The Annals of Statistics
1986, Vol. 14, No. 2, 579–589

ESTIMATION FOR A SEMIMARTINGALE REGRESSION MODEL USING THE METHOD OF SIEVES[1]

BY IAN W. MCKEAGUE

Florida State University

Estimation by the method of sieves for a semimartingale regression model introduced by Aalen (1980) is studied. It is of interest to estimate functions which describe the influence of the covariates over time. An estimator for these functions is introduced and conditions which ensure consistency of the estimator in L^2-norm are given. Applications to diffusion processes and point processes with censored data are also discussed.

1. Introduction. The method of sieves (Grenander, 1981) has proved to be a powerful technique in nonparametric estimation. It has recently been applied to stochastic processes for the estimation of such time dependent functions as the mean of a translate of the Wiener process (Grenander, 1981; Geman and Hwang, 1982), the drift coefficient of a linear diffusion process (Nguyen and Pham 1982), the hazard function in the multiplicative intensity model for point processes (Karr, 1983), and the mean of a Gaussian process (Antoniadis, 1985).

In the present paper we study estimation by the method of sieves for the following semimartingale regression model which was introduced by Aalen (1980). It contains diffusion processes and the multiplicative intensity model for point processes as important examples. Suppose that n subjects and p covariates for each subject are observed over the time interval $[0,1]$. Let $X_i(t)$ denote the state of the ith subject at time t, and suppose that $X = (X_1, \dots, X_n)'$ satisfies

$$X(t) = X(0) + \int_0^t Y(s)\alpha(s)\,ds + M(t), \qquad t \in [0,1], \tag{1.1}$$

where $\alpha = (\alpha_1, \dots, \alpha_p)'$ is a vector of unknown nonrandom functions, $Y = (Y_{ij})$ is the $n \times p$ matrix of covariate processes, with Y_{ij} being the jth covariate for the ith subject, and $M = (M_1, \dots, M_n)'$ where each M_i is a square integrable martingale.

It is of interest to estimate the functions $\alpha_1, \dots, \alpha_p$ and so provide detailed information on changes in the influence of the covariates over time. In Section 2 we introduce an estimator $\hat{\alpha}^{(n)}$ for α and state conditions under which $\int_0^1 [\hat{\alpha}_j^{(n)}(t) - \alpha_j(t)]^2\,dt \to 0$ in probability as $n \to \infty$, for $j = 1, \dots, p$. Our approach is based on the sieve method developed for linear diffusion processes by Nguyen and Pham (1982) and is similar to the well known orthogonal series technique for nonparametric density estimation, first used by Cencov (1962). We

---

Received October 1984; revised July 1985.

[1] Research supported by the U.S. Army Research Office under Grant DAAG 29-82-K-0168.

AMS 1980 *subject classifications.* Primary 62M09, 62G05; secondary 60G44.

Key words and phrases. Inference for stochastic processes, method of sieves, regression analysis, semimartingales, point process, diffusion process, censoring.

use an increasing sequence of finite dimensional subspaces of $L^2[0,1]$ and define the estimator $\hat{\alpha}_j^{(n)}$ to be an element of the nth subspace. The dimension d_n of the nth subspace is allowed to tend to infinity at the rate $d_n = o(n)$ as $n \to \infty$. This improves on the rate $d_n = o(n^{1/2})$ given by Nguyen and Pham (1982). Various moment conditions ((A1)–(A5), Section 2) are imposed on the p covariate processes. These conditions are easily satisfied, unless $p \geq 3$, in which case condition (A4) becomes more severe as p increases. Several examples of our model (1.1) are discussed in Section 3; proofs are contained in Section 4.

2. Estimation of α. We begin by stating some technical assumptions needed in the semimartingale regression model (1.1). $(\Omega, \mathscr{F}, P)$ will denote a complete probability space and for each $i = 1, \ldots, n$, $(\mathscr{F}_{it}, t \in [0,1])$ is a nondecreasing right-continuous family of sub-σ-fields of $\mathscr{F}$ where $\mathscr{F}_{i0}$ contains all P-null sets in $\mathscr{F}$. All processes are indexed by $t \in [0,1]$. Each process $(M_i(t), \mathscr{F}_{it})$, $i = 1, \ldots, p$, is assumed to be a square integrable martingale such that almost all paths of M_i are right-continuous on $[0,1)$ with left limits on $(0,1]$. The predictable variation of M_i is the unique increasing, $(\mathscr{F}_{it})$ predictable process $\langle M_i \rangle_t$ such that $\langle M_i \rangle_0 = M_i^2(0)$ and $M_i^2 - \langle M_i \rangle$ is a martingale; refer to Meyer (1976).

The covariate process Y_{ij} is assumed to be $(\mathscr{F}_{it})$ predictable, that is measurable with respect to the σ-field on $[0,1] \times \Omega$ generated by all left-continuous, $(\mathscr{F}_{it})$ adapted processes. The σ-field $\mathscr{F}_{it}$ represents the state and covariate history of the ith subject up to time t. It is assumed that $(X_n, n \geq 1)$, $(M_n, n \geq 1)$, and $(Y_{nj}, n \geq 1)$ for $j = 1, \ldots, p$ are strictly stationary sequences. In particular, this will be the case if the subjects are iid.

The method of sieves consists in taking an estimator from an increasing sequence of sets of functions indexed by the sample size. For each $j = 1, \ldots, p$, let $(\phi_{jr}, r \geq 1)$ be a complete orthonormal sequence in $L^2[0,1]$. Define the estimator $\hat{\alpha}_j^{(n)}$ of α_j to be the element of $\operatorname{span}\{\phi_{jr}, r = 1, \ldots, d_n\}$ given by

$$\hat{\alpha}_j^{(n)}(t) = \sum_{r=1}^{d_n} \hat{\alpha}_{jr}^{(n)} \phi_{jr}(t), \tag{2.1}$$

where (d_n) is an increasing sequence of positive integers, the $p \times d_n$ matrix $\hat{\alpha}^{(n)} \equiv (\hat{\alpha}_{jr}^{(n)})$ satisfies

$$\operatorname{vec}(\hat{\alpha}^{(n)}) = A^{(n)-1} \operatorname{vec}(B^{(n)}), \tag{2.2}$$

where the vec operator takes a matrix and places the elements in lexicographical order to form a large column vector, $B^{(n)}$ is the $p \times d_n$ matrix given by

$$B_{jr}^{(n)} = \sum_{i=1}^{n} \int_0^1 \phi_{jr}(t) Y_{ij}(t)\, dX_i(t), \tag{2.3}$$

and $A^{(n)}$ is the $pd_n \times pd_n$ matrix partitioned into p^2 submatrices $A_{jk}^{(n)}$ of order $d_n \times d_n$ with

$$A_{jkrl}^{(n)} = \sum_{i=1}^{n} \int_0^1 \phi_{jr}(t) \phi_{kl}(t) Y_{ij}(t) Y_{ik}(t)\, dt. \tag{2.4}$$

In (2.2), $A^{(n)-1}$ denotes a generalized inverse of $A^{(n)}$. The choice of generalized inverse here does not affect any of our results since, by the proof of Lemma 4.2,

$$P(A^{(n)} \text{ is invertible}) \to 1, \quad \text{as } n \to \infty.$$

In the case of diffusion processes, $\hat{\alpha}_j^{(n)}$ can be derived as a restricted maximum likelihood estimator; see Nguyen and Pham (1982). However, for dependent observations and arbitrary square integrable martingales no such interpretation is available. A rationale for using $\hat{\alpha}_j^{(n)}$ comes from the following result which establishes the L^2 consistency of the estimator under some assumptions (A1)–(A6) stated after the theorem.

THEOREM 2.1. *Under* (A1)–(A6), *for any* $d_n \uparrow \infty$ *such that* $d_n = o(n)$,

$$\int_0^1 \left[\hat{\alpha}_j^{(n)}(t) - \alpha_j(t)\right]^2 dt \to_p 0, \quad as\ n \to \infty.$$

ASSUMPTIONS.

(A1) $$\int_0^1 \alpha_j^2(t)\,dt < \infty, \quad \text{for } j = 1, \dots, p.$$

(A2) $$\sup_{t \in [0,1]} EY_{1j}^4(t) < \infty, \quad \text{for } j = 1, \dots, p.$$

(A3) $$\inf_{t \in [0,1]} EY_{1j}^2(t) > 0, \quad \text{for } j = 1, \dots, p.$$

(A4) $$\sup_{t \in [0,1]} \frac{\left[EY_{1j}(t)Y_{1k}(t)\right]^2}{EY_{1j}^2(t)EY_{1k}^2(t)} < (p-1)^{-2},$$

for all $1 \le j < k \le p$, applicable for $p \ge 2$.

(A5) The function

$$\nu_j(t) = E\left[\int_0^t Y_{1j}^2(s)\,d\langle M_1\rangle_s\right], \qquad t \in [0,1],$$

is absolutely continuous with bounded derivative (Lebesgue a.e.) for $j = 1, \dots, p$.

(A6) For $1 \le k \le l \le \infty$, let $\mathscr{G}_k^l$ denote the σ-algebra generated by $\{\mathscr{F}_{i1}: k \le i \le l\}$, and denote

$$\varphi(n) = \sup_{k \ge 1} \sup_{\substack{A \in \mathscr{G}_1^k,\, P(A) > 0 \\ B \in \mathscr{G}_{k+n}^\infty}} |P(B|A) - P(B)|.$$

Assume that the following φ-mixing condition holds

$$\sum_{n \ge 1} \varphi^{1/2}(n) < \infty.$$

REMARKS.

(i) Assumptions (A3) and (A4) can be regarded as identifiability criteria. It is easy to construct examples which violate each of these assumptions and for which α is nonidentifiable. Note that the expression inside the supremum in (A4) is bounded above by 1 (Cauchy–Schwarz inequality) so that for $p = 2$ (A4) is a very weak requirement. As p increases it quickly becomes a rather severe condition on the covariates.

(ii) The fourth moment assumption (A2) is also required in the analysis of some Cox-type regression models; see Prentice and Self (1983, page 812).

(iii) Assumptions (A1), (A2), (A5), and stationarity ensure the existence of the integrals in (2.3) and (2.4). The martingale integral $\int_0^1 \phi_{jr}(t) Y_{ij}(t)\, dM_i(t)$ is defined since

$$E\left[\int_0^1 \phi_{jr}(t) Y_{ij}(t)\, dM_i(t)\right]^2 = E\left[\int_0^1 \phi_{jr}^2(t) Y_{ij}^2(t)\, d\langle M_i\rangle_t\right]$$
$$= \int_0^1 \phi_{jr}^2(t)\, d\nu_j(t) < \infty.$$

(iv) Examples of processes satisfying the assumptions (A1)–(A6) are described in Section 3. In the important cases of diffusion processes and point processes (A5) as a consequence of (A2).

3. Example.

3.1. Diffusion processes. Let $\alpha(t)$, $t \in [0,1]$ be a continuous function and $b(x)$, $\sigma(t,x)$, $t \in [0,1]$, and $x \in \mathscr{R}$ satisfy the following Lipschitz and growth conditions:

$$\text{(C1)} \qquad |b(x) - b(y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \le K|x-y|^2,$$

$$\text{(C2)} \qquad b^2(x) + \sigma^2(t,x) \le K(1 + x^2),$$

where K is a constant. Let $W = (W_t, \mathscr{F}_t)$ be a Wiener process and η an $\mathscr{F}_0$-measurable random variable. Under these conditions the stochastic differential equation

$$\text{(3.1)} \qquad dX_t = \alpha(t) b(X_t)\, dt + \sigma(t, X_t)\, dW_t, \qquad t \in [0,1],\ X_0 = \eta,$$

has a unique solution $X = (X_t, \mathscr{F}_t)$. If the function b is known and n iid copies of X are observed, then (3.1) can be expressed in the form of the semimartingale regression model (1.1) where $M_t = \int_0^t \sigma(s, X_s)\, dW_s$ and $Y(t) = b(X_t)$. Conditions (C1), (C2) and the following additional conditions (C3) and (C4) are sufficient for Theorem 2.1 to be applicable.

$$\text{(C3)} \qquad E\eta^4 < \infty,$$

$$\text{(C4)} \qquad b(X_t) \text{ vanishes a.s. for no } t \in [0,1].$$

Assumptions (A2) and (A5) can be checked using (C3) from which a result of Liptser and Shiryayev (1973, Theorem 4.6) gives $\sup_{t \in [0,1]} EX_t^4 < \infty$. (A5) then follows from (C2) and the fact that $\langle M \rangle_t = \int_0^t \sigma^2(s, X_s)\, ds$. (A3) follows from (C4), the a.s. path continuity of X and the continuity of b.

Special cases of (3.1) were treated by Grenander (1981) and Geman and Hwang (1982) who considered the case $b(x) = 1$, $\sigma(t, x) = 1$, and Nguyen and Pham (1983) who took $b(x) = x$, $\sigma(t, x) = 1$.

3.2. Point processes. Let $N = (N(t), \mathscr{F}_t)$ be a point process with intensity

$$\lambda(t) = \sum_{j=1}^{p} \alpha_j(t) Y_j(t), \tag{3.2}$$

where α_j is an unknown, continuous, nonnegative function and Y_j is an observable, nonnegative, $(\mathscr{F}_t)$-predictable process, $j = 1, \dots, p$. Assuming that $EN(1) < \infty$, there is a square integrable martingale $(M_t, \mathscr{F}_t)$ such that

$$N_t = \int_0^t \lambda(s)\, ds + M_t, \qquad t \in [0,1], \tag{3.3}$$

(see Aalen (1978)) and this is also a form of the semimartingale regression model (1.1). Assumption (A5) is a consequence of (A2) in this case since $\langle M \rangle_t = \int_0^t \lambda(s)\, ds$.

A practical example of this model might arise in which $\lambda(t)$ is the hazard rate for the incidence of cancer in a subject who at age t has had a cumulative exposure $Y_j(t)$ to each of $j = 1, \dots, p$ carcinogens and for whom N is the point process with a single jump at the time of initial detection of cancer. λ is set to zero after cancer is detected. The functions $\alpha_1, \dots, \alpha_p$ in this example represent the change in the relative hazard rates for the p carcinogens with age.

The model (3.2) was introduced by Aalen (1978, 1980) as an alternative to the proportional-hazard regression model of Cox (1972). Aalen provided an estimator for the cumulative hazard function $\int_0^t \alpha_j(s)\, ds$ rather than α_j itself. For the case $p = 1$, Ramlau-Hansen (1983) has used kernel function methods from density estimation and Karr (1983) has used the method of sieves to obtain estimators of α_1. It is not clear that these two approaches can be extended to $p \geq 1$.

3.3. Processes with both diffusion process and point process components. Let $\beta(t)$, $t \in [0,1]$ be a continuous function, $N = (N(t), \mathscr{F}_t)$ the point process of Section 3.2, $b(x), \sigma(t, x), \eta$ as in Section 3.1, and $\varepsilon > 0$. Then the equation

$$X_t = \eta + \int_0^t \beta(s) b(X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s + \varepsilon N_t, \tag{3.4}$$

has a unique solution $X = (X_t)$ which behaves as a diffusion process between the jump times of the point process. The size of the (positive) jumps of X is given by ε, which is assumed to be known. By substituting (3.3) into (3.4) we obtain another example of the semimartingale regression model (1.1) from which the functions $\beta, \alpha_1, \dots, \alpha_p$ can be estimated.

3.4. Censoring. In many practical situations the available data have been randomly censored. The possibility of censoring is easily incorporated into the semimartingale regression model (1.1) as follows. Suppose that the state X_t and covariates Y_{ij} of the ith subject are observable only up to an $(\mathscr{F}_{it})$ stopping time τ_i and that $(\tau_i, i \geq 1)$ is a stationary sequence of random variables. Define new state and covariate processes (which are observable over the whole of $[0,1]$) by the stopped processes $\tilde{X}_i(t) = X_i(t \wedge \tau_i)$ and $\tilde{Y}_{ij}(t) = Y_{ij}(t \wedge \tau_i)$, respectively. Equivalently, $\overline{Y}_{ij}(t) = I(t \leq \tau_i)Y_{ij}(t)$ could be used in place of $\tilde{Y}_{ij}(t)$. Also define a new square integrable martingale, $\tilde{M}_i(t) = M_i(t \wedge \tau_i)$. The censored version of the model is formed by replacing X, Y, and M in (1.1) by $\tilde{X}$, $\tilde{Y}$, and $\tilde{M}$, respectively. The assumptions of Theorem 2.1 should now be checked for the stopped processes $\tilde{X}$, $\tilde{Y}$, and $\tilde{M}$.

For $\tilde{M}_1$ to satisfy (A5) it is necessary that $P(\tau_1 > t) > 0$, for all $0 \leq t < 1$. This follows from the fact that $\langle \tilde{M}_1 \rangle_t = \tilde{M}_1^2(\tau_1)$ on $\{\omega\colon t \geq \tau_1\}$, P a.s. In some applications it is reasonable to assume that the censoring is independent of the subject (i.e., τ_i is independent of X_i, Y_{ij}, and M_i). In this case, by using the covariate $\overline{Y}_{ij}(t) = I(t \leq \tau_i)Y_{ij}(t)$, for which the quantity $P(\tau_1 \geq t)$ factors out of expressions in (A2)–(A5), it suffices to check (A2)–(A5) for the unstopped processes and have $P(\tau_1 \geq 1) > 0$.

Estimation for an example of the censored semimartingale regression model arising in neurophysiology is discussed by Habib and McKeague (1985).

4. Proofs. The measures μ_j, $j = 1, \dots, p$ defined below play an important role in the proof of Theorem 2.1. Define μ_j by $d\mu_j(t) = EY_j^2(t)\,dt$. Under assumptions (A2) and (A3) we have $L^2([0,1], dt) = L^2([0,1], d\mu_j)$ as sets and the norms are equivalent. As in Nguyen and Pham (1982), there exists a complete orthonormal sequence $(\Psi_{jr}, r \geq 1)$ in $L^2([0,1], d\mu_j)$ such that

$$\operatorname{span}\{\Psi_{jr}, r = 1, \dots, d_n\} = \operatorname{span}\{\phi_{jr}, r = 1, \dots, d_n\}.$$

The coordinates of α_j and $\hat{\alpha}_j^{(n)}$ in the basis $(\Psi_{jr}, r \geq 1)$ are denoted ξ_{jr}, $r \geq 1$ and $\hat{\xi}_{jr}^{(n)}$, $r = 1, \dots, d_n$, respectively. Let $\xi_j^{(n)} = (\xi_{jr}, r = 1, \dots, d_n)'$, $\hat{\xi}_j^{(n)} = (\hat{\xi}_{jr}^{(n)}, r = 1, \dots, d_n)'$. It is clear that to establish Theorem 2.1 it suffices to show that

$$\left\| \hat{\xi}_j^{(n)} - \xi_j^{(n)} \right\| \to_p 0, \quad \text{as } n \to \infty. \tag{4.1}$$

By (2.2) the $p \times d_n$ matrix $\hat{\xi}^{(n)} \equiv (\hat{\xi}_{jr}^{(n)})$ satisfies

$$\operatorname{vec}(\hat{\xi}^{(n)}) = a^{(n)-1}\operatorname{vec}(b^{(n)}), \tag{4.2}$$

where $b^{(n)}$ is the $p \times d_n$ matrix given by

$$b_{jr}^{(n)} = n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t) Y_{ij}(t)\, dX_i(t), \tag{4.3}$$

$a^{(n)}$ is the $pd_n \times pd_n$ matrix partitioned into p^2 submatrices $a_{jk}^{(n)}$ of order $d_n \times d_n$ with

$$a_{jkrl}^{(n)} = n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t)\Psi_{kl}(t) Y_{ij}(t) Y_{ik}(t)\, dt, \tag{4.4}$$

and $a^{(n)-1}$ is a generalized inverse of $a^{(n)}$. Let $\xi^{(n)}$ denote the $p \times d_n$ matrix with elements ξ_{jr} and $\alpha_j^{(n)}$ the protection of α_j onto $\operatorname{span}\{\Psi_{jr}, r = 1, \dots, d_n\}$, so that $\alpha_j^{(n)} = \sum_{r=1}^{d_n} \xi_{jr} \Psi_{jr}$. Using (1.1) to expand (4.3) it is easily checked that

$$\operatorname{vec}\big(\hat{\xi}^{(n)} - \xi^{(n)}\big) = a^{(n)-1} \operatorname{vec} c^{(n)}, \tag{4.5}$$

where $c^{(n)}$ is the $p \times d_n$ matrix with

$$\begin{aligned} c_{jr}^{(n)} = {} & n^{-1} \sum_{k=1}^{p} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t) Y_{ij}(t) Y_{ik}(t) \{\alpha_k(t) - \alpha_k^{(n)}(t)\}\, dt \\ & + n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t) Y_{ij}(t)\, dM_i(t). \end{aligned} \tag{4.6}$$

By (4.1) and (4.5) we have that Theorem 2.1 follows from the next two results which hold under the conditions of the theorem.

LEMMA 4.1. $\|\operatorname{vec} c^{(n)}\| \to_p 0$, *as* $n \to \infty$.

LEMMA 4.2. $\{\|a^{(n)-1}\|, n \geq 1\}$ *is a tight sequence of random variables* ($\|\cdot\|$ *denotes operator norm*).

The following elementary inequality is used in the proof of Lemmas 4.1 and 4.2.

LEMMA 4.3. *Let* $(Z_t, t \in [0,1])$ *be a measurable stochastic process such that* $K \equiv \sup_{t \in [0,1]} EZ_t^2 < \infty$. *Then for any integrable function* h,

$$E\left[\int_0^1 h_t Z_t\, dt\right]^2 \leq K \left[\int_0^1 |h_t|\, dt\right]^2.$$

PROOF.

$$\begin{aligned} E\left[\int_0^1 h_t Z_t\, dt\right]^2 &= \int_0^1 \int_0^1 h_s h_t E[Z_s Z_t]\, ds\, dt \\ &\leq \int_0^1 \int_0^1 |h_s|\,|h_t| [EZ_s^2]^{1/2} [EZ_t^2]^{1/2}\, ds\, dt \\ &\leq K \left[\int_0^1 |h_t|\, dt\right]^2. \square \end{aligned}$$

PROOF OF LEMMA 4.1. From (4.6) we can write

$$c_{jr}^{(n)} = \gamma_{jr}^{(n)} + \eta_{jr}^{(n)} + \rho_{jr}^{(n)},$$

where

$$\begin{aligned} \gamma_{jr}^{(n)} &= n^{-1} \sum_{k=1}^{p} \sum_{i=1}^{n} \big(\tilde{\gamma}_{jrik}^{(n)} - E\tilde{\gamma}_{jr1k}^{(n)}\big), \\ \tilde{\gamma}_{jrik}^{(n)} &= \int_0^1 \Psi_{jr}(t) Y_{ij}(t) Y_{ik}(t) \{\alpha_k(t) - \alpha_k^{(n)}(t)\}\, dt, \\ \eta_{jr}^{(n)} &= \sum_{k=1}^{p} E\tilde{\gamma}_{jr1k}^{(n)}, \\ \rho_{jr}^{(n)} &= n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t) Y_{ij}(t)\, dM_i(t). \end{aligned}$$

Using (A3) and since ($\Psi_{jr}, r = 1, \dots, d_n$) are orthonormal in $L^2([0,1], d\mu_j)$, we have

$$\begin{aligned}
\|\mathrm{vec}\,\eta^{(n)}\|^2 &\le p \sum_{k=1}^{p} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left\{ E\tilde{\gamma}_{jr1k}^{(n)} \right\}^2 \\
&= p \sum_{j,k=1}^{p} \sum_{r=1}^{d_n} \left\{ \int_0^1 \Psi_{jr}(t) \left(\frac{EY_{1j}(t)Y_{1k}(t)}{EY_{1j}^2(t)} \right) \left(\alpha_k(t) - \alpha_k^{(n)}(t) \right) EY_{1j}^2(t)\,dt \right\}^2 \\
&\le p \sum_{j,k=1}^{p} \int_0^1 \left(\frac{EY_{1j}(t)Y_{1k}(t)}{EY_{1j}^2(t)} \right)^2 \left(\alpha_k(t) - \alpha_k^{(n)}(t) \right)^2 EY_{1j}^2(t)\,dt \\
&\qquad\qquad \text{(by Bessel's inequality)} \\
&\le p^2 \sum_{k=1}^{p} \int_0^1 \left(\alpha_k(t) - \alpha_k^{(n)}(t) \right)^2 EY_{1k}^2(t)\,dt,
\end{aligned}$$

by the Cauchy–Schwarz inequality. It follows that $\|\mathrm{vec}\,\eta^{(n)}\| \to 0$, as $n \to \infty$, since $\alpha_k^{(n)} \to \alpha_k$ in $L^2([0,1], \mu_k)$ for each $k = 1, \dots, p$. Next, by stationarity of $(Y_{ij}, i \ge 1)$,

$$(4.7)\quad E\left(\gamma_{jr}^{(n)}\right)^2 \le p \sum_{k=1}^{p} \left[n^{-1} \mathrm{var}\left(\tilde{\gamma}_{jr1k}^{(n)} \right) + 2n^{-2} \sum_{i=2}^{n} (n - i + 1) \mathrm{cov}\left(\tilde{\gamma}_{jr1k}^{(n)}, \tilde{\gamma}_{jrik}^{(n)} \right) \right].$$

By Lemma 4.3 and (A2) there are constants K_1, K_2, such that

$$\begin{aligned}
E\left(\tilde{\gamma}_{jr1k}^{(n)} \right)^2 &\le K_1 \left[\int_0^1 \left| \Psi_{jr}(t) \right| \left| \alpha_k(t) - \alpha_k^{(n)}(t) \right| dt \right]^2 \\
&\le K_2 \int_0^1 \left[\alpha_k(t) - \alpha_k^{(n)}(t) \right]^2 dt \to 0,
\end{aligned}$$

as $n \to \infty$. By a result of Ibragimov and Linnik (1971, Lemma 17.2.3),

$$\mathrm{cov}\left(\tilde{\gamma}_{jr1k}^{(n)}, \tilde{\gamma}_{jr1k}^{(n)} \right) \le 2\varphi^{1/2}(i-1) E\left(\tilde{\gamma}_{jr1k}^{(n)} \right)^2,$$

so that from (4.7) and (A6), $E(\gamma_{jr}^{(n)})^2 = o(n^{-1})$, uniformly in j and r, as $n \to \infty$. Thus, since $d_n = o(n)$, we have

$$E\|\mathrm{vec}\,\gamma^{(n)}\|^2 = \sum_{j=1}^{p} \sum_{r=1}^{d_n} E\left(\gamma_{jr}^{(n)} \right)^2 = o(1).$$

Finally consider $\rho_{jr}^{(n)}$. Using a property of stochastic integrals with respect to square integrable martingales, we have

$$\begin{aligned}
E\left[\int_0^1 \Psi_{jr}(t) Y_{ij}(t)\, dM_i(t) \right]^2 &= E\left[\int_0^1 \Psi_{jr}^2(t) Y_{ij}^2(t)\, d\langle M_i \rangle_t \right] \\
&= \int_0^1 \Psi_{jr}^2(t)\, d\nu_j(t),
\end{aligned}$$

which is uniformly bounded in j and r by (A5). Then, using the mixing condition

(A6) as before, it can be checked that $E\|\mathrm{vec}\,\rho^{(n)}\|^2 = o(1)$. Collecting terms we obtain

$$E\|\mathrm{vec}\,c^{(n)}\|^2 \le 3\big(E\|\mathrm{vec}\,\gamma^{(n)}\|^2 + \|\mathrm{vec}\,\eta^{(n)}\|^2 + E\|\mathrm{vec}\,\rho^{(n)}\|^2\big)$$
$$\to 0, \quad \text{as } n \to \infty. \ \square$$

PROOF OF LEMMA 4.2. We can write $\alpha^{(n)} = \beta^{(n)} + \zeta^{(n)}$, where $\beta^{(n)}$ and $\zeta^{(n)}$ are partitioned in the same way as $\alpha^{(n)}$,

$$\beta_{jkrl}^{(n)} = n^{-1} \sum_{i=1}^{n} \big(\tilde{\beta}_{jkrl}^{(i)} - E\tilde{\beta}_{jkrl}^{(1)}\big),$$
$$\tilde{\beta}_{jkrl}^{(i)} = \int_0^1 \Psi_{jr}(t)\Psi_{kl}(t)Y_{ij}(t)Y_{ik}(t)\,dt,$$
$$\zeta_{jkrl}^{(n)} = E\tilde{\beta}_{jkrl}^{(1)}.$$

Using a similar argument to the estimation of $E(\gamma_{jr}^{(n)})^2$ in the proof of Lemma 4.1, there is a constant K such that

$$E\big(\beta_{jkrl}^{(n)}\big)^2 \le \frac{K}{n} E\big(\tilde{\beta}_{jkrl}^{(1)}\big)^2.$$

Thus,

$$\begin{aligned}
E\|\beta^{(n)}\|^2 &\le \sum_{j,k=1}^{p} \sum_{r,l=1}^{d_n} E\big(\beta_{jkrl}^{(n)}\big)^2 \\
&\le \frac{K}{n} \sum_{j,k=1}^{p} \sum_{r,l=1}^{d_n} E\big(\tilde{\beta}_{jkrl}^{(1)}\big)^2 \\
&= \frac{K}{n} \sum_{j,k=1}^{p} \sum_{r,l=1}^{d_n} E\Big(\int_0^1 \Psi_{jr}(t)\Psi_{kl}(t)Y_{1j}(t)Y_{1k}(t)\,dt\Big)^2 \\
&= \frac{K}{n} \sum_{j,k=1}^{p} \sum_{l=1}^{d_n} E\left\{ \sum_{r=1}^{d_n} \left[\int_0^1 \Psi_{jr}(t) \left(\frac{\Psi_{kl}(t)Y_{1j}(t)Y_{1k}(t)}{EY_{1j}^2(t)} \right) EY_{1j}^2(t)\,dt \right]^2 \right\} \\
&\le \frac{K}{n} \sum_{j,k=1}^{p} \sum_{l=1}^{d_n} E\left\{ \int_0^1 \left(\frac{\Psi_{kl}(t)Y_{1j}(t)Y_{1k}(t)}{EY_{1j}^2(t)} \right)^2 EY_{1j}^2(t)\,dt \right\} \\
&\qquad\qquad \text{(by Bessel's inequality)} \\
&= \frac{K}{n} \sum_{j,k=1}^{p} \sum_{l=1}^{d_n} \int_0^1 \Psi_{kl}^2(t) \left\{ \frac{EY_{1j}^2(t)Y_{1k}^2(t)}{EY_{1j}^2(t)} \right\} dt \\
&= O\left(\frac{d_n}{n}\right),
\end{aligned}$$

by (A2), (A3), and the Cauchy–Schwarz inequality. It follows that

$$E\|\beta^{(n)}\|^2 = o(1), \quad \text{as } n \to \infty. \tag{4.8}$$

Now we consider the behavior of $\zeta^{(n)}$ as $n \to \infty$. First note that the diagonal submatrices of $\zeta^{(n)}$ are identity matrices. Thus, letting $I^{(n)}$ denote the $pd_n \times pd_n$ identity matrix, we have directly from the definition of the operator norm that

$$\|I^{(n)} - \zeta^{(n)}\|^2 = \sup_{\|\operatorname{vec} x^{(n)}\| \le 1} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left[\sum_{\substack{k=1\\k\ne j}}^{p} \sum_{l=1}^{d_n} \zeta^{(n)}_{jkrl} x_{kl} \right]^2$$

$$(4.9) \qquad = \sup_{\|\operatorname{vec} x^{(n)}\| \le 1} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left[\sum_{\substack{k=1\\k\ne j}}^{p} \int_0^1 \Psi_{jr}(t) \left\{ \sum_{l=1}^{d_n} x_{kl} \Psi_{kl}(t) \right\} \times EY_{1j}(t) Y_{1k}(t)\, dt \right]^2,$$

where the supremum is over the set of $p \times d_n$ matrices $x^{(n)}$ with $\|\operatorname{vec} x^{(n)}\| \le 1$. Let $\|\cdot\|_k$ denote the norm in $L^2([0,1], d\mu_k)$ and

$$H = \left\{ h = (h_1, \dots, h_p) \colon h_k \in L^2([0,1], d\mu_k) \text{ for } k = 1, \dots, p \text{ and } \sum_{k=1}^{p} \|h_k\|_k^2 \le 1 \right\}.$$

Then it follows from (4.9) that

$$\|I^{(n)} - \zeta^{(n)}\|^2 \le \sup_{h \in H} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left[\int_0^1 \Psi_{jr}(t) \left\{ \sum_{\substack{k=1\\k\ne j}}^{p} h_k(t) \frac{EY_{1j}(t) Y_{1k}(t)}{EY_{1j}^2(t)} \right\} d\mu_j(t) \right]^2$$

$$(4.10) \qquad \le \sup_{h \in H} \sum_{j=1}^{p} \int_0^1 \left(\sum_{\substack{k=1\\k\ne j}}^{p} h_k(t) \frac{EY_{1j}(t) Y_{1k}(t)}{EY_{1j}^2(t)} \right)^2 d\mu_j(t)$$

$$\le (p-1) \sup_{h \in H} \sum_{\substack{j,k=1\\j\ne k}}^{p} \int_0^1 h_k^2(t) \frac{\left[EY_{1j}(t) Y_{1k}(t)\right]^2}{EY_{1j}^2(t) EY_{1k}^2(t)}\, d\mu_k(t)$$

$$\le (p-1)^2 \delta,$$

where

$$\delta = \sup_{\substack{t \in [0,1]\\ j \ne k}} \frac{\left[EY_{1j}(t) Y_{1k}(t)\right]^2}{EY_{1j}^2(t) EY_{1k}^2(t)}.$$

By (A4) there exists a constant c such that $(p-1)^2 \delta < c < 1$. Then, by (4.8) and (4.10),

$$P\{\|I^{(n)} - a^{(n)}\| < c\} \to 1, \quad \text{as } n \to \infty.$$

But if V is any $pd_n \times pd_n$ matrix such that $\|I^{(n)} - V\| < 1$ then V is invertible and $\|V^{-1}\| \le (1 - \|I^{(n)} - V\|)^{-1}$. It follows that

$$P(a^{(n)} \text{ is invertible}) \to 1, \quad \text{as } n \to \infty$$

and

$$P\Big(\|a^{(n)-1}\| < (1-c)^{-1}\Big) \to 1, \quad \text{as } n \to \infty. \square$$

Acknowledgments. I would like to thank Muhammad Habib for several helpful conversations during the preparation of this paper. Thanks also to a referee for detailed comments which led to a significant improvement in the paper.

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. In *Lecture Notes in Statistics* **2** 1–25. Springer, New York.

ANTONIADIS, A. (1985). Parametric estimation for the mean of a Gaussian process by the method of sieves. Unpublished.

CENCOV, N. N. (1962). Estimation of an unknown distribution density from observations. *Soviet Math.* **3** 1559–1562.

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

GEMAN, S. and HWANG, C. R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.

HABIB, M. K. and MCKEAGUE, I. W. (1985). Parameter estimation for nonstationary diffusion models of neurons. O.N.R. Technical Report 4, Department of Biostatistics, Univ. North Carolina, Chapel Hill.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Walters-Noordhoff, Netherlands.

KARR, A. F. (1983). Maximum likelihood estimation in the multiplicative intensity model. Technical Report 46, Center for Stochastic Processes, Department of Statistics, Univ. North Carolina, Chapel Hill.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes* **1**. Springer, New York.

MEYER, P. A. (1976). Un cours sur les integrales stochastiques. *Lecture Notes in Math.* **511** 245–400. Springer, Berlin.

NGUYEN, H. T. and PHAM, T. D. (1982). Identification of nonstationary diffusion model by the method of sieves. *SIAM J. Control Optim.* **20** 603–611.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306