ESTIMATION FOR A SEMIMARTINGALE REGRESSION MODEL
USING THE METHOD OF SIEVES

by

Ian W. McKeague

Florida State University
Department of Statistics
Tallahassee, Florida 32306

FSU Statistics Report M686
USARO Technical Report No.D-70

August, 1984

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

Key Words: Inference for stochastic processes, method of sieves, regression analysis, semimartingales, point process, diffusion process, censoring.

AMS(1980) Subject classifications. Primary 62M09, 62G05; secondary 60G44.
ESTIMATION FOR A SEMIMARTINGALE REGRESSION MODEL
USING THE METHOD OF SIEVES

by

Ian W. McKeague

ABSTRACT

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), and the hazard function in the multiplicative intensity model for point processes (Karr 1983).

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 prominent 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 \( i \)th subject at time \( t \), and suppose that \( X = (X_1, \ldots, X_n)^T \) satisfies

\[
X(t) = X(0) + \int_0^t Y(s)\alpha(s)ds + M(t), \quad t \in [0,1],
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_p)^T \) is a vector of unknown non-random functions, \( Y = (Y_{ij}) \) is the \( n \times p \) matrix of covariate processes, with \( Y_{ij} \) being the \( j \)th covariate for the \( i \)th subject, and \( M = (M_1, \ldots, M_n)^T \) where each \( M_i \) is square integrable martingale.

It is of interest to estimate the functions \( \alpha_1, \ldots, \alpha_p \) 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 \( \alpha \) and state conditions under which \( \int_0^1 [\hat{\alpha}_{j}^{(n)}(t) - \alpha_j(t)]^2dt \to 0 \) in probability as \( n \to \infty \), for \( j=1, \ldots, p \). Our approach is based on the orthogonal functions sieve technique developed for linear diffusion processes by Nguyen and Pham (1982). Applications are discussed in section 3. All proofs are contained in section 4.
2. Estimation of $\alpha$.

We begin by stating some technical assumptions needed in the semimartingale regression model (1.1). $(\Omega, F, P)$ will denote a complete probability space and for each $i=1, \ldots, n$, $(F_{it}, t \in [0,1])$ is a nondecreasing right-continuous family of sub-$\sigma$-fields of $F$ where $F_{i0}$ contains all $P$-null sets in $F$. All processes are indexed by $t \in [0,1]$. Each process $(M_i(t), F_{it})$, $i=1, \ldots, p$ 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, $(F_{it})$ 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 $(F_{it})$ predictable, that is measurable with respect to the $\sigma$-field on $[0,1] \times \Omega$ generated by all left-continuous, $(F_{it})$ adapted processes. The $\sigma$-field $F_{it}$ represents the state and covariate history of the $i$th 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 $\alpha_j$ to be the element of span $\{\phi_{jr}, r=1, \ldots, d_n\}$ given by

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

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

$$
A^{(n)} \text{vec}(\hat{\alpha}^{(n)}) = \text{vec}(B^{(n)}),
$$
where the vec operator takes a matrix and stacks the columns one on top of the other to form a large column vector, $B^{(n)}$ is the $p \times d_n$ matrix given by

$$B_{jr}^{(n)} = \sum_{i=1}^{n} \int_{0}^{1} \phi_{jr}(t) Y_{ij}(t) dX_i(t),$$

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

$$A_{jk}^{(n)} = \sum_{i=1}^{n} \int_{0}^{1} \phi_{jr}(t) \phi_{kl}(t) Y_{ij}(t) Y_{ik}(t) dt.$$

It follows from the proof of Lemma 4.3 that $A^{(n)}$ is invertible for $n$ sufficiently large almost surely, so that for $n$ sufficiently large, (2.2) has a unique solution

$$\text{vec}(\hat{\alpha}^{(n)}) = A^{(n)^{-1}} \text{vec}(B^{(n)}).$$

In the case of diffusion processes, $\hat{\alpha}^{(n)}_j$ 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 $\hat{\alpha}^{(n)}_j$ 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 $d_n \rightarrow \infty$ such that $d_n = o(n^{1/2})$,

$$\int_{0}^{1} [\hat{\alpha}^{(n)}_j(t) - \alpha_j(t)]^2 dt \Rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

**Assumptions**

(A1) \( \int_{0}^{1} \alpha^2_j(t) dt < \infty, \) for j=1, ..., p.

(A2) \( \sup_{t \in [0,1]} \mathbb{E} \alpha^4_j(t) < \infty, \) for j=1, ..., p.

(A3) \( \inf_{t \in [0,1]} \mathbb{E} \alpha^2_j(t) > 0, \) for j=1, ..., p.
\[ (A4) \quad \sup_{t \in [0,1]} \frac{[EY_{1j}(t)Y_{1k}(t)]^2}{EY_{1j}(t)EY_{1k}(t)} < (p-1)^{-2}, \]

for all \( 1 \leq j < k \leq p \), applicable for \( p \geq 2 \).

\[ (A5) \quad \text{Let } \nu_j \text{ be a measure defined on } L[0,1] \text{ by} \]
\[ \nu_j(A) = E[\int_A Y^2(t) \, d\mathcal{L}_j(t)], \quad A \in L[0,1], \]

and let \( m \) denote Lebesgue measure on \( L[0,1] \). Assume that \( \nu_j \ll m \) and there exists a version of \( d\nu_j / dm \) which is bounded on \([0,1]\), for \( j=1, \ldots, p \).

\[ (A6) \quad \text{For } 1 \leq k \leq \ell \leq \infty, \text{ let } G^\ell_k \text{ denote the } \sigma\text{-algebra generated by} \]
\[ \{F_{i1} \mid k \leq i \leq \ell\}, \text{ and denote} \]
\[ \varphi(n) = \sup_{k \geq 1} \sup_{A \in G^k_1, P(A) > 0} \frac{|P(AnB) - P(A)P(B)|}{P(A)}. \]

Assume that the following \( \ell \)-mixing condition holds
\[ \sum_{n \geq 1} \varphi^{\ell}_n(n) < \infty. \]

**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 \( \alpha \) is non-identifiable. 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 forth moment assumption A2 is also required in the analysis of some Cox-type regression models, see Prentice and Self (1983, p.812).
(iii) Assumptions A1, A2, A5 and stationarity ensure the existence of the integrals in (2.3) and (2.4). The martingale integral \( \int_0^1 \phi_i(t)Y_{ij}(t)dM_i(t) \) is defined since
\[
E[\int_0^1 \phi_i(t)Y_{ij}(t)d<M_i>_t]^2 = \int_0^1 \phi_i^2(t)d\nu_j(t) < \infty.
\]

(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 is a consequence of A2.

3. Examples

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

(C1) \[ |b(x) - b(y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \leq K|x-y|^2 \]
(C2) \[ b^2(x) + \sigma^2(t,x) \leq K(1+x^2), \]

where K is a constant. Let \( W = (W_t, F_t) \) be a Wiener process and \( \eta \) an \( F_0 \)-measurable random variable. Under these conditions the stochastic differential equation

(3.1) \[ dX_t = a(t)b(X_t)dt + \sigma(t,X_t)dW_t, \quad t \in [0,1], \quad X_0 = \eta, \]

has a unique solution \( X = (X_t, F_t) \). 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, C4 are sufficient for Theorem 2.1 to be applicable.

(C3) \[ E\eta^4 < \infty \]
(C4) \[ b(X_t) \] 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 Shiryaev (1978, Theorem 4.6) gives \( \sup_{t \in [0,1]} \mathrm{EX}_t^h < \infty \). 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 \( N = (N(t), F_t) \) be a point process with intensity

\[
\lambda(t) = \sum_{j=1}^{p} \alpha_j(t) Y_j(t),
\]

where \( \alpha_j \) is an unknown, continuous, nonnegative function and \( Y_j \) is an observable, nonnegative, \( (F_t) \)-predictable process, \( j = 1, \ldots, p \). Assuming that \( EN(1) < \infty \), there is a square integrable martingale \( (M_t, F_t) \) such that

\[
N_t = \int_0^t \lambda(s) ds + M_t, \quad t \in [0,1],
\]

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 \( j = 1, \ldots, p \) carcinogens and for whom \( N \) is the point process with a single jump at the time of initial detection of cancer. The functions \( \alpha_1, \ldots, \alpha_p \) in this example represent the change in the relative hazard rates for the p carcinogens with age.

The model (3.3) 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 \( \int_0^t a_j(s)ds \) rather than \( a_j \) 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 \( a_1 \). 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, \( N = (N(t), F_t) \) the point process of section (3.2), \( b(x), \sigma(t,x), \eta \) as in section (3.1), \( \epsilon > 0 \). Then the equation

\[
X_t = \eta + \int_0^t \beta(s)b(X_s)ds + \int_0^t \sigma(s,X_s)dW_s + \epsilon N_t,
\]

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 \( \epsilon \), which is assumed to be known. By substituting (3.4) into (3.5) we obtain another example of the semimartingale regression model (1.1) from which the functions \( \beta, a_1, \ldots, a_p \) can be estimated.

(3.4) Censoring. In many practical situations the available data has been randomly censored. The possibility of censoring is easily incorporated into the semimartingale regression model (1.1) as follows. Suppose that the state \( X_i \) and covariates \( Y_{ij} \) of the \( i \)th subject are observable only up to an \( (F_{it}) \) stopping time \( \tau_i \) 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, \( \tilde{Y}_{ij}(t) = I(t \leq \tau_i)Y_{ij}(t) \) 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 \( M_1 \) to satisfy A5 it is necessary that 
\( P(\tau_1 > t) > 0 \), for all \( 0 \leq t < 1 \). This follows from the fact that 
\( \langle \tilde{M}_1 \rangle_t = \tilde{M}_1(\tau_1) \) on \( \{ \omega : t \geq \tau_1 \} \), P a.s.. In 
some applications it is reasonable to assume that the censoring is independent 
of the subject (i.e. \( \tau_1 \) is independent of \( X_i, Y_{ij} \) and \( M_1 \)). In this case, by 
using the covariate \( \tilde{Y}_{ij}(t) = I(t \leq \tau_1)Y_{ij}(t) \), 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 (1984).

4. Proofs.

The measures \( \mu_j, j = 1, \ldots, p \) defined below play an important role in 
the proof of Theorem 2.1. Define \( \mu_j \) on \( L[0,1] \) 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 
ortonormal sequence \( (\psi_{jr}, r \geq 1) \) in \( L^2([0,1], d\mu_j) \) such that 
\[
\text{span } \{ \psi_{jr}, r=1, \ldots, d_n \} = \text{span } \{ \phi_{jr}, r=1, \ldots, d_n \}.
\]
The coordinates of \( \alpha_j \) and \( \alpha_j^{(n)} \) in the basis \( (\psi_{jr}, r \geq 1) \) are denoted \( \xi_{jr}, r \geq 1 \) 
and \( \xi_j^{(n)} = (\xi_{jr}, r=1, \ldots, d_n)^\top \), 
\( \xi_j^{(n)} = (\xi_{jr}^{(n)}, r=1, \ldots, d_n)^\top \). It is clear that to establish Theorem 2.1 it 
suffices to show that 
\[
(4.1) \quad \| \xi_j^{(n)} - \xi_j \|_E \to 0, \text{ as } n \to \infty.
\]
By (2.5), for \( n \) sufficiently large the \( p \times d_n \) matrix \( \xi_j^{(n)} = (\xi_{jr}^{(n)}) \) satisfies 
\[
(4.2) \quad \text{vec}(\xi^{(n)}) = a(n)^{-1}\text{vec}(b^{(n)}),
\]
where \( b^{(n)} \) is the \( p \times d_n \) matrix given by 
\[
(4.3) \quad b_{jr}^{(n)} = n^{-1} \sum_{i=1}^n \int_0^{\tau_j(t)} Y_{ij}(t)dX_i(t),
\]
and \( a^{(n)} \) is the \( p d_n \times pd_n \) matrix partitioned into \( p^2 \) submatrices \( a^{(n)}_{jk} \) of order \( d_n \times d_n \) with

\[
(4.4) \quad a^{(n)}_{jk} = n^{-1} \sum_{i=1}^{n} \int_{0}^{t} \psi_{jr}(t) \psi_{k\ell}(t) Y_{ij}(t) Y_{ik}(t) dt.
\]

For convenience assume that \( a^{(n)} \) is invertible for all \( n \geq 1 \). Let \( \xi^{(n)} \) denote the \( p \times d_n \) matrix with elements \( \xi_{jr} \) and \( a^{(n)}_j \) the projection of \( a_j \) onto \( \text{span} \{ \psi_{jr}, r=1, \ldots, d_n \} \), so that \( a^{(n)}_j = \sum_{r=1}^{d_n} \xi_{jr} \psi_{jr} \). Using (1.1) to expand (4.3) it is easily checked that

\[
(4.5) \quad \text{vec}(\xi^{(n)} \cdot \xi^{(n)}) = a^{(n)} - 1 \cdot \text{vec} \ c^{(n)},
\]

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

\[
(4.6) \quad c^{(n)}_{jr} = n^{-1} \sum_{k=1}^{P} \sum_{i=1}^{n} \int_{0}^{t} \psi_{jr}(t) Y_{ij}(t) Y_{ik}(t) (a_k(t) - a_k^{(n)}(t)) dt
\]

\[
+ n^{-1} \int_{0}^{t} \psi_{jr}(t) Y_{ij}(t) dM_{ij}(t).
\]

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.** \( \| \text{vec} \ c^{(n)} \| \to 0 \), 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 stochastic process such that

\[ K = \sup_{t \in [0,1]} \mathbb{E} Z_t^2 < \infty. \]

Then for any integrable function \( h \),

\[ E \left[ \int_{0}^{1} h_t Z_t dt \right]^2 \leq K \int_{0}^{1} |h_t|^2 dt^2. \]
Proof. \[
\mathbb{E}[\int_0^t h_s Z_t \, dt]^2 = \int_0^1 \int_0^1 h_s h_t \mathbb{E}[Z_s Z_t] \, dsdt
\]
\[
\leq \int_0^1 \int_0^1 |h_s| |h_t| [\mathbb{E}Z_s^2]^{\frac{h_s}{2}} [\mathbb{E}Z_t^2]^{\frac{h_t}{2}} \, dsdt
\]
\[
\leq K [\int_0^1 |h_t| \, dt]^2
\]

Proof of Lemma 4.1. From (4.6) we can write

\[
c_{jr}^{(n)} = \gamma_{jr}^{(n)} + \eta_{jr}^{(n)} + \rho_{jr}^{(n)},
\]
where

\[
\gamma_{jr}^{(n)} = n^{-1} \frac{p}{\sum} \sum_{k=1}^{n} (\gamma_{jrik}^{(n)} - \mathbb{E}v_{jrik}^{(n)}),
\]
\[
\gamma_{jrik}^{(n)} = \int_0^1 \psi_{jr}(t) Y_{ij}(t) Y_{ik}(t) \{a_k(t) - a_k^{(n)}(t)\} \, dt,
\]
\[
\eta_{jr}^{(n)} = \frac{p}{\sum} \mathbb{E}v_{jrik}^{(n)},
\]
\[
\rho_{jr}^{(n)} = n^{-1} \frac{n}{\sum} \int_0^1 \psi_{jr}(t) Y_{ij}(t) \, dM_i(t).
\]

Using A3 and since \((\psi_{jr}, r=1, \ldots, d_n)\) are orthonormal in \(L^2([0,1], d\nu_j)\), we have

\[
||\text{vec } \eta^{(n)}||^2 \leq \frac{p}{\sum} \sum_{k=1}^{p} \sum_{j=1}^{d_n} [\mathbb{E}v_{jrik}^{(n)}]^2
\]
\[
\leq \frac{p}{\sum} \sum_{k=1}^{p} \sum_{j=1}^{d_n} \int_0^1 \frac{1}{\mathbb{E}Y_{1j}(t) Y_{1k}(t)} \{a_k(t) - a_k^{(n)}(t)\}^2 \, dt
\]
\[
\leq p^2 \frac{p}{\sum} \int_0^1 \{a_k(t) - a_k^{(n)}(t)\}^2 \, dt \to 0,
\]
as \(n \to \infty\). Next, by stationarity of \((Y_{ij}, j \geq 1)\),
(4.7) \( E(\gamma_{jr}^{(n)} )^2 \leq \frac{p}{k=1} \sum \frac{1}{n} \text{Var}(\gamma_{jrk}^{(n)}) + 2n^{-2} \sum_{i=2}^{n} (n-i+1) \text{Cov}(\gamma_{jrk}^{(n)}, \gamma_{jrik}^{(n)})]. \)

By Lemma 1 and A2 there are constants \( K_1, K_2, \) such that

\[
E(\gamma_{jrk}^{(n)})^2 \leq K_1 \left[ \int_0^1 |\gamma_{jrk}(t)| \alpha_k(t) - \alpha_k^{(n)}(t) | dt \right]^2
\]

\[
\leq K_2 \int_0^1 (\alpha_k(t) - \alpha_k^{(n)}(t))^2 dt \to 0,
\]

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

\[
\text{Cov}(\gamma_{jrk}^{(n)}, \gamma_{jrik}^{(n)}) \leq 2 \gamma_{jrk}^{(i)} E(\gamma_{jrk}^{(n)})^2,
\]

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 \( d_n = o(n^{k_2}), \) we have

\[
E \| \text{vec} \gamma_{jr}^{(n)} \|^2 = \sum_{j=1}^{p} \sum_{r=1}^{d_n} E(\gamma_{jr}^{(n)})^2 = o(n^{-k_2}).
\]

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

\[
E[\int_0^1 \gamma_{jr}(t) \gamma_{ij}(t) dM_i(t)]^2 = E[\int_0^1 \gamma_{jr}^2(t) \gamma_{ij}^2(t) d< M_i > t]
\]

\[
= \frac{1}{0} \int_0^1 \gamma_{jr}^2(t) dt
\]

which is uniformly bounded in \( j \) and \( r \) by A5. Then, using the mixing condition A6 as before, it can be checked that \( E \| \text{vec} \rho_{jr}^{(n)} \|^2 = o(n^{-k_2}). \) Collecting terms we obtain

\[
E \| \text{vec} c^{(n)} \|^2 \leq 3(E \| \text{vec} \gamma_{jr}^{(n)} \|^2 + \| \text{vec} n^{(n)} \|^2 + E \| \text{vec} \rho_{jr}^{(n)} \|^2)
\]

\(+ 0, \text{ as } n \to \infty. \)

\[]
Proof of Lemma 4.2. We can write \( a^{(n)} = \beta^{(n)} + \xi^{(n)} \), where \( \beta^{(n)} \) and \( \xi^{(n)} \) are partitioned in the same way as \( a^{(n)} \),

\[
\beta^{(n)}_{jkr\ell} = n^{-1} \sum_{i=1}^{n} \left( \beta^{(i)}_{jkr\ell} - E \beta_{jkr\ell}^{(1)} \right),
\]

\[
\beta^{(i)}_{jkr\ell} = \int_{0}^{\psi_{jr}(t) \psi_{kl}(t) Y_{ij}(t) Y_{ik}(t) dt},
\]

\[
\xi^{(n)}_{jkr\ell} = E \beta_{jkr\ell}^{(1)}.
\]

Using A2, A6 and a similar argument in the proof of Lemma 4.1, it can be checked that \( E(\beta^{(n)}_{jkr\ell})^2 = o(n^{-1}) \), uniformly in \( j, k, r, \) and \( \ell \), as \( n \to \infty \). Thus, since \( d_n = o(n^{k/2}) \),

\[
E \| \beta^{(n)} \|^2 \leq \sum_{j,k=1}^{p} \sum_{r,\ell=1}^{d_n} E(\beta^{(n)}_{jkr\ell})^2 = o(1),
\]

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

\[
\| I^{(n)} - \xi^{(n)} \|^2 = \sup_{\| \text{vec} x^{(n)} \| \leq 1} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left[ \sum_{k=1}^{p} \sum_{\ell=1}^{d_n} \xi^{(n)}_{jkr\ell} x_{k\ell} \right]^2
\]

\[
= \sup_{\| \text{vec} x^{(n)} \| \leq 1} \sum_{j=1}^{p} \sum_{r=1}^{d_n} \left[ \sum_{k=1}^{p} \int_{0}^{\psi_{jr}(t)} \psi_{kl}(t) \left\{ \sum_{\ell=1}^{d_n} \xi_{jkr\ell} \psi_{k\ell}(t) \right\} E Y_{ij}(t) Y_{ik}(t) dt \right]^2,
\]

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

$$H = \{ h = (h_1, \ldots, h_p) : h_k \in L^2([0,1],d\nu_k) \text{ for } k=1, \ldots, p \text{ and } \sum_{k=1}^{p} \|h_k\|_{k}^{2} \leq 1 \}. $$

Then it follows from (4.9) that

$$\| I^{(n)} - \zeta^{(n)} \|^{2} \leq $$

$$\sup_{h \in H} \sum_{j=1}^{p} \sum_{r=1}^{d_{n}} \int_{0}^{1} \psi_{jr}(t) \left\{ \sum_{k=1}^{p} \frac{EY_{ij}(t)Y_{1k}(t)}{EY_{1j}(t)} h_k(t) \right\}^{2} d\nu_{j}(t) $$

$$\leq \sup_{h \in H} \sum_{j=1}^{p} \int_{0}^{1} \left\{ \sum_{k=1}^{p} \frac{EY_{ij}(t)Y_{1k}(t)}{EY_{1j}(t)} h_k(t) \right\}^{2} d\nu_{j}(t) $$

$$\leq (p-1) \sup_{h \in H} \sum_{j=1}^{p} \int_{0}^{1} \frac{[EY_{ij}(t)Y_{1k}(t)]^{2}}{EY_{1j}(t)EY_{1k}(t)} d\nu_{k}(t)$$

(4.10) \leq (p-1)^2 \delta ,

where $\delta = \sup_{t \in [0,1]} \frac{[EY_{ij}(t)Y_{1k}(t)]^{2}}{EY_{1j}(t)EY_{1k}(t)}$. By A4 there exists a constant $c$ such that $(p-1)^2 \delta < c < 1$. Then, by (4.8) and (4.10),

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

Now using the inequality $\| V^{-1} \| \leq (1 - \| I^{(n)} - V \|)^{-1}$, valid for any $pd_n \times pd_n$ matrix $V$ such that $\| I^{(n)} - V \| < 1$, we conclude that

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

Acknowledgment. I would like to thank Muhammad Habib for several helpful conversations during the preparation of this paper.
References


ESTIMATION FOR A SEMIMARTINGALE REGRESSION MODEL USING THE METHOD OF SIEVES

Ian W. McKeage

Department of Statistics
Florida State University
Tallahassee, Florida 32306

U.S. Army Research Office-Durham
P.O. Box 12211
Research Triangle Park, NC 27709

Distribution unlimited

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

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.