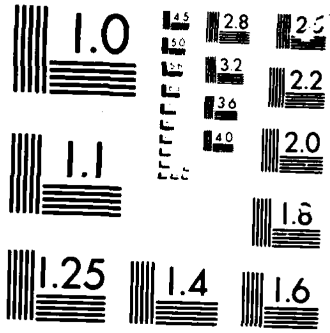
AD-A169 207   ASYMPTOTIC THEORY FOR SIEVE ESTIMATORS IN   1/1
SEMIMARTINGALE REGRESSION MODEL..(U) FLORIDA STATE UNIV
TALLAHASSEE DEPT OF STATISTICS I W MCKEAGUE MAR 86

UNCLASSIFIED   FSU-STATISTICS-M722 ARO-19367.35-MA   F/G 12/1   NL



MICROCOP' HART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

AD-A169 207

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 19367.35-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Asymptotic Theory for Sieve Estimators in Semimartingale Regression Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ian W. McKeague		8. CONTRACT OR GRANT NUMBER(s) DAAG29-82-K-0168
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University Tallahassee, Florida 32306-3033		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE March, 1986
		13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE JUN 27 1986 S D

D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper studies the estimation of functions $\alpha_1, \ldots, \alpha_p$ describing the temporal influence of p covariate processes in a regression model for semimartingales. McKeague (1986) introduced sieve estimators for $\alpha_1, \ldots, \alpha_p$ and

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

86

20. ABSTRACT CONTINUED

established consistency in L^2-norm. In the present paper the asymptotic distribution theory for the integrated sieve estimators is developed. Smoothed sieve estimators are shown to be pointwise consistent and rates of convergence are provided.

ARo 19367.35-MA

ASYMPTOTIC THEORY FOR SIEVE ESTIMATORS IN
SEMIMARTINGALE REGRESSION MODELS

by

Ian W. McKeague

FSU Technical Report No. M 722
USARO Technical Report No. D-86

March, 1986

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

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

Key words and phrases: Inference for stochastic processes, semimartingale, diffusion process, point process, method of sieves, kernel function smoothing.

AMS (1980) Subject Classification: Primary 62M09, 62G05; Secondary 60G44.

ASYMPTOTIC THEORY FOR SIEVE ESTIMATORS IN SEMIMARTINGALE REGRESSION MODELS

by

Ian. W. McKeague

ABSTRACT

This paper studies the estimation of functions $\alpha_1, \ldots, \alpha_p$ describing the temporal influence of p covariate processes in a regression model for semimartingales. McKeague (1986) introduced sieve estimators for $\alpha_1, \ldots, \alpha_p$ and established consistency in L^2-norm. In ~~the present~~ paper the asymptotic distribution theory for the integrated sieve estimators is developed. Smoothed sieve estimators are shown to be pointwise consistent and rates of convergence are provided.

1. Introduction

This work is a sequel to McKeague (1986) in which statistical estimation for a nonparametric regression model for semimartingales was introduced. The model is given by

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

where M is a square integrable martingale and

$$\lambda(s) = \Sigma_{j=1}^{p} \alpha_j(s) Y_j(s). \tag{1.2}$$

Here $\alpha_1, \ldots, \alpha_p$ are deterministic functions of time and $Y_1, \ldots, Y_p$ are predictable covariate processes. Grenander's method of sieves was used to obtain estimators, denoted $\hat{\alpha}_j^{(n)}$, of α_j, $j = 1, \ldots, p$ based on n replicates of X and its covariates. These estimators were shown to be consistent in L^2-norm as $n \to \infty$.

Our aim in this paper is to obtain some results on the asymptotic distirbutions and rates of convergence of the sieve estimators. As in density estimation a satisfactory distribution theory is possible only for the integrated estimator given by

$$\hat{A}_j^{(n)}(t) = \int_0^t \hat{\alpha}_j^{(n)}(s)ds, \tag{1.3}$$

where $1 \le j \le p$. Also, to obtain rate of convergence results we need to look at a smoothed sieve estimator

$$\tilde{\alpha}_j^{(n)}(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) \hat{\alpha}_j^{(n)}(s)ds, \tag{1.4}$$

where K is a function with integral 1, called the kernel function, and $b_n > 0$ is a bandwidth parameter. Our results make it possible to test hypotheses concerning $\alpha_1, \ldots, \alpha_p$ and to construct confidence bands for

$$A_j(t) = \int_0^t \alpha_j(s)ds, \qquad j=1, \ldots, p. \tag{1.5}$$

The model (1.1) contains a number of important special cases. Grenander (1981), Ibragimov and Khasminskii (1981) and Geman and Hwang (1982) have studied the case where $p = 1$, $Y_1 \equiv 1$ and M is a Wiener process. Nguyen and Pham (1982) have treated the linear diffusion process with $p = 1$. Aalen (1980) has studied the point process case for general $p \geq 1$. Aalen provided estimators of $A_1, \ldots, A_p$ but it has not been possible to obtain consistency or asymptotic distribution results for these estimators except when $p = 1$. The importance of the point process version of (1.1) is that it provides an alternative to the regression model of Cox (1972) for the analysis of censored survival data. A practical example might arise in which $\lambda(t)$ is the hazard rate for the incidence of cancer in a subject who has been exposed to p carcinogens, where $Y_j(t)$ is the cumulative exposure to the jth carcinogen and $\alpha_j(t)$ represents the relative hazard rate of the jth carcinogen at age t.

Our asymptotic distribution results are given in section 2. In section 3 we state results on the rate of convergence of the smoothed sieve estimators. All proof are contained in section 4.

2. Asymptotic distributions.

Let $(X_i, M_i, Y_{ij}, j = 1, \ldots, p)$, $i = 1, \ldots, n$, denote n independent copies of the generic processes X, M and Y_j, $j = 1, \ldots, p$ which satisfy the model (1.1). For each $j = 1, \ldots, p$ let $(\phi_{jr}, r \geq 1)$ be a complete orthonormal sequence in $L^2[0,1]$. The sieve estimator $\hat{\alpha}_j^{(n)}$ for α_j, which was introduced by McKeague (1986), is defined as follows. Let

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

where (d_n) is an increasing sequence of positive integers and $\hat{\alpha}_{jr}^{(n)}$ is the jr element in the $p \times d_n$ matrix $\hat{\alpha}^{(n)}$ defined by

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

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

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

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

$$A_{jkr\ell}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \phi_{jr}(t) \phi_{k\ell}(t) Y_{ij}(t) Y_{ik}(t) dt.$$

In equation (2.2), $A^{(n)-1}$ is a generalized inverse of $A^{(n)}$ whose choice does not affect the asymptotic behavior of $\hat{\alpha}_j^{(n)}$.

Define measures μ_j, $j = 1, \ldots, p$ by $d\mu_j(t) = EY_j^2(t)dt$. The projections of α_j onto span $\{\phi_{jr}, r = 1, \ldots, d_n\}$ in $L^2([0, 1], d\mu_j)$ and $L^2([0, 1], dt)$ are denoted $\alpha_j^{(n)}$ and $\alpha_j^{\{n\}}$ respectively.

Write $A_j^{(n)}(t) = \int_0^t \alpha_j^{(n)}(s)ds$, $A_j^{\{n\}}(t) = \int_0^t \alpha_j^{\{n\}}(s)ds$. A function f is said to be Lipschitz of order γ, where $0 < \gamma \le 1$, if there is a constant C such that for all s, t in the domain of f, $|f(t) - f(s)| \le C|t - s|^\gamma$. If f is Lipschitz of order 1 we simply say it is Lipschitz. The predictable quadratic variation of M is denoted $<M>$.

The following assumptions were used in McKeague (1986) to obtain consistency of $\hat{\alpha}_j^{(n)}$ in L^2-norm.

(A1) $\int_0^1 \alpha_j^2(t)dt < \infty$ for $j = 1, \ldots, p$.

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

(A3) $\inf_{t\in[0,1]} EY_j^2(t) > 0$ for $j = 1, \ldots, p$.

(A4) $$\sup_{t\in[0,1]} \frac{|EY_j(t)Y_k(t)|}{[EY_j^2(t)]^{1/2}[EY_k^2(t)]^{1/2}} < \frac{1}{p-1}$$

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

(A5) The function

$$\nu_j(t) = E\left[\int_0^t Y_j^2(s)d<M>_s\right], \qquad t \in [0,1]$$

is Lipschitz, for $j = 1, \ldots, p$.

Additional assumptions needed for our various weak convergence results are now stated.

(S1) $d_n\int_0^1[\alpha_j(t) - \alpha_j^{\{n\}}(t)]^2dt \longrightarrow 0,$ as $n \to \infty$.

(S2) $n\int_0^1[\alpha_j(t) - \alpha_j^{\{n\}}(t)]^2dt \longrightarrow 0,$ as $n \to \infty$.

(B1)
$$\frac{\sup_{t\epsilon[0,1]} |EY_j(t)Y_k(t)|}{[\inf_{t\epsilon[0,1]} EY_j^2(t)]^{\frac{1}{2}}[\inf_{t\epsilon[0,1]} EY_k^2(t)]^{\frac{1}{2}}} < \frac{1}{p-1}$$

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

(B2) $\langle M\rangle$ is absolutely continuous and

$$\sup_{t\epsilon[0,1]} E[Y_j^4(t)(\frac{d\langle M\rangle}{dt})^2] < \infty, \qquad \text{for } j=1, \ldots, p.$$

Our first result deals with weak convergence of finite dimensional distributions.

Theorem 2.1. Suppose that (A1)-(A5), (S1) hold and $d_n \to \infty$, $d_n = o(n^{\frac{1}{2}})$. Then for $h_j \epsilon L^2[0,1]$, $j=1, \ldots, p$

$$(\sqrt{n}\int_0^1 h_j(t)[\hat{\alpha}_j^{(n)}(t) - \alpha_j^{(n)}(t)]dt)_{j=1}^p \xrightarrow{\mathcal{D}} N(0,\Sigma)$$

where $\Sigma = (\sigma_{jk})$ is the p×p matrix defined by

$$\sigma_{jk} = E\int_0^1 H_j(t)H_k(t)Y_j(t)Y_k(t)d\langle M\rangle_t \tag{2.3}$$

$$H = (H_1, \ldots, H_p)', \ h = (h_1, \ldots, h_p)'$$

$$H(t) = L(t)h(t), \qquad L(t) = [K(t)]^{-1} \tag{2.4}$$

$K(t)$ is the $p \times p$ matrix with components

$$K_{jk}(t) = E[Y_j(t)Y_k(t)]. \tag{2.5}$$

Corollary 2.2. Under the conditions of Theorem 2.1

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j^{(n)}) \xrightarrow{\mathcal{D}_f} m_j,$$

where $\mathcal{D}_f$ denotes convergence of finite dimensional distributions and m_j is a continuous Gaussian martingale with $m_j(0) = 0$,

$$\mathrm{Cov}(m_j(s),\ m_j(t)) = E\int_0^{s\wedge t} L^2{}_{jj}(u) Y^2{}_j(u)\, d\langle M\rangle_u. \tag{2.6}$$

Corollary 2.3. Under the conditions of Theorem 2.1 but with (A4) replaced by the stronger condition (B1)

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j^{\{n\}}) \xrightarrow{\mathcal{D}_f} m_j.$$

Corollary 2.4. Under the conditions of Theorem 2.1 but with (S1) replaced by the stronger condtion (S2)

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j) \xrightarrow{\mathcal{D}_f} m_j.$$

If X is assumed to be a continuous process (equivalently the martingale M is continuous) then it is possible to strengthen these results to weak convergence in the function space C[0,1].

<u>Theorem 2.5</u>. Suppose that (A1)-(A4), (S1) and (B2) hold, $d_n \to \infty$, $d_n = o(n^{\frac{1}{2}})$ and M is a continuous martingale. Then

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j^{(n)}) \xrightarrow{\mathcal{D}} m_j \quad \text{in } C[0,1].$$

<u>Corollary 2.6</u>. Under the conditions of Theorem 2.5 but with (A4) replaced by the stronger condition (B2)

$$\sqrt{n}(\hat{A}_j^{(n)} - A_n^{\{n\}}) \xrightarrow{\mathcal{D}} m_j \quad \text{in } C[0,1].$$

<u>Corollary 2.7</u>. Under the conditions of Theorem 2.5 but with (S1) replaced by the stronger condition (S2)

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j) \xrightarrow{\mathcal{D}} m_j \quad \text{in } C[0,1].$$

Under appropriate smoothness conditions on α_j condition (S1) or (S2) can be satisfied by a careful choice of the sieve and (d_n). We mention two important sieves for which this is possible.

<u>(1) The Fourier sieve</u>. Take $\phi_{jr} = \phi_r$, $j = 1, \ldots, p$ where $\phi_1(t) \equiv 1$ and for $r \geq 2$,

$$\phi_r(t) = \begin{cases} \sqrt{2}\cos(\pi r t), & r \text{ even} \\ \sqrt{2}\sin(\pi(r+1)t), & r \text{ odd}. \end{cases} \tag{2.7}$$

Then (S1) is satisfied under the following conditons:

(F1) $\alpha_j(0) = \alpha_j(1)$;

(F2) The extension of α_j to a function of period 1 on $(-\infty,\infty)$ is Lipschitz of order $\gamma > \frac{1}{2}$.

The stronger assumption (S2) is satisfied provided (F1) and the following conditions hold:

(F3) The extension of α_j to a function of period 1 on $(-\infty,\infty)$ has Lipschitz first derivative;

(F4) $d_n^4/n \longrightarrow \infty$.

These facts are consequences of a result in approximation theory known as Jackson's inequality, see Lemma 4.4.

(2) The Walsh sieve. The following definition of the Walsh functions is due to Paley (1932), other definitions can be found in Beauchamp (1975). First define the Rademacher functions $(\psi_r,\ r \geq 0)$ on $[0,1)$ by

$$\psi_r(t) = (-1)^k \qquad \text{if} \qquad k2^{-(r+1)} \leq t < (k+1)2^{-(r+1)}$$

Then put $\phi_0(t) \equiv 1$ and for $r = 2^{r_1} + 2^{r_2} + \ldots + 2^{r_\nu}$, with $r_1 > r_2 > \ldots > r_\nu \geq 0$ let

$$\phi_r(t) = \psi_{r_1}(t)\psi_{r_2}(t) \ \ldots\ \psi_{r_\nu}(t) \qquad \text{for } t \in [0,1).$$

ϕ_r is called the rth Walsh function in Paley ordering. Results of Fine (1955, p. 394) show that the Walsh sieve, defined by taking ϕ_{jr} as the rth Walsh function, satisfies assumption (S1) provided α_j is Lipschitz of order $\gamma > \frac{1}{2}$. Unfortunately, in general (S2) is not satisfied for the Walsh sieve unless the rate of increase of d_n is prohibitively large. However some examples discussed in Beauchamp (1975, p. 34) show that the Walsh sieve would be preferable to the Fourier sieve if the α_j's have a rectangular form. Also the Walsh sieve has an advantage over the Fourier sieve in terms of computational simplicity.

In order to use the results of this section to obtain confidence intervals and confidence bands for $A_j(t)$ it is first necessary to estimate the function $G_j(t) = Em_j^2(t)$. The matrix $K(t)$ given in (2.5) can be estimated by $\hat{K}^{(n)}(t) = (\hat{K}_{jk}^{(n)}(t))$, where $\hat{K}_{jk}^{(n)}(t) = n^{-1}\sum_{i=1}^{n} Y_{ij}(t)Y_{ik}(t)$.

Assuming that the covariate processes have paths in $D[0,1]$ and

$$E \sup_{t\in[0,1]} |Y_j(t)Y_k(t)| < \infty$$

it follows from Ranga Rao (1963, Theorem 1) that

$$\sup_{t\in[0,1]} |\hat{K}_{jk}^{(n)}(t) - K_{jk}(t)| \xrightarrow{a.s.} 0. \tag{2.8}$$

If M is a standard Wiener process $G_j(t)$ can be estimated by

$$\hat{G}_j^{(n)}(t) = \int_0^t \hat{L}_{jj}^{(n)}(s))\,\hat{K}_{jj}^{(n)}(s)ds,$$

where $\hat{L}^{(n)}(s) = (\hat{L}_{jk}^{(n)}(s))$ is a generalized inverse of $K^{(n)}(s)$. Then from (2.8)

$$\sup_{t\in[0,1]} |\hat{G}_j^{(n)}(t) - G_j(t)| \xrightarrow{a.s.} 0. \tag{2.9}$$

If X is a point process the predictable quadratic variation process $\langle M\rangle_t$ is $\int_0^t \lambda(s)ds$ which involves $\alpha_1, \ldots, \alpha_p$. In turn $G_j(t)$ involves the unknown

$\alpha_1, \ldots, \alpha_p$. However, as we have seen, $\hat{\alpha}_j^{(n)}$ is a consistent estimator of α_j in L^2-norm, so it is possible to consistently estimate $G_j(t)$ in this case as well. It follows from Corollary 2.4 that upper and lower limits of a $100(1-\alpha)$ percent confidence interval for $A_j(t)$ are given by $\hat{A}_j^{(n)}(t) \pm z_{\alpha/2} n^{-\frac{1}{2}} \hat{G}_j^{(n)}(t)^{\frac{1}{2}}$, where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

If M is a standard Wiener process we can give a confidence *band* for A_j. First note that $m_j(t) G_j(1)^{\frac{1}{2}} (G_j(1) + G_j(t))^{-1}$ is distributed as $W^o\left(\frac{G_j(t)}{G_j(1) + G_j(t)}\right)$, where W^o is the Brownian bridge on $[0,1]$. It follows from (2.9) and Corollary 2.7 that upper and lower limits of a $100(1-\alpha)$ percent confidence band for A_j are given by

$$\hat{A}_j^{(n)}(t) \pm c_\alpha n^{-\frac{1}{2}} \hat{G}_j^{(n)}(1)^{\frac{1}{2}} \left(1 + \frac{\hat{G}_j^{(n)}(t)}{\hat{G}_j^{(n)}(1)}\right), \qquad t \in [0,1]$$

where c_α is the upper α quantile of the distribution of $\sup_{t \in [0,1]} |W^o(t)|$. A table for this distribution can be found in Hall and Wellner (1980).

The assumptions (A4) and (B1) can be weakened by assuming that the sequence of eigenvalues of the matrices $R^{(n)}$, $n \geq 1$, defined in Lemma 4.3, is bounded away from 0. However in practice it probably would not be necessary to go to the trouble of checking this given a lack of any obvious collinearity in the covariates.

3. Smoothing the sieve estimators

Ramlau-Hansen (1983) has used the methods of kernel density estimation to smooth the Nelson-Aalen estimator and obtain a pointwise consistent estimator of the hazard function. We now apply kernel function smoothing methods to the sieve estimators. Consider the smoothed sieve estimator $\tilde{\alpha}_j^{(n)}$ defined by (1.4) where, for simplicity, we assume that the kernel function K has support $[-1,1]$. The following result shows how the sieve dimension d_n and the bandwidth parameter b_n can be specified so that $\tilde{\alpha}_j^{(n)}$ is a pointwise consistent estimator of α_j.

Theorem 3.1. Suppose that the Fourier sieve is used, conditions (A2), (A3), (A5), (B1), (F1) and (F3) hold and K is differentiable with Lipschitz derivative. Let

$$d_n = \lfloor n^{\delta} \rfloor, \qquad \text{where } \frac{5}{12} < \delta < \frac{1}{2},$$

$$b_n = n^{-\beta}, \qquad \text{where } \frac{1}{3} < \beta < \frac{4}{5}\delta.$$

Then for each $t \in (0,1)$

$$\tilde{\alpha}_j^{(n)}(t) - \alpha_j(t) = O_p(n^{-\frac{1}{2}(1-\beta)}). \tag{3.1}$$

If it is assumed that X is a continuous process then we can establish uniform consistency for $\tilde{\alpha}_j^{(n)}$. The rate of convergence in uniform metric is naturally slower than the pointwise rate given by (3.1).

Theorem 3.2. Suppose that (A1)-(A4), (S2) and (B2) hold, α_j and K are Lipschitz, $d_n \to \infty$, $d_n = o(n^{1/2})$, $b_n = n^{-\beta}$ where $\frac{1}{3} < \beta < \frac{1}{2}$, and M is a continuous martingale.

Then

$$\sup_{t\in[0,1]} | \tilde{\alpha}_j^{(n)}(t) - \alpha_j(t) | = O_p(n^{-\frac{1}{2}(1-2\beta)}).$$

4. Proofs of Theorems.

The following notation was used in McKeague (1986) where a more detailed discussion can be found.

Notation

$(\Psi_{jr}, r \geq 1)$ denotes a complete orthonormal sequence in $L^2([0,1], d\mu_j)$ such that $\text{span}\{\Psi_{jr}, r=1, \ldots, d_n\} = \text{span}\{\phi_{jr}, r=1, \ldots, d_n\}$ for all $n \geq 1$.

ξ_{jr}, $\hat{\xi}_{jr}^{(n)}$ denote the coordinates of $\alpha_j^{(n)}$ and $\hat{\alpha}_j^{(n)}$ with respect to the basis $(\Psi_{jr}, r \geq 1)$ in $L^2([0,1], d\mu_j)$, respectively.

$$a_{jkr\ell}^{(n)} = n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t)\Psi_{k\ell}(t)Y_{ij}(t)Y_{ik}(t)dt$$

$a^{(n)}$ denotes the $pd_n \times pd_n$ matrix partitioned into the p^2 submatrices

$$a_{jk}^{(n)} = (a_{jkr\ell}^{(n)}: r, \ell = 1, \ldots, d_n).$$

$$c_{jr}^{(n)} = n^{-1} \sum_{k=1}^{p} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t)Y_{ij}(t)Y_{ik}(t)\{\alpha_k(t) - \alpha_k^{(n)}(t)\}dt$$

$$+ n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t)Y_{ij}(t)dM_i(t)$$

$$\rho_{jr}^{(n)} = n^{-1} \sum_{i=1}^{n} \int_0^1 \Psi_{jr}(t) Y_{ij}(t)\, dM_i(t)$$

$$\zeta_{jkr\ell}^{(n)} = \int_0^1 \Psi_{jr}(t) \Psi_{k\ell}(t) E[Y_j(t) Y_k(t)]\, dt$$

$\zeta^{(n)}$ denotes the $pd_n \times pd_n$ matrix partitioned into the p^2 submatrices $\zeta_{jk}^{(n)} = (\zeta_{jkr\ell}^{(n)}: r, \ell = 1, \ldots, d_n)$.

$\xi^{(n)}$, $\hat{\xi}^{(n)}$, $c^{(n)}$, $\rho^{(n)}$ are $p \times d_n$ matrices defined by their entries given above.

It is easily checked that

$$\text{vec } (\hat{\xi}^{(n)} - \xi^{(n)}) = a^{(n)-1} \text{ vec } c^{(n)} \tag{4.1}$$

where $a^{(n)-1}$ is a generalized inverse of $a^{(n)}$. The next two lemmas collect various facts proved in McKeague (1986).

Lemma 4.1. Suppose that (A1)-(A3), (A5) hold. Then

(i) $E \| a^{(n)} - \zeta^{(n)} \|^2 = O(\frac{d_n}{n})$, where $\| \cdot \|$ denotes operator norm;

(ii) $E \| \text{vec } \rho^{(n)} \|^2 = O(\frac{d_n}{n})$;

(iii) $E \| \text{vec } (c^{(n)} - \rho^{(n)}) \|^2 = O(\sum_{j=1}^{p} \int_0^1 [\alpha_j(t) - \alpha_j^{\{n\}}(t)]^2 dt) + o(\frac{d_n}{n})$.

Lemma 4.2. (i)(A2)-(A4) imply that $\zeta^{(n)}$ is invertible for all $n \geq 1$ and $\sup_{n \geq 1} \| \zeta^{(n)-1} \| < \infty$.

(ii) (A1)-(A5) imply that $P(a^{(n)}$ is invertible) $\to 1$ and $\{ \| a^{(n)-1} \| , n \geq 1\}$ is a tight sequence of random variables.

<u>Proof of Theorem 2.1</u>. Writing $\hat{\alpha}_j^{(n)}$ and $\alpha_j^{(n)}$ in terms of the basis $(\Psi_{jr}, r \geq 1)$ we obtain

$$\sqrt{n} \sum_{j=1}^{p} \int_0^1 h_j(t)[\hat{\alpha}_j^{(n)}(t) - \alpha_j^{(n)}(t)]dt = \sqrt{n} \sum_{j=1}^{p} \sum_{r=1}^{d_n} h_{jr}(\hat{\xi}_{jr}^{(n)} - \xi_{jr}^{(n)}) \tag{4.2}$$

where $h_{jr} = \int_0^1 h_j(t)\Psi_{jr}(t)dt$. By Lemma 4.2 we may define a $p \times d_n$ matrix $\lambda^{(n)} = (\lambda_{jr}^{(n)})$ satisfying

$$vec\ (\lambda^{(n)}) = \zeta^{(n)-1}\ vec\ (h^{(n)}), \tag{4.3}$$

where $h^{(n)} = (h_{jr}^{(n)})$. Let $I^{(n)}$ denote the $pd_n \times pd_n$ identity matrix. Then using (4.1) we can split (4.2) into four parts

$$\sqrt{n}\ vec\ (\lambda^{(n)})'\ vec\ \rho^{(n)} \tag{4.4}$$

$$+ \sqrt{n}\ vec\ (\lambda^{(n)})'\ vec\ (c^{(n)} - \rho^{(n)}) \tag{4.5}$$

$$+ \sqrt{n}\ vec\ (\lambda^{(n)})'(a^{(n)}a^{(n)-1} - I^{(n)})\ vec\ c^{(n)} \tag{4.6}$$

$$+ \sqrt{n}\ vec\ (\lambda^{(n)})'(\zeta^{(n)} - a^{(n)})a^{(n)-1}\ vec\ c^{(n)}. \tag{4.7}$$

By Lemma 4.2(ii) (4.6) converges to zero in probability. (4.7) is bounded in absolute value by

$$\| \text{ vec } \lambda^{(n)} \| \quad [(\frac{n}{d_n})^{\frac{1}{2}} \| \zeta^{(n)} - a^{(n)} \|] \quad \| a^{(n)-1} \| [d_n^{\frac{1}{2}} \| \text{vec } c^{(n)} \|]$$

and this tends to zero in probability since

$$\sup_{n\geq 1} \| \text{vec } \lambda^{(n)} \| \leq [\sum_{j=1}^{p} \int_0^1 \frac{h_j^2(t)}{EY_j^2(t)} dt]^{\frac{1}{2}} \sup_{n\geq 1} \| \zeta^{(n)-1} \| < \infty \tag{4.8}$$

by Lemma 4.2(i), $\{\| a^{(n)-1} \| , n \geq 1\}$ is tight and $d_n^{\frac{1}{2}} \| \text{vec } c^{(n)} \| \xrightarrow{p} 0$ by Lemma 4.1 (ii), (iii), (S1) and the assumption $d_n = o(n^{\frac{1}{2}})$. Similarly (4.5) converges to zero in probability. It remains to consider (4.4) which we write in the form

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ni} \tag{4.9}$$

where

$$Z_{ni} = \sum_{j=1}^{p} \int_0^1 u_j^{(n)}(t) Y_{ij}(t) dM_j(t) \tag{4.10}$$

$$u_j^{(n)}(t) = \sum_{r=1}^{d_n} \lambda_{jr}^{(n)} \psi_{jr}(t). \tag{4.11}$$

Introduce

$$Z_{\infty i} = \sum_{j=1}^{p} \int_0^1 H_j(t) Y_{ij}(t) dM_i(t) \tag{4.12}$$

where H_j is defined by (2.4). In an Appendix (Lemma 4.6) we have used some operator theory to show that $u_j^{(n)} \to H_j$ in $L^2[0,1]$ as $n \to \infty$. Consequently $Z_{ni} \to Z_{\infty i}$ in $L^2(\Omega, F, P)$ as $n \to \infty$:

$$E(Z_{ni} - Z_{\infty i})^2 \le p \sum_{j=1}^{p} E\{\int_0^1 (u_j^{(n)}(t) - H_j(t))Y_j(t)dM(t)\}^2$$

$$= p \sum_{j=1}^{p} \int_0^1 (u_j^{(n)}(t) - H_j(t))^2 d\nu_j(t)$$

$\to 0$, by (A5). Now apply the Lindeberg-Feller Theorem to (4.9). Note that for $\varepsilon > 0$,

$$\frac{1}{n} \sum_{i=1}^{n} E(Z_{ni}^2 I(|Z_{ni}| > \varepsilon \sqrt{n})) = E(Z_{n1}^2 I(Z_{n1}^2 > \varepsilon^2 n))$$

$\to 0$, since $\{Z_{n1}^2, n \ge 1\}$ is uniformly integrable, so the Lindeberg condition is satisfied. Thus (4.9) converges in distribution to $N(0,\sigma^2)$, where $\sigma^2 = EZ_{\infty 1}^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} \sigma_{jk}$. An application of the Cramér-Wold device completes the proof of the theorem. □

Of the corollaries to Theorem 2.1 only Corollary 2.3 needs some explanation. The following lemma can be used to rework the proof of Theorem 2.1 in terms of the basis $(\phi_{jr}, r \ge 1)$ instead of $(\Psi_{jr}, r \ge 1)$ to yield Corollary 2.3. Note that we have replaced (A4) by the stronger condition (B1) in order to do this.

Lemma 4.3. Suppose that (A2), (A3), (B1) hold and let $R^{(n)}$ denote the $pd_n \times pd_n$ matrix partitioned into the p^2 submatrices $R_{jk}^{(n)}$, $j, k = 1, \ldots, p$ with entries

$$R_{jkr\ell}^{(n)} = \int_0^1 \phi_{jr}(t)\phi_{k\ell}(t)E[Y_j(t)Y_k(t)]dt. \qquad (4.13)$$

Then $R^{(n)}$ is invertible for all $n \geq 1$ and $\sup_{n\geq 1} \| R^{(n)-1} \| < \infty$.

<u>Proof</u>. Let $c_j = \inf_{t\in[0,1]} EY_j^2(t)$, $\bar{Y}_j(t) = c_j^{-1} Y_j(t)$ for $j=1,\ldots,p$ and let $\bar{R}^{(n)}$ be a $pd_n \times pd_n$ matrix partitioned the same way as $R^{(n)}$ but with entries

$$\bar{R}^{(n)}_{jkr\ell} = \int_0^1 \phi_{jr}(t)\phi_{k\ell}(t)E[\bar{Y}_j(t)\bar{Y}_k(t)]dt.$$

It suffices to show that $\bar{R}^{(n)}$ is invertible for all $n \geq 1$ and $\sup_{n\geq 1} \| \bar{R}^{(n)-1} \| < \infty$. For then the matrix with entries $D^{(n)}_{jkr\ell} = c_j^{-1} c_k^{-1} (\bar{R}^{(n)-1})_{jkr\ell}$ is an inverse for $R^{(n)}$ and

$$\sup_{n\geq 1} \| R^{(n)-1} \| \leq (\max_{j=1,\ldots,p} c_j^{-2}) \sup_{n\geq 1} \| \bar{R}^{(n)-1} \| < \infty.$$

Condition (B1) implies that

$$\sup_{t\in[0,1]} | E[\bar{Y}_j(t)\bar{Y}_k(t)] | < \frac{1}{p-1} \quad \text{for } j \neq k. \tag{4.14}$$

Also note that

$$\inf_{t\in[0,1]} E\bar{Y}_j^2(t) = 1 \qquad \text{for } j=1,\ldots,p.$$

Let $b_j = \sup_{t\in[0,1]} E\bar{Y}_j^2(t)$, $f_j(t) = 1 - b_j^{-1} E\bar{Y}_j^2(t)$ and I_n denote the $d_n \times d_n$ identity matrix.

Then

$$\| I_n - b_j^{-1}\bar{R}^{(n)} \|^2 \le \sup_{\substack{h\in L^2[0,1] \\ \|h\| \le 1}} \sum_{r=1}^{d_n} \left(\int_0^1 h(t)\phi_{jr}(t)f_j(t)dt\right)^2$$

$$\le \sup_{\|h\| \le 1} \int_0^1 h^2(t)f_j^2(t)dt \qquad \text{(by Bessel's inequality)}$$

$$\le \sup_{t\in[0,1]} f_j^2(t) < 1.$$

It follows that $\bar{R}_{jj}^{(n)}$ is invertible and

$$\| \bar{R}_{jj}^{(n)-1} \| = \frac{1}{b_j} \| (b_j^{-1}R_{jj}^{(n)})^{-1} \|$$

$$\le \frac{1}{b_j} \left(\frac{1}{1- \sup_{t\in[0,1]} f_j(t)}\right)$$

$$= \frac{1}{\inf_{t\in[0,1]} E\bar{Y}_j^2(t)} = 1.$$

Let $F^{(n)}$ denote the $pd_n\times pd_n$ matrix obtained by replacing all off diagonal submatrices of $\bar{R}^{(n)}$ by zero matrices. Let $G^{(n)} = \bar{R}^{(n)} - F^{(n)}$ and $I^{(n)}$ denote the $pd_n \times pd_n$ identity matrix. Then $F^{(n)}$ is invertible and

$$\| F^{(n)-1} \| \le \max_{j=1,\dots,p} \| \bar{R}_{jj}^{(n)-1} \| \le 1. \qquad (4.15)$$

Using (4.14) and an argument from McKeague (1986, 4.10)

$$\sup_{n\geq 1} \| G^{(n)} \| < 1,$$

so that

$$\sup_{n\geq 1} \| I^{(n)} - \bar{R}^{(n)}F^{(n)-1} \| \leq \sup_{n\geq 1}(\| G^{(n)} \| \, \| F^{(n)-1} \|) < 1.$$

Thus $\bar{R}^{(n)}F^{(n)-1}$ is invertible for all $n \geq 1$,

$$\sup_{n\geq 1} \| (\bar{R}^{(n)}F^{(n)-1})^{-1} \| < \infty,$$

and using (4.15) we conclude that $\bar{R}^{(n)}$ is invertible for all $n \geq 1$ and

$\sup_{n\geq 1} \| \bar{R}^{(n)-1} \| < \infty$. □

Proof of Theorem 2.5. Condition (B2) implies that (A5) holds, so that the conditions of Corollary 2.2 are satisfied and the finite dimensional distributions of $\sqrt{n}(\hat{A}_j^{(n)} - A_j^{(n)})$ converge. From the proof of Theorem 3.1 it can be seen that

$$\sqrt{n}(\hat{A}_j^{(n)}(t) - A_j^{(n)}(t)) = U_n(t) + V_n(t) \tag{4.16}$$

where

$$\sup_{t\in[0,1]} | U_n(t) | \xrightarrow{p} 0, \tag{4.17}$$

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ni}(t)$$

$$Z_{ni}(t) = \sum_{k=1}^{p} \int_0^1 u_k^{(n)}(s,t) Y_{ik}(s) dM_i(s)$$

$$u_k^{(n)}(s,t) = \sum_{r=1}^{d_n} \lambda_{kr}^{(n)}(t) \Psi_{kr}(s)$$

$$\text{vec}\,[\lambda^{(n)}(t)] = \zeta^{(n)-1}\ \text{vec}\,[h^{(n)}(t)]$$

$$h_{kr}^{(n)}(t) = \begin{cases} \int_0^1 1_{[0,t]}(s) \Psi_{jr}(s) ds & \text{for } k=j,\ r=1,\ \ldots,\ d_n \\ 0 & \text{for } k \neq j,\ r=1,\ \ldots,\ d_n. \end{cases}$$

It remains to show that the sequence of processes $\{V_n,\ n \geq 1\}$ is tight in $C[0,1]$. In what follows C denotes a postive constant independent of n whose value may change at each occurrence. From Billingsley (1968, Theorem 12.3) the required tightness is implied by the following condition: there exist constants $q \geq 0$, $\gamma > 1$ such that for all $n \geq 1$, t_1, $t_2 \in [0,1]$

$$E \mid V_n(t_2) - V_n(t_1) \mid^q \leq C \mid t_2 - t_1 \mid^\gamma. \tag{4.18}$$

The processes $Z_{n1}, \ldots, Z_{nn}$ are independent and have zero mean so that by the Marcinkiewicz-Zygmund inequality (see Chow and Teicher, 1978, p. 356) for $q \geq 1$

$$E\,|\,V_n(t_2) - V_n(t_1)\,|^q \le C\ E\{\frac{1}{n}\sum_{i=1}^{n}[Z_{ni}(t_2) - Z_{ni}(t_1)]^2\}^{q/2}$$

$$\le C\ E\,|\,Z_{n1}(t_2) - Z_{n1}(t_1)\,|^q$$

$$\le C\sum_{k=1}^{p} E\,|\int_0^1\{u_k^{(n)}(s,t_2) - u_k^{(n)}(s,t_1)\}Y_k(s)dM_s\,|^q. \qquad (4.19)$$

For fixed t_1, t_2, k, n define the process

$$N_t = \int_0^t\{u_k^{(n)}(s,t_2) - u_k^{(n)}(s,t_1)\}Y_k(s)dM_s, \qquad t \in [0,1]$$

which is a square integrable martingale. By the Burkholder-Davis-Gundy inequality (see Dellacherie and Meyer, 1982, p. 287), for $q \ge 1$

$$E(\sup_{t\in[0,1]} |N_t|)^q \le C\ E[N]_1^{q/2}. \qquad (4.20)$$

Also, by Dellacherie and Meyer (1982, Theorem VIII.30) and the assumption that M is a continuous martingale,

$$[N]_t = \int_0^t\{u_k^{(n)}(s,t_2) - u_k^{(n)}(s,t_1)\}^2 Y_k^2(s)d\langle M\rangle_s, \qquad (4.21)$$

so that combining (4.19), (4.20) and (4.21) with $q = 4$

$$E\,|\,V_n(t_2) - V_n(t_1)\,|^4 \le C\sum_{k=1}^{p} E\{\int_0^1[u_k^{(n)}(s,t_2) - u_k^{(n)}(s,t_1)]^2 Y_k^2(s)d\langle M\rangle_s\}^2$$

$$\le C \sum_{k=1}^{p} E\{\int_0^1 [u_k^{(n)}(s,t_2) - u_k^{(n)}(s,t_1)]^2 ds\}^2$$

(by Lemma 4.3 of McKeague (1986) and (B2))

$$= C \sum_{k=1}^{p} \{ \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)]^2\}^2$$

$$\le C\{ \sum_{k=1}^{p} \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)]^2\}^2$$

$$= C \, \| \mathrm{vec}\, [\lambda^{(n)}(t_2) - \lambda^{(n)}(t_1)] \, \|^4$$

$$\le C \, \| \mathrm{vec}\, [h^{(n)}(t_2) - h^{(n)}(t_1)] \, \|^4 \qquad \text{(by Lemma 4.2(i))}$$

$$\le C \left\{ \int_0^1 \frac{(1_{[0,t_2]}(s) - 1_{[0,t_1]}(s))^2}{EY_j^2(s)} ds \right\}^2 \qquad \text{(by Bessel's inequality)}$$

$$\le C \, | t_2 - t_1 |^2. \qquad \text{(by (A3))}$$

Thus (4.18) is satisfied for $q = 4$ and $\gamma = 2$. □

The proof of Theorem 3.1 depends on a result from approximation theory known as Jackson's inequality. The following L^2-version of this result (see Cernyh, 1969) is suitable for our purposes.

Lemma 4.4. Let $f: \mathbb{R} \to \mathbb{R}$ have period 1 and denote the partial sums of the Fourier series of f by $f^{\{d\}}(t) \equiv \sum_{r=1}^{d} f_r \phi_r(t)$, $d \geq 1$, where ϕ_r is defined by (2.7) and $f_r = \int_0^1 f(t)\phi_r(t)dt$. There exist universal constants C_1, C_2 such that

(i) for all $d \geq 1$

$$\int_0^1 [f(t) - f^{\{d\}}(t)]^2 dt \leq C_1 \sup_{|t-s| < \frac{C_2}{d}} |f(t) - f(s)|^2;$$

(ii) if f is differentiable then for all $d \geq 1$

$$\int_0^1 [f(t) - f^{\{d\}}(t)]^2 \leq \frac{C_1}{d^2} \sup_{|t-s| < \frac{C_2}{d}} |f'(t) - f'(s)|^2.$$

Proof of Theorem 3.1. Introduce

$$\bar{\alpha}_j^{(n)}(t) = \frac{1}{b_n} \int_0^1 K(\frac{t-s}{b_n})\alpha_j(s)ds$$

and note that

$$\bar{\alpha}_j^{(n)}(t) - \alpha_j(t) = \int_{(t-1)/b_n}^{t/b_n} K(u)[\alpha_j(t-b_n u) - \alpha_j(t)]du.$$

Since α_j is Lipschitz, this implies that

$$\begin{aligned} \sup_{t\in[0,1]} |\bar{\alpha}_j^{(n)}(t) - \alpha_j(t)| &= O(b_n) \\ &= O(n^{-\frac{1}{2}(1-\beta)}), \qquad (4.22) \end{aligned}$$

because $\beta > \frac{1}{2}(1-\beta)$ for $\beta > 1/3$. Define

$$\alpha_j^{*(n)}(t) = \frac{1}{b_n} \int_0^1 K(\frac{t-s}{b_n}) \alpha_j^{\{n\}}(s) ds ,$$

which we shall use to approximate $\bar{\alpha}_j^{(n)}$. By (F1), (F3) and Lemma 4.4 (ii), for $\delta > 1/4$,

$$\int_0^1 [\alpha_j^{\{n\}}(s) - \alpha_j(s)]^2 \, ds = o(\frac{1}{n})$$

so that, using the Cauchy-Schwarz inequality,

$$|\alpha_j^{*(n)}(t) - \bar{\alpha}_j^{(n)}(t)| \doteq |\frac{1}{b_n} \int_0^1 K(\frac{t-s}{b_n}) [\alpha_j^{\{n\}}(s) - \alpha_j(s)] ds|$$

$$\leq \frac{1}{b_n} [\int_0^1 K^2(\frac{t-s}{b_n}) ds]^{\frac{1}{2}} [\int_0^1 [\alpha_j^{\{n\}}(s) - \alpha_j(s)]^2 ds]^{\frac{1}{2}}$$

$$= O(\frac{1}{\sqrt{b_n}}) \; o(\frac{1}{\sqrt{n}})$$

$$= o(n^{-\frac{1}{2}(1-\beta)}) . \tag{4.23}$$

In view of (4.22) and (4.23) , to complete the proof of the theorem it suffices to show that $\{\sqrt{nb_n} \, (\tilde{\alpha}_j^{(n)}(t) - \alpha_j^{*(n)}(t)), \; n \geq 1\}$ is tight for all $t \in (0,1)$. Fix $t \in (0,1)$. Using Lemma 4.3 we can rework the proof of Theorem 2.1 replacing the basis $(\Psi_{kr}, \; r \geq 1)$ by $(\phi_{kr}, \; r \geq 1)$ and h_k by

$$h_k^{(n)}(s) = \begin{cases} \frac{1}{\sqrt{b_n}} K(\frac{t-s}{b_n}) & \text{if } k=j \\ 0 & \text{if } k \neq j \end{cases}$$

for $k=1,\ldots, p$. Provided we can show that

$$\sup_{n\geq 1} \sum_{r=1}^{d_n} h_{jr}^{(n)2} < \infty \tag{4.24}$$

where $h_{kr}^{(n)} = \int_0^1 h_k^{(n)}(s)\phi_k(s)ds$, it follows that

$$\sqrt{nb_n}(\tilde{\alpha}_j^{(n)}(t) - \alpha_j^{*(n)}(t)) = \sqrt{n} \int_0^1 h_j^{(n)}(s)(\hat{\alpha}_j^{(n)}(s) - \alpha_j^{\{n\}}(s))ds$$

$$= U_n + V_n,$$

where $U_n \xrightarrow{p} 0$, $V_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ni}$,

$$Z_{ni} = \sum_{k=1}^{p} \int_0^1 u_k^{(n)}(s) Y_{ik}(s) dM_i(s),$$

$$u_k^{(n)}(s) = \sum_{r=1}^{d_n} \lambda_{kr}^{(n)} \phi_{kr}(s),$$

$$\text{vec } (\lambda^{(n)}) = R^{(n)-1} \text{ vec } (h^{(n)}),$$

and $h^{(n)}$ is the $p \times d_n$ matrix with entries $h_{kr}^{(n)}$. The condition $d_n = o(n^{\frac{1}{2}})$ of Theorem 2.1 is satisfied here since $\delta < \frac{1}{2}$. Next

$$EZ_{n1}^2 \le p \sum_{k=1}^{p} E\left(\int_0^1 u_k^{(n)}(s) Y_k(s) dM_s\right)^2$$

$$= p \sum_{k=1}^{p} E\int_0^1 u_k^{(n)2}(s) Y_k^2(s) d\langle M\rangle_s$$

$$\le C \sum_{k=1}^{p} \int_0^1 u_k^{(n)2}(s) ds \qquad \text{(by (A5))}$$

$$= C \| \operatorname{vec} \lambda^{(n)} \|^2$$

$$\le C \| \operatorname{vec} h^{(n)} \|^2, \qquad \text{(by Lemma 4.3)}$$

where C is a constant independent of n throughout. Given (4.24), this implies $\sup_{n\ge 1} EZ_{n1}^2 < \infty$ and that $\{V_n, n \ge 1\}$ is tight. It remains to check (4.24). First note that the derivative of $h_n^{(n)}$ satisfies the Lipschitz condition

$$|h_j^{(n)\prime}(s_1) - h_j^{(n)\prime}(s_2)|^2 \le \frac{C}{b_n^5} |s_1 - s_2|^2, \qquad (4.25)$$

for all $s_1, s_2 \in [0,1]$. Since K has compact support there exist $\varepsilon > 0$, $n_1 \ge 1$ such that the support of $h_j^{(n)}$ is contained in the interval $(\varepsilon, 1-\varepsilon)$ for all $n \ge n_1$. For $n \ge n_1$ extend $h_j^{(n)}$ to a periodic function on the whole real line. This extension is differentiable and satisfies the Lipschitz condition (4.25) for all $s_1, s_2 \in \mathbb{R}$.

Now applying Lemma 4.4 (ii)

$$0 \le \int_0^1 [h_j^{(n)}(s)]^2 ds - \sum_{r=1}^{d_n} h_{jr}^{(n)2} \le \frac{C_1}{b_n^5} \sup_{|s_1 - s_2| < \frac{C_2}{d_n}} |h_j^{(n)\prime}(s_1) - h_j^{(n)\prime}(s_2)|^2$$

$$= O\left(\frac{1}{b_n^5 d_n^4}\right) = o(1)$$

since $\beta < \frac{4}{5}\delta$. Also

$$\lim_{n\to\infty} \int_0^1 [h_j^{(n)}(s)]^2 ds = \int_{-1}^1 K^2(v)dv$$

so that

$$\lim_{n\to\infty} \sum_{r=1}^{d_n} h_{jr}^{(n)2} = \int_{-1}^1 K^2(v)dv$$

and this proves (4.24). □

<u>Proof of Theorem 3.2</u>. Since K is Lipschitz it is of bounded variation. Denote its total variation by V(K). Then

$$\sup_{t\in[0,1]} |\tilde{\alpha}_j^{(n)}(t) - \bar{\alpha}_j^{(n)}(t)| = \sup_{t\in[0,1]} \left| \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) d(\hat{A}_j^{(n)} - A_j)(s) \right|$$

$$\le \frac{2}{b_n} V(K) \sup_{s\in[0,1]} |\hat{A}_j^{(n)}(s) - A_j(s)|$$

$$= O_p\left(\frac{1}{b_n\sqrt{n}}\right), \qquad (4.26)$$

since $\{\sqrt{n}(\hat{A}_j^{(n)} - A_j),\ n \geq 1\}$ is tight in $C[0,1]$ by Corollary 2.7. Combining (4.26) with (4.22) completes the proof of the theorem. □

Appendix to the proof of Theorem 2.1.

Let V be a Hilbert space and T a bounded linear operator on V. Suppose that T is invertible, i.e. T^{-1} exists and is continuous. We shall need the following Euler-Knopp series representation for T^{-1}. Let $0 < \lambda < \|T\|^{-2}$ and define

$$S^{(k)} = \lambda \sum_{i=0}^{k} (I - \lambda T^*T)^i T^*, \tag{4.27}$$

where I is the identity operator. Then $S^{(k)}$ converges to T^{-1} in the uniform topology. From Groetsch (1977, p. 69) an error bound is given by

$$\| S^{(k)} - T^{-1} \| \leq \| T^{-1} \| \beta^{k+1}, \tag{4.28}$$

where $0 < \beta < 1$ and

$$\beta = 1 - \lambda \min \{ \|T\|^2, \ \|T^{-1}\|^{-2} \}. \tag{4.29}$$

In the following result we shall apply the Euler-Knopp representation to obtain an approximation to T^{-1} in terms of projections of T onto finite dimensional subspaces. Let $\{V_n, n \geq 1\}$ be an increasing sequence of finite dimensional subspaces of V such that $\bigcup_{n\geq 1} V_n$ is dense in V. The projection of V onto V_n is denoted P_n. Define the operator $T_n : V_n \to V_n$ by $T_n = P_n T P_n^*$.

<u>Lemma 4.5</u>. Suppose that T_n is invertible for all $n \geq 1$ (T is not assumed to be invertible here). Then the following statements are equivalent.

(i) T is invertible and $P_n^* T_n^{-1} P_n \to T^{-1}$ in the strong operator topology;

(ii) $\sup_{n\geq 1} \| T_n^{-1} \| < \infty$.

<u>Proof</u>. (i) $\Rightarrow$(ii) by the principle of uniform boundedness. Conversely, suppose that $M \equiv \sup_{n\geq 1} \| T_n^{-1} \| < \infty$. Put $I_n = P_n^* P_n$ and note that I_n converges strongly to I. By the definition of T_n, $\| I_n T I_n x \| = \| P_n^* T_n P_n x \| \geq M^{-1} \| I_n x \|$ and letting $n \to \infty$ we obtain $\| Tx \| \geq M^{-1} \| x \|$, for all $x \in V$, so that T is invertible. Let $0 < \lambda < \| T \|^{-2}$. Then we also have $0 < \lambda < \| T_n \|^{-2}$ for all $n \geq 1$, so we may use the same λ in the Euler-Knopp approximations $S^{(k)}$, $S_n^{(k)}$ of T^{-1} and T_n^{-1} respectively. It is easily checked from (4.29) that the $\mathcal{E}$'s corresponding to T and T_n are bounded above by $\gamma = 1 - \lambda M^{-2}$. It follows from (4.28) applied to T and T_n that

$$\| S^{(k)} - T^{-1} \| \leq M\gamma^{k+1} \tag{4.30}$$

and

$$\| S_n^{(k)} - T_n^{-1} \| \leq M\gamma^{k+1}$$

for all $n \geq 1$, $k \geq 1$. In particular

$$\| P_n^* S_n^{(k)} P_n - P_n^* T_n^{-1} P_n \| \leq M\gamma^{k+1}, \qquad (4.31)$$

for all $n \geq 1$, $k \geq 1$. But, for k fixed, an induction argument shows that $P_n^* S_n^{(k)} P_n$ converges strongly to $S^{(k)}$ as $n \to \infty$. The triangle inequality, (4.30) and (4.31) then show that $P_n^* T_n^{-1} P_n$ converges strongly to T^{-1}. □

Lemma 4.6. Suppose that conditions (A2)-(A4) are satisfied. Then the matrix $K(t)$ is invertible a.e. (dt) and $u_j^{(n)}$ converges to H_j in $L^2[0,1]$ as $n \to \infty$. Here $K(t)$, $u_j^{(n)}$ and H_j are defined by (2.5), (4.11) and (2.4) respectively.

Proof. Let $V = \bigoplus_{j=1}^{p} L^2([0,1], d\mu_j)$ and $V_n = \bigoplus_{j=1}^{p} \mathrm{span}(\Psi_{jr}, r=1, \ldots, d_n)$. Define $T: V \to V$ by $T(u)(t) = K(t)u(t)$, $u \in V$, $t \in [0,1]$. The matrix representation of $T_n = P_n T P_n^*$ with respect to the basis $\{(\Psi_{1r}, 0,\ldots,0),\ldots,(0,\ldots,0,\Psi_{pr}), r=1,\ldots,d_n\}$ is precisely $\zeta^{(n)}$. It follows that $u^{(n)} = P_n^* T_n^{-1} P_n h$, where $u^{(n)} = (u_1^{(n)},\ldots,u_p^{(n)})$. By Lemma 4.2(i), T_n is invertible for all $n \geq 1$ and $\sup_{n\geq 1} \| T_n^{-1} \| < \infty$. Thus, in view of Lemma 4.5, T is invertible and $u^{(n)} \to T^{-1}h$ in V as $n \to \infty$.

Thus, since the norms in $L^2([0,1], d\mu_j)$ and $L^2([0,1], dt)$ are equivalent under (A2) and (A3), $K(t)$ is nonsingular a.e. (dt), $T^{-1}h = H$ given by (2.4) and $u_j^{(n)} \to H_j$ in $L^2[0,1]$. □

REFERENCES

Aalen, O. O. (1980). "A model for nonparmetric regression analysis of counting processes." Lecture Notes in Statistics 2, 1-25. Springer-Verlag, New York.

Beauchamp, K. G. (1975). Walsh functions and their applications. Academic Press, New York.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Cernyh, N. I. (1969). "On Jackson's inequality in L_2." Proc. Steklov Inst. Math. 88, 75-78.

Chow, W. W. and Teicher, H. (1978). Probability Theory: Independence, Interchangeability, Martingales. Springer-Verlag, New York.

Cox, D. R. (1972). "Regression models with life tables (with discussion)". J. Roy. Statist. Soc. B34, 187-220

Dellacherie, C. and Meyer, P.-A. (1982). Probabilities and Potential B. North-Holland, Amsterdam.

Fine, N. J. (1949). "On the Walsh functions." Trans. Amer. Math. Soc. 65, 372-414.

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.

Groetsch, C. W. (1977). Generalized Inverses of Linear Operators. Marcel Dekker, New York.

Hall, W. J. and Wellner, J. A. (1980). "Confidence bands for a survival curve with censored data." Biometrika 67, 133-143.

Ibragimov, I. A. and Khasminskii, R. Z. (1981). "Estimates of the signal, its derivatives and point of maximum for Gaussian distributions." Theory Prob. Appl. 25, 703-719.

McKeague, I. W. (1986). "Estimation for a semimartingale regression model using the method of sieves." Ann. Statist. 14, No. 2.

Nguyen, H. T. and Pham, T. D. (1982). "Identification of nonstationary diffusion model by the method of sieves." Siam J. on Control and Optimization 20, 603-611.

Paley, R. E. A. C. (1932). "A remarkable series of orthogonal functions." Proc. London Math. Soc. 34, 241-279.

Ramlau-Hansen, H. (1983). "Smoothing counting process intensities by means of kernel functions." Ann. Statist. 11, 453-466.

Rao, R. R. (1963). "The law of large numbers for D[0,1]-valued random variables." Theor. Probab. Appl. 8, 70-74.

END

DTIC

1-86