{i=1}^{K} ((n\ \ell_i)^{-1} \sum_{j=1}^{n} \int_0^T I_i(s)[X_s(j)-E_n(\beta_0^n,s)]\ dN_s(j))^2 \tag{5.1}$$

$$\leq 2 \sum_{i=1}^{K} ((n\ \ell_i)^{-1} \sum_{j=1}^{n} \int_0^T I_i(s)[X_s(j)-E_n(\beta_0^n,s)]\ dM_s(j))^2$$

$$+ 2 \sum_{i=1}^{K} (\ell_i^{-1} \int_0^T I_i(s)[E_n(\beta_0^n,s)$$

$$- E_n(\beta_0,s)]S_n^0(\beta_0,s)\lambda_0(s)\ ds)^2$$

Consider

$$Z_t = \sum_{i=1}^{K} ((n\,\ell_i)^{-1} \sum_{j=1}^{n} \int_0^t I_i(s)\,[X_s(j)-E_n(\beta_0^n,s)]\,dM_s(j))^2$$

The compensator of Z, is

$$C_t = \sum_{i=1}^{K} (n\,\ell_i)^{-2} \sum_{j=1}^{n} \int_0^t I_i(s)[X_s(j) - E_n(\beta_0^n,s)]^2 Y_s(j) e^{\beta_0(s)X_s(j)} \lambda_0(s)\,ds$$

$$= \int_0^t \sum_{i=1}^{K} I_i(s)\,\ell_i^{-2}\,n^{-1}[S_n^2(\beta_0,s) - 2S_n^1(\beta_0,s)\,E_n(\beta_0^n,s)$$

$$+ E_n^2(\beta_0^n,s)S_n^0(\beta_0,s)]\lambda_0(s)ds$$

To show that Z_t has the same limit in probability as its' compensator C_t, it is sufficient (by Lenglart's inequality, Lenglart (1977)) to show that the quadratic variation of $\|\ell\|^4 n(Z_t - C_t)$ goes to zero in probability. Denoting the endpoints of interval I_i by a_i and a_{i+1}, and defining $M^*(a_i,s) = 2 \sum_{j=1}^{n} n^{-1} \int_{a_i}^{s-} \ell_i^{-1}[X_u(j)-E_n(\beta_0^n,u)]\,dM_u(j)$, the optional variation of $\|\ell\|^4 n(Z-C)$ is (Kopp, 1984, pg. 148),

$$[\|\ell\|^4 n(Z-C)]_t = \sum_{s\leq t} \|\ell\|^8\, n^2 (\Delta(Z-C)_s)^2$$

$$= \|\ell\|^8 n^2 \sum_{j=1}^{n} \int_0^t \sum_{i=1}^{K} I_i(s)\,\{\,M^*(a_i,s)^2$$

$$\cdot [X_s(j) - E_n(\beta_0^n,s)]^2\, n^{-2}\ell_i^{-2} + [X_s(j)-E_n(\beta_0^n,s)]^4 n^{-4}\ell_i^{-4}$$

$$+ 2\,M^*(a_i,s) \cdot [X_s(j) - E_n(\beta_0^n,s)]^3 n^{-3}\ell_i^{-3}\}\,dN_s(j).$$

Then the compensator of $[\|\ell\|^4 n(Z-C)]$ is given by,

$$\langle \|\ell\|^4 n\,(Z-C)\rangle_t = \|\ell\|^8 n^2 \int_0^t \sum_{i=1}^{K} I_i(s)$$

$$\left\{ M^*(a_i,s)^2\, n^{-1}\ell_i^{-2} \left[\frac{1}{n} \sum_{j=1}^{n} (X_s(j)-E_n(\beta_0^n,s))^2\, e^{\beta_0(s)X_s(j)}\, Y_s(j)\right]\right.$$

$$+ n^{-3} \ell_i^{-4} \left[\frac{1}{n} \sum_{j=1}^{n} (X_s(j) - E_n(\beta_0^n,s))^4 e^{\beta_0(s)X_s(j)} Y_s(j)\right]$$

$$+ 2 M^*(a_i,s) n^{-2} \ell_i^{-3} \left[\frac{1}{n} \sum_{j=1}^{n} (X_s(j)\right.$$

$$\left.\left. - E_n(\beta_0^n,s))^3 e^{\beta_0(s)X_s(j)} Y_s(j)\right]\right\} \lambda_0(s)ds$$

$$= \|\ell\|^4 n \int_0^t \sum_{i=1}^{K} I_i(s) M^*(a_i,s)^2 ds\ O_p(1)$$

$$+ n^{-1} O_p(1) + \|\ell\|^2 \int_0^t \sum_{i=1}^{K} I_i(s) |M^*(a_i,s)| ds\ O_p(1)$$

by A3, C1.

Now,
$$\max_{1 \le i \le K} \sup_{s \in I_i} |M^*(a_i,s)| \le \max_{1 \le i \le K} \sup_{s \in I_i} |M^*(0,s]| + \max_{1 \le i \le K} |M^*(0,a_i]|$$

$$\le 4 \sup_{s \in [0,T]} \left| \sum_{i=1}^{K} \sum_{j=1}^{n} n^{-1} \ell_i^{-1} \int_0^s I_i(u)[X_u(j)-E_n(\beta_0^n,u)]\, dM_u(j)\right|$$

and using Lenglart's inequality (1977) for $B > 0$,

$$P\left[\sup_{0 \le t \le T} \left| \sum_{i=1}^{K} \sum_{j=1}^{n} n^{-1} \ell_i^{-1} \int_0^s I_i(u)[X_u(j) - E_n(\beta_0^n,u)]\, dM_u(j)\right|^2 \right.$$

$$\left. > B(\|\ell\|^4 n)^{-1}\right] \le \frac{(\|\ell\|^4 n)^{-1} B\frac{\epsilon}{2}}{(\|\ell\|^4 n)^{-1} B}$$

$$+ P\left[\|\ell\|^4 n \sum_{i=1}^{K} n^{-1} \ell_i^{-2} \int_0^T I_i(s)[S_n^2(\beta_0,s) - 2 S_n^1(\beta_0,s)E_n(B_n,s) \right. \tag{5.2}$$

$$\left. + E_n^2(\beta_0^n,s) S_n^0(\beta_0,s)]\lambda_0(s)\, ds \ge B\frac{\epsilon}{2} \right]$$

$\le \epsilon$ for B large and n large (use A3, C1).

Therefore $\|\ell\|^2 \int_0^t \sum_{i=1}^K I_i(s)|M^*(a_i,s)|ds = O_p(n^{-1/2})$.

Consider the process, $\int_0^t \sum_{i=1}^K I_i(s)M^*(a_i,s)^2 ds$ for t belonging to $\{a_1 = 0, a_1, \ldots, a_{K+1} = T\}$ and the family $\{\mathcal{F}_{a_i}\}_{i=0,K+1}$. On $\{\mathcal{F}_{a_i}\}_{i=0,K+1}$,

$$\int_0^\cdot \sum_{i=1}^K I_i(s)\, M^*(a_i,s)^2 ds$$

$$- \int_0^\cdot \sum_{i=1}^K I_i(s) \sum_{j=1}^n n^{-2}\, \ell_i^{-2} \int_{a_i}^s [X_u(j)-$$

$$E_n(\beta_0^n,u)]^2\, e^{\beta_0(u)X_u(j)}\, Y_u(j)\lambda_0(u)\, du\, ds$$

is a local martingale. Therefore by Lenglart's inequality (1977) for $B > 0$, $\epsilon > 0$,

$$P[\|\ell\|^2\, n \int_0^T \sum_{i=1}^K I_i(s)\, M^*(a_i,s)^2\, ds \geq B]$$

$$(5.3) \qquad \leq \frac{B(\|\ell\|^2 n)^{-1}\, \frac{\epsilon}{2}}{B(\|\ell\|^2 n)^{-1}} + P\Big[\|\ell\|^2 n \int_0^T \sum_{i=1}^K I_i(s) n^{-1}\, \ell_i^{-2} \int_{a_i}^s [S_n^2(\beta_0,u)$$

$$+ E_n^2(\beta_0^n,u)S_n^0(\beta_0,u)$$

$$- 2\, E_n(\beta_0^n,u)S_n^1(\beta_0,u)]\lambda_0(u)\, du\, ds \geq B\, \frac{\epsilon}{2}\Big]$$

$\leq \epsilon$ for B and n large (use A1, C1, and lemma 2).

Therefore $\langle \|\ell\|^4\, n(Z-C)\rangle_T = \|\ell\|^2\, O_p(1) + n^{-1}\, O_p(1) + n^{-1/2}\, O_p(1)$ (by (5.2), (5.3)). This, as mentioned earlier, implies that

$$\sup_{0 \leq t \leq T} \|\ell\|^4 n\, |Z_t - C_t| = o_p(1).$$

Since,

$$\sup_{0 \leq t \leq T} \|\ell\|^4\, n|C_t - \int_0^t \sum_{i=1}^K I_i(s)\, \ell_i^{-2}\, n^{-1}\, V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)\, ds|$$

$$= \|\ell\|^4 \sum_{i=1}^{K} \ell_i^{-1} \, o_p(1) \quad \text{by A1 and C1,}$$

$$\sup_{0\leq t\leq T} \|\ell\|^4 n \left| Z_t - \int_0^t \sum_{i=1}^{K} I_i(s)\, \ell_i^{-2} n^{-1} V(\beta_0,s) S^0(\beta_0,s)\lambda_0(s)\, ds \right| = o_p(1).$$

This concludes the proof for the first term on the RHS of (5.1).

Consider the second term on the RHS of (5.1),

$$\sum_{i=1}^{K} \left(\ell_i^{-1} \int_0^T I_i(s)[E_n(\beta_0^n,s) - E_n(\beta_0,s)]S_n^0(\beta_0,s)\lambda_0(s)ds\right)^2.$$

Let $\|x\|^2 = \sum_{i=1}^{K} \left(\ell_i^{-1} \int_0^T I_i(s)(\beta_0^n(s) - \beta_0(s))(V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)\, S^0(\beta_0,s))\, \lambda_0(s)\, ds\right)^2$

and,

$$\|y\|^2 = \sum_{i=1}^{K} \left(\ell_i^{-1} \int_0^T I_i(s)(\beta_0^n(s) - \beta_0(s))^2 S_n^0(\beta_0,s)\lambda_0(s)\, ds\right)^2 \cdot \sup_{0\leq s\leq T} \sup_{\substack{\beta(s)\in\mathbb{R} \\ |\beta(s)-\beta_0(s)|<\gamma}} \frac{1}{4}\left(\frac{\partial}{\partial\beta(s)} V_n(\beta,s)\right)^2.$$

Using a Taylor series for fixed s yields:

$$E_n(\beta_0^n,s) = E_n(\beta_0,s) + (\beta_0^n(s) - \beta_0(s))\, V_n(\beta_0,s) + \frac{1}{2}(\beta_0^n(s) - \beta_0(s))^2 \frac{\partial}{\partial\beta(s)} V_n(\beta,s) \quad \text{where}$$

$$|\beta(s)-\beta_0(s)| \leq |\beta_0^n(s) - \beta_0(s)|$$

Then, since $\sup_{0\leq s\leq T} |\beta_0^n(s) - \beta_0(s)| = o(1)$,

$$\sum_{i=1}^{K} \left(\ell_i^{-1} \int_0^T I_i(s)[E_n(\beta_0^n,s) - E_n(\beta_0,s)]S_n^0(\beta_0,s)\lambda_0(s)ds\right)^2$$

$$\leq 2\|x\|^2 + 2\|y\|^2.$$

It turns out that, $\|x\|^2 = O_p(\frac{1}{n})$ and $\|y\|^2 = O(\|\ell\|^6)$ so that the second term on the RHS of (5.1) is equal to $O_p(\frac{1}{n}) + O_p(\|\ell\|^6)$. By A, C1, and D1, one gets,

$$\|x\|^2 = \sum_{i=1}^{K} \left(\int_0^T I_i(s)\,|V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)|\,ds\right)^2 O(1)$$

$$= \sum_{i=1}^{K} \left(\ell_i^{1/2} \sqrt{\frac{1}{n}}\,\right)^2 O_p(1)$$

$$= O_p\left(\frac{1}{n}\right), \text{ and}$$

$$\|y\|^2 = \sum_{i=1}^{K} (\ell_i^2)^2\, O_p(1) = \|\ell\|^6\, O_p(1).$$

2) $$\left|(\ell_i\, n)^{-1} \frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta_0^n) + \ell_i^{-1}\sigma_i^2\right|$$

$$= \left|\ell_i^{-1}\int_0^T I_i(s)\, V_n(\beta_0^n,s)\, d\bar{N}_s(\cdot) - \ell_i^{-1}\int_0^1 I_i(s)V(\beta_0,s)S^0(\beta_0,s)\lambda_0(s)\, ds\right|$$

$$\leq \left|\ell_i^{-1} \int_0^T I_i(s)(V_n(\beta_0^n,s) - V(\beta_0,s))\, d\bar{N}_s(\cdot)\right|$$

$$+ \ell_i^{-1} \left|\int_0^T I_i(s)V(\beta_0,s)\, d\bar{M}_s(\cdot)\right|$$

$$+ \ell_i^{-1} \left|\int_0^T I_i(s)V(\beta_0,s)(S_n^0(\beta_0,s) - S^0(\beta_0,s))\lambda_0(s)ds\right|$$

$$\leq \sup_{0\leq s\leq T} |V_n(\beta_0^n,s) - V(\beta_0,s)|\, \max_i\, \ell_i^{-1} \int_0^T I_i(s)\, d\bar{N}_s(\cdot)$$

$$+ \max_{1\leq i\leq K}\, \ell_i^{-1} \left|\int_0^T I_i(s)V(\beta_0,s)\, d\bar{M}_s(\cdot)\right|$$

$$+ \sup_{0\leq s\leq T} |S_n^0(\beta_0,s) - S^0(\beta_0,s)| \cdot \max_{1\leq i\leq K} \ell_i^{-1} \int_0^T I_i(s)V(\beta_0,s)\lambda_0(s)ds$$

So by lemma 2 and C1,

$$\max_{1\leq i\leq K} \left|(\ell_i\, n)^{-1} \frac{\partial^2}{\partial\beta_i^2} \mathcal{L}_n(\beta_b) - \ell_1^{-1}\, \sigma_i^2\right| = O_p((\sqrt{n}\, \|\ell\|^2)^{-1}) + o_p(1)$$

3) $$\max_{1\le i\le K} \sup_{\substack{\beta^*\in\Theta_n \\ \|\beta^*-\beta_0^n\|<.5\gamma}} \left|(n\,\ell_i)^{-1} \frac{\partial^3}{\partial\beta_i^3} \mathcal{L}_n(\beta^*)\right| =$$

$$\max_{1\le i\le K} \sup_{\substack{\beta^*\in\Theta_n \\ \|\beta^*-\beta_0^n\|<.5\gamma}} \left|\ell_i^{-1} \int_0^T I_i(s) \left[\frac{S_n^3(\beta^*,s)}{S_n^0(\beta^*,s)} - \frac{3\,S_n^2(\beta^*,s)}{S_n^0(\beta^*,s)^2} E_n(\beta^*,s) + 2\,E_n(\beta^*,s)^3\right] d\bar{N}(\cdot)\right|$$

By assumption A3, C1, and lemma 2, the above is,

$$\le O_p(1) \max_{1\le i\le K} \ell_i^{-1} \int_0^T I_i(s)\, d\bar{N}_s(\cdot) = O_p(1) + O_p((\sqrt{n}\|\ell\|^2)^{-1}) \qquad \square$$

Lemma 2. Assume A1, A3, C1, and D1,

then,

1) $$\max_{1\le i\le K} \left|\int_0^T I_i(s)\, d\bar{M}_s(\cdot)\right| = O_p\left(\frac{1}{\sqrt{n}}\right), \text{ and}$$

2) $$\sup_{0\le s\le T} \left|S_n^i(\beta_0^n,s) - S_n^i(\beta_0,s)\right| = O_p(\ell_{(K)}) \quad i=0,1,2.$$

PROOF:

1) Let $B > 0$ and consider,

$$\max_{1\le i\le K} \left|\sqrt{n} \int_0^T I_i(s)\, d\bar{M}_s(\cdot)\right| \le 2 \sup_{t\in[0,T]} \left|\sqrt{n}\, \bar{M}_t(\cdot)\right|.$$

Using the version of Rebolledo's central limit theorem present in Anderson and Gill (1982) it is easily proved that for $Z_t^n = \sqrt{n}\,\bar{M}_t(\cdot)$, Z^n converges weakly to a Gaussian martingale with variance function $\int_0^t S^0(\beta_0,s)\lambda_0(s)ds$. An application of the continuous mapping theorem (Theorem 5.1 in Billingsley, 1968) suffices to prove 1.

2) Fix s, then using a Taylor series about $\beta_0^n(s)$ results in,

$$S_n^i(\beta_0, s) - S_n^i(\beta_0^n, s) = S_n^{i+1}(b, s)(\beta_0(s) - \beta_0^n(s)) \quad \text{where}$$

$$|b-\beta_0^n(s)| \le |\beta_0(s) - \beta_0^n(s)| \quad i=0,1,2.$$

Therefore,

$$\sup_{0\le s\le T} |S_n^i(\beta_0, s) - S_n^i(\beta_0^n, s)|$$

$$= \sup_{0\le s\le T} |\beta_0(s) - \beta_0^n(s)|\, O_p(1) \quad \text{(by A3)}$$

$$= O_p(\ell_{(k)}) \quad \text{(by D1)}. \qquad \square$$

Lemma 3. Assume A, B1, C1, and $\overline{\lim}\, \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$,

then,

$$\sum_{i=1}^K (\sigma_i^2 - \hat{\sigma}_i^2)^2 = \|\ell\|^4 [O_p(\frac{1}{\|\ell\|^4 n}) + O_p(\|\ell\|^2) + O(\|\hat{\beta}-\beta_0^n\|^2)].$$

PROOF:

$$\sum_{i=1}^K (\sigma_i^2 - \hat{\sigma}_i^2)^2 \le 2 \sum_{i=1}^K \Big[\int_0^T I_i(s)\, V_n(\beta_0^n, s)\, d\bar{N}_s(\cdot)$$

$$- \int_0^T I_i(s) V(\beta_0, s) S^0(\beta_0, s)\lambda_0(s)\, ds\Big]^2$$

$$+ 2 \sum_{i=1}^K (\frac{1}{2n} \frac{\partial^3}{\partial \beta_i^3} \mathcal{L}_n(\beta^*)(\hat{\beta}_i - \beta_0(i))^2$$

$$\text{where } \|\beta^* - \beta_0^n\| \le \|\beta_0^n - \hat{\beta}\|.$$

On $\|\beta_0^n - \hat{\beta}\| < .5\gamma$,

$$\sum_{i=1}^K (\sigma_i^2 - \hat{\sigma}_i^2)^2 \le 4 \sum_{i=1}^K (\int_0^T I_i(s)\, V_n(\beta_0^n, s)\, d\bar{M}_s(\cdot))^2 +$$

$$+ 4 \sum_{i=1}^K (\int_0^T I_i(s)[V_n(\beta_0^n, s) S_n^0(\beta_0, s) - V(\beta_0, s) S^0(\beta_0, s)]\lambda_0(s) ds)^2$$

$$+ .5\ell_{(K)}^2 \ \|\hat{\beta}-\beta_0^n\|^2 \max_{1\le i\le K} \sup_{\substack{\beta^*\in\Theta_n \\ \|\beta^*-\beta_n\|<.5\gamma}} \left|(\ell_i n)^{-1} \frac{\partial^3}{\partial\beta_i^3} \mathcal{L}_n(\beta^*)\right|^2$$

Using lemma 1 results in,

$$\sum_{i=1}^K (\sigma_i^2-\hat{\sigma}_i^2)^2 \le 4 \sum_{i=1}^K \left(\ell_i^{-1} \int_0^T I_i(s)V_n(\beta_0^n,s)\, d\overline{M}_s(\cdot)\right)^2 O(\|\ell\|^4)$$

$$+ 4 \sum_{i=1}^K \Big(\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0^n,s)S_n^0(\beta_0,s)$$

$$- V(\beta_0,s)S^0(\beta_0,s)]\, \lambda_0(s)ds\Big)^2 O(\|\ell\|^4)$$

$$+ \|\hat{\beta}-\beta_0^n\|^2 O_p(\|\ell\|^4).$$

Using Lenglart's inequality (Lenglart, 1977) it is easy to show (using lemma 2) that

$$\sum_{i=1}^K (\sigma_i^2-\hat{\sigma}_i^2)^2 = O_p\left(\frac{1}{\|\ell\|^4 n}\right) O(\|\ell\|^4) + \|\hat{\beta}-\beta_0^n\|^2 O(\|\ell\|^4)$$

$$+ O(\|\ell\|^4) \sum_{i=1}^K \Big(\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0^n,s)S_n^0(\beta_0,s)$$

$$- V(\beta_0,s)S^0(\beta_0,s)]\, \lambda_0(s)ds\Big)^2$$

All that is left is to prove that

$$\sum_{i=1}^K \Big(\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0^n,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)]\, \lambda_0(s)ds\Big)^2$$

(5.4)
$$= \sum_{i=1}^K \Big(\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0^n,s) - V_n(\beta_0,s)]S_n^0(\beta_0,s)\lambda_0(s)ds\Big)^2$$

$$+ \sum_{i=1}^K \Big(\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)]\lambda_0(s)ds\Big)^2$$

$$= O_p(\|\ell\|^2) + O_p(\frac{1}{n\|\ell\|^4}).$$

Using lemma 2, it is easy to show that the first term on the RHS of (5.4) is $O_p(\|\ell\|^2)$. The second term,

$$\sum_{i=1}^K (\ell_i^{-1} \int_0^T I_i(s)[V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)]\lambda_0(s)ds)^2,$$

can be divided up into terms such as

$$\sum_{i=1}^K (\ell_i^{-1} \int_0^T I_i(s)|S_n^j(\beta_0,s) - S^j(\beta_0,s)|ds)^2 \; O_p(1)$$

$j = 0,1,2$, by lemma 2, and the fact that $\inf_{0\leq s\leq T} S^0(\beta_0,s) > 0$.

The proof will be concluded if for $j = 0,1,2$,

$$\sum_{i=1}^K (\ell_i^{-1} \int_0^T I_i(s)|S_n^j(\beta_0,s)- S^j(\beta_0,s)|ds)^2 = O_p(\frac{1}{n\|\ell\|^4}).$$

The above LHS is less than or equal to,

$$\sum_{i=1}^K (\ell_i^{-2} \ell_i \int_0^T(S_n^j(\beta_0,s) - S^j(\beta_0,s))^2 ds)$$

$$= \ell_{(1)}^{-2} \int_0^T (S_n^j(\beta_0,s) - S^j(\beta_0,s))^2 ds \; O(1).$$

Using A2, and $\overline{\lim_n} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$ yields the desired result. □

Lemma 4. Assume A, C, D and $\overline{\lim} \frac{\ell_{(K)}}{\ell_{(1)}} < \infty$,

then,

$$\sup_{0\leq t\leq T} | \sqrt{n} \int_0^t \left[\sum_{i=1}^K I_i(s) \frac{\ell_i}{\sigma_i^2}\right] (E_n(\beta_0,s) - E_n(\beta_0^n,s))S_n^0(\beta_0,s)\lambda_0(s) \, ds|$$

$$= O_p(\sqrt{n} \; \|\ell\|^4) + O_p(\|\ell\|^2) \; .$$

PROOF:

Using a Taylor series on $\beta_0^n(s)$ about $\beta_0(s)$ at each s results in

$$E_n(\beta_0^n,s) - E_n(\beta_0,s) = (\beta_0^n(s) - \beta_0(s))\, V_n(\beta_0,s)$$

$$+ .5(\beta_0^n(s) - \beta_0(s))^2 \frac{\partial}{\partial\beta(s)} V_n(\beta,s)$$

where $|\beta(s) - \beta_0(s)| \le |\beta_0(s) - \beta_0^n(s)|$ and subsequently for $|\beta_0(s) - \beta_0^n(s)|<\gamma$,

$$\sup_{0\le s\le T}\ \sup_{\substack{\beta(s)\in\mathbb{R} \\ |\beta(s)-\beta_0(s)|<\gamma}} \left|\frac{\partial}{\partial\beta(s)} V_n(\beta,s)\right| = O_p(1) \quad \text{by A3.}$$

Then $\sup\limits_{0\le t\le T} \left|\sqrt{n} \int_0^t \left[\sum_{i=1}^K I_i(s) \frac{\ell_i}{\sigma_i^2}\right] (E_n(\beta_0,s) - E_n(\beta_0^n,s))S_n^0(\beta_0,s)\lambda_0(s)ds\right|$

$$\le \sup_{0\le t\le T} \left|\sqrt{n} \int_0^t \left[\sum_{i=1}^K I_i(s) \frac{\ell_i}{\sigma_i^2}\right] (\beta_0(s) - \beta_0^n(s))\, V_n(\beta_0,s)S_n^0(\beta_0,s)\lambda_0(s)ds\right|$$

$$+ \sqrt{n} \int_0^T (\beta_0(s) - \beta_0^n(s))^2 ds\ O_p(1) \quad \text{(by A, C1).}$$

Using the definition of β_0^n it is easy to see that the second term above is $O_p(\sqrt{n}\ \|\ell\|^4)$. As for the first term,

$$\sup_{0\le t\le T} \left|\sqrt{n} \int_0^t \left[\sum_{i=1}^K I_i(s) \frac{\ell_i}{\sigma_i^2}\right] (\beta_0(s) - \beta_0^n(s))\, V_n(\beta_0,s)S_n^0(\beta,s)\lambda_0(s)\ ds\right|$$

$$\le \sup_{0\le t\le T} \left|\sqrt{n} \int_0^t (\beta_0(s) - \beta_0^n(s))ds\right|$$

$$+ \sup_{0\le t\le T} \left|\sqrt{n} \int_0^t |\beta_0(s) - \beta_0^n(s)| \left[\sum_{i=1}^K I_i(s) \frac{\ell_i}{\sigma_i^2} V_n(\beta_0,s)S_n^0(\beta_0,s)\lambda_0(s) - 1\right]ds.\right.$$

The first term above, $\sup\limits_{0\le t\le T} \left|\sqrt{n} \int_0^t (\beta_0(s) - \beta_0^n(s))ds\right|$, has already been shown to be $O(\sqrt{n}\|\ell\|^4)$. The second term is equal to,

$$O(1)\cdot\sqrt{n}\int_0^T |\beta_0(s) - \beta_0^n(s)|\ |V_n(\beta_0,s)S_n^0(\beta_0,s)\lambda_0(s) - \left[\sum_{i=1}^K I_i(s)\frac{\sigma_i^2}{\ell_i}\right]|ds$$

$$\leq O(1)\ \sqrt{n}\int_0^T|\beta_0(s) - \beta_0^n(s)||V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)|\lambda_0(s)ds$$

$$+ O(1)\ \sqrt{n}\int_0^T|\beta_0(s) - \beta_0^n(s)||V_n(\beta_0,s)S^0(\beta_0,s)\lambda_0(s) - \left[\sum_{i=1}^K I_i(s)\frac{\sigma_i^2}{\ell_i}\right]|ds$$

$$= O_p(\sqrt{n}\|\ell\|^2)\int_0^T |V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)\ S^0(\beta_0,s)|ds$$

$$+ O_p(\sqrt{n}\|\ell\|^4)$$ by D4.

Using A2 and C1, results in,

$$\int_0^T |\ V_n(\beta_0,s)S_n^0(\beta_0,s) - V(\beta_0,s)S^0(\beta_0,s)|ds = O_p(\frac{1}{\sqrt{n}}).$$ □