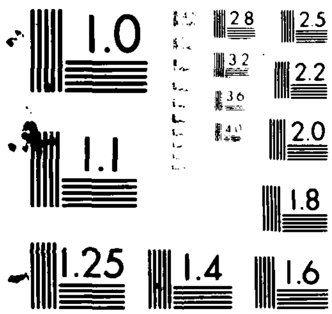
AD-A109 546 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
ASYMPTOTICALLY OPTIMAL DESIGNS FOR SOME TIME SERIES MODELS.(U)
NOV 81 R L EUBANK, P L SMITH, P W SMITH N00014-75-C-0439
UNCLASSIFIED TR-151 NL

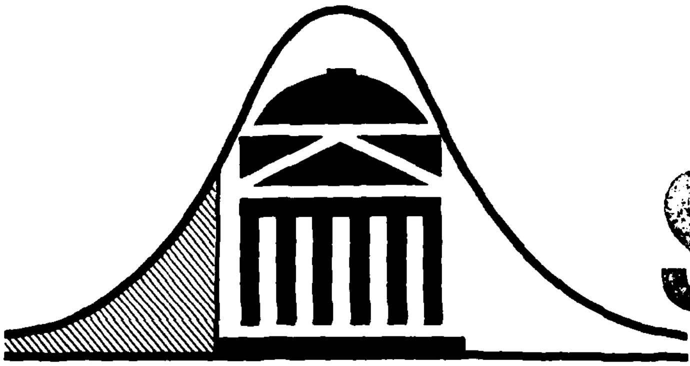
END
DATE
FILMED
2 82
DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

SOUTHERN METHODIST UNIVERSITY



DEPARTMENT OF STATISTICS

DALLAS, TEXAS 75275

ASYMPTOTICALLY OPTIMAL DESIGNS FOR SOME TIME SERIES MODELS

by

R. L. Eubank, P.L. Smith and P.W. Smith

Technical Report No. 151
Department of Statistics ONR Contract

November, 1981

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

Asymptotically Optimal Designs for Some
Time Series Models

By Randall L. Eubank, Patricia L. Smith and Philip W. Smith
Southern Methodist University and Old Dominion University

Short title: Asymptotically Optimal Designs

Summary. The limiting behaviour (as the sample size increases) of the BLUE of the regression coefficients is investigated in the case that derivative information is not available. Using the results of Barrow and Smith (1979) on the asymptotic properties of optimal quadrature formulas several results obtained by Eubank, Smith and Smith (1981) are extended to a multiparameter setting and a wider class of processes which includes multiple integrals of Brownian motion. The asymptotic behaviour of the variance of the BLUE is characterized in terms of the density defining the designs and densities which generate asymptotically optimal design sequences are provided for several optimality criteria.

1. INTRODUCTION. Consider the linear regression model in which a stochastic process, Y, is observed having the form

$$Y(t) = \sum_{j=1}^{J} \beta_j f_j(t) + X(t), \quad t \in [0,1] , \tag{1.1}$$

where $\beta=(\beta_1,\ldots,\beta_J)'$ is a vector of unknown parameters, the f_j are known regression functions and $X(\cdot)$ is a zero mean process with known covariance kernel R. The X process is assumed to admit $k-1$ quadratic mean derivatives at each point $t \in [0,1]$.

If the Y process is sampled over all of [0,1], the linear estimation of β may be accomplished through the use of the reproducing kernel Hilbert space (RKHS) techniques developed by Parzen (1961a, 1961b). We denote this estimator by $\hat{\beta}$ and its corresponding variance-covariance matrix by A^{-1}. When observations are taken at only a finite number of distinct design points on [0,1] the best linear unbiased estimator (BLUE) of β can be constructed through the use of generalized least squares. Various aspects of the problem of optimal design selection for the BLUE have been addressed by Sacks and Ylvisaker (1966, 1968, 1970), Wahba (1971, 1974), and Eubank, Smith, and Smith (1981).

Denote the set of possible $n+1$ point designs for model (1.1) by

(1.2) $$D_n := \{(t_o, t_1, \ldots, t_n) \mid 0 = t_o < t_1 < \ldots < t_n = 1\},$$

where ":=" means "is defined as". Also for $T \varepsilon D_n$ let $\hat{\beta}_T$ represent the BLUE of β based on the observation set $Y_T = \{Y(t) \mid t \varepsilon T\}$ with corresponding variance-covariance matrix denoted by A_T^{-1}. When $k = 1$, i.e., the Y process does not admit derivatives, Sacks and Ylvisaker (1968) considered the problem of selecting a $T^* \varepsilon D_n$ so that

(1.3) $$\psi(A_{T^*}^{-1}) = \inf_{T \varepsilon D_n} \psi(A_T^{-1})$$

or, alternately

(1.4) $$\psi(A_{T^*}) = \sup_{T \varepsilon D_n} \psi(A_T)$$

where ψ is some criterion function which measures the size of A_T^{-1} (e.g., the trace or determinant function). A design which satisfies (1.3) is termed ψ1-optimum whereas a design satisfying (1.4) is termed ψ2-optimum.

As optimal designs are difficult to construct, Sacks and Ylvisaker (1968) instead developed approximate solutions to the optimal design problems (1.3) and (1.4) which entailed the use of design sequences. The method they used for constructing such sequences was to choose the elements of the nth design to be the n-tiles of a continuous density function, h, with support on [0,1]. A design sequence constructed in this manner is called the <u>regular sequence generated by h</u>. This relationship is abbreviated $\{T_n\}$ is RS(h).

A design sequence, $\{T_n^*\}$, is said to be asymptotically ψ1-optimum if

$$\lim_{n\to\infty} [\inf_{T\varepsilon D_n} \psi(A_T^{-1}) - \psi(A^{-1})][\psi(A_{T_n^*}^{-1}) - \psi(A^{-1})]^{-1} = 1 \tag{1.5}$$

and is asymptotically ψ2-optimum if

$$\lim_{n\to\infty} [\psi(A) - \sup_{T\varepsilon D_n} \psi(A_T)][\psi(A) - \psi(A_{T_n^*})]^{-1} = 1. \tag{1.6}$$

Sacks and Ylvisaker (1968) derived densities having the property that the corresponding regular sequence is asymptotically ψ1 or ψ2-optimum. Then if h^* is such a density and $\{T_n^*\}$ is RS (h^*), their approximate solution to the optimal design problem is T_n^*.

For $k > 2$ it has not been possible to characterize the asymptotic variance behavior or obtain asymptotically optimal design sequences for $\hat{\beta}_T$. Instead, it has been necessary to use the less natural observation set $Y_{k,T} = \{Y^{(j)}(t) \mid j=0,\ldots,k-1,\ t\varepsilon T\}$ and the corresponding best linear unbiased estimator $\hat{\beta}_{k,T}$. Hence, for $k > 2$, most of the available literature pertains to $\hat{\beta}_{k,T}$ rather than $\hat{\beta}_T$ (c.f. Sacks and Ylvisaker (1970) and Wahba (1971, 1974)). This is unfortunate as derivative information on the Y process will frequently not be available.

In this paper we examine the asymptotic behaviour of the variance of $\hat{\beta}_T$ where k is allowed to be any fixed finite positive integer. Using results pertaining to the approximation of functions by splines subject to boundary conditions it is possible to extend the asymptotic results of Eubank, Smith and Smith (1981) to a wider class of processes and multiparameter situations. The class of processes considered includes the case when the error process is a multiple integral of Brownian motion whose importance in terms of (1.1) is well known (c.f. Sacks and Ylvisaker (1970)).

Our results stem from certain asymptotic properties of optimal quadrature formula derived by Barrow and Smith (1979). The implications of their work to the approximation of functions by splines under boundary constraints are explored in Section 3. These results are related to the optimal design problem in Section 4. As an illustration of the relationship between these two problems consider the instance when $X(\cdot)$ is a (k-1)-fold multiple integral of Brownian motion. In this case optimal designs for the one parameter case (J = 1) are obtained by minimizing

$$\int_0^1 \left(f^{(k)}(s) - \sum_{i=0}^{n} a_i \frac{\partial^k R(s,t_i)}{\partial s^k} \right)^2 ds$$

with respect to $(t_0,\ldots,t_n)$ and $(a_0,\ldots,a_n)$. It will be seen that $r_t^{(k)}(s) = \frac{\partial^k R(s,t)}{\partial s^k}$ is a spline of order k with a knot at t and that both $f^{(k)}$ and $r_t^{(k)}$ will satisfy

$$f^{(k+j)}(1) = r_t^{(k+j)}(1) = 0$$

for $j = 0,\ldots,k-1$ (except for $t = 1$ when $j = k-1$). Consequently, for this type of process, the optimal design problem is equivalent to a variable knot spline approximation problem where both the function and splines must satisfy boundary conditions.

We present asymptotic results in Section 2 for the special case of J = 1 with proofs in Section 4. In Section 5 the case of J > 1 is examined. A theorem is given which allows for the proof of results analogous to those of Sacks and Ylvisaker (1970) for the multiparameter setting except without the need for derivative information.

2. The case of one parameter. For the purpose of this section we will restrict our attention to the case of J = 1. Consequently, model (1.1) can now be written as

$$Y(t) = \beta f(t) + X(t) \ , \ t \in [0,1]. \tag{2.1}$$

For model (2.1) an optimal n+1 point design is a $T^* \in D_n$ which satisfies

$$\mathrm{Var}(\hat{\beta}_{T^*}) = \inf_{T \in D_n} \mathrm{Var}(\hat{\beta}_T) \tag{2.2}$$

and an asymptotically optimal design sequence, $\{T_n^*\}$, must satisfy

$$\lim_{n \to \infty} \frac{\mathrm{Var}(\hat{\beta}_{T_n^*}) - \mathrm{Var}(\hat{\beta})}{\inf_{T \in D_n} \mathrm{Var}\ (\hat{\beta}_T) - \mathrm{Var}(\hat{\beta})} = 1 \quad . \tag{2.3}$$

The optimal design problem can now be formulated as a nonlinear minimum norm approximation problem for f. Let H(R) denote the RKHS generated by the covariance kernel R which is isometrically isomorphic to the Hilbert space spanned by the Y process (c.f. Parzen (1961a, 1961b)) and let $||\cdot||_R$ denote the norm in H(R). It has been shown by Sacks and Ylvisaker (1966) that

$$\mathrm{Var}(\hat{\beta}_T) = ||R_T f||_R^{-2} \tag{2.4}$$

where $\mathcal{R}_T$ denotes the H(R) orthogonal projector for the subspace $R_T = \text{span}\{R(\cdot,t) \mid t \in T\}$. As $||\mathcal{R}_T f||_R^2 = ||f||_R^2 - ||f - \mathcal{R}_T f||_R^2$ it follows that T* is an optimal design if and only if

$$(2.5) \qquad ||f - \mathcal{R}_{T^*} f||_R = \inf_{T \in D_n} ||f - \mathcal{R}_T f||_R .$$

Consequently, the optimal design problem is equivalent to the problem of finding (when it exists) the best H(R) approximation to f from the nonlinear manifold $R_n = \bigcup_{T \in D_n} R_T$. Although this latter problem is, in general, quite difficult to solve, in certain instances, H(R) and R_n consist respectively of functions and splines which satisfy certain boundary conditions. In this event the problem becomes amenable to analysis as will be shown in Section 4.

Now and in further discussions we will consider a specific class of Y processes. Define the covariance kernel

$$(2.6) \qquad K(s,t) := \int_0^1 \frac{(t-u)_+^{k-1} (s-u)_+^{k-1}}{(k-1)!^2} \, du$$

where $x_+^k = x^k$ if $x \geq 0$ and is zero otherwise. Denote by W(t) the zero mean, normal process corresponding to K, i.e., the (k-1)-st multiple integral of Brownian motion, and let

$$(2.7) \qquad Z(t) := \begin{cases} W(t) - E[W(t) \mid W^{(i)}(1),\ i = k-q,\dots,k-1], & 1 \leq q \leq k \\ W(t) & q = 0 . \end{cases}$$

where q is some fixed but arbitrary integer between 0 and k. We now take R to be

$$(2.8) \qquad R(s,t) := \text{Cov}(Z(s), Z(t)).$$

More specifically, let

$$K^{(i,j)}(s,t) := \frac{\partial^{i+j} K(s,t)}{\partial s^i \partial t^j} .$$

Then, for $q \geq 1$

(2.9) $\quad R(s,t) = K(s,t) - \nu(s)' B^{-1} \nu(t)$

where $\nu(s) = (K^{(0,k-q)}(s,1), \ldots, K^{(0,k-1)}(s,1))'$ and the ij^{th} element of the matrix B is $K^{(i,j)}(1,1)$, i, j = k-q,..., k-1. One consequence of this choice for R is that the class of processes to be considered includes those which have covariance structures like that of (k-1)-fold integrated Brownian motion (q = 0) and (k-1)-fold integrated Brownian bridge (q = 1) processes.

We now state two theorems regarding the behaviour of $\hat{\beta}_T$ for which the proofs will be presented in Section 4.

<u>Theorem 2.1</u> Let $f \epsilon H(R) \cap C^{2k}[0,1]$. If h is a continuous density on [0,1] and $\{T_n\}$ is RS(h) then

(2.10) $$\lim_{n\to\infty} n^{2k} ||f - R_T f||_R^2 = C_k^2 \left[\int_0^1 \frac{(f^{(2k)}(x))^2}{h^{2k}(x)} dx \right]$$

where $C_k^2 = |B_{2k}|/2k!$ and B_{2k} is the 2k-th Bernoulli number.

In view of (2.4) and (2.5) it is clear that equation (2.10) characterizes the asymptotic behaviour of the variance of $\hat{\beta}_T$ in terms of the density defining the designs. The next theorem concerns the limiting behaviour of $\inf_{T \epsilon D_n} \text{Var}(\hat{\beta}_T)$ and provides a density which generates an asymptotically optimal design sequence.

Theorem 2.1. Let $f \in C^{2k}[0,1] \cap H(R)$ and define

$$(2.11)\qquad h(x) := \frac{|f^{(2k)}(x)|^{2/2k+1}}{\int_0^1 |f^{(2k)}(t)|^{2/2k+1} dt} .$$

Then, if $\{T_n\}$ is RS(h)

$$(2.12)\quad \lim_{n\to\infty} n^{2k}\{\mathrm{Var}(\hat\beta_{T_n}) - \mathrm{Var}(\hat\beta)\} = \lim_{n\to\infty} n^{2k}\{\inf_{T\in D_n} \mathrm{Var}(\hat\beta_T) - \mathrm{Var}(\hat\beta)\} =$$

$$C_k^{\,2}\left(\int_0^1 \left(f^{(2k)}(x)\right)^{2/2k+1} dx\right)^{2k+1} \cdot \mathrm{Var}^2(\hat\beta) .$$

It is important to note that when $k > 2$ the bound $\inf_{T\in D_n} \mathrm{Var}(\hat\beta_T)$ may not be obtainable without the use of derivative information. This point will arise as a result of the proof of Theorem 2.2 and has been discussed by Sacks and Ylvisaker (1966). However, by sampling according to the density (2.11) an approximate solution, T_n, to the optimal design problem can be obtained which, for large n, will behave like an optimal design with regards to the corresponding variance of the BLUE. It is also of interest to note that the limit in (2.12) was not previously known to exist although certain results given by Sacks and Ylvisaker (1970) had the consequence of bounding $\lim_{n\to\infty} n^{2k}\{\mathrm{Var}(\hat\beta_{T_n}) - \mathrm{Var}(\hat\beta)\}$ between two numbers. The case of $q = k$ was considered in Eubank, Smith and Smith (1978).

3. Variable knot spline approximation with boundary conditions. In this section we develop certain mathematical preliminaries which will be used in latter sections. For $T\in D_n$ let S_T^k denote the linear manifold of piecewise polynomials of order k (degree < k) on [0,1] in C^{k-2} with breakpoints at $t_1,\ldots,t_{n-1}$. S_T^k is usually called the set of splines of order k with knots at T. Also define the nonlinear manifold of all splines with n - 1

distinct knots by $S_n^k = \bigcup_{T\epsilon D_n} S_T^k$. Since the results which follow involve approximation in the $L^2[0,1]$ norm we adopt the notation

$$||f||: = (\int_0^1 f^2(t)dt)^{1/2} .$$

Given $f \epsilon L^2[0,1]$ and a fixed set of knots, T, the best $L^2[0,1]$ approximation to f by the corresponding splines of order k is the projection of f onto S_T^k, denoted $S_T^k f$. If the knots are allowed to vary the best approximation problem becomes nonlinear as one is attempting to find the best approximation to f from S_n^k . Thus, in this latter case, one is attempting to find $s_n^* \epsilon S_n^k$ so that

$$(3.1) \qquad ||f-s_n^*|| = \inf_s \{||f-s|| \;|s \epsilon S_n^k\} .$$

Finding s_n^* in (3.1) is a variable knot spline approximation problem and is, in general, quite difficult to solve. However, when $f \epsilon C^k[0,1]$, Barrow and Smith (1978) have been able to describe the asymptotic behaviour of $||f - s_n^*||$ and suggest a scheme for knot selection which is asymptotically optimal.

In the next section results will be needed regarding variable knot spline approximation when both f and s_n^* are required to satisfy certain boundary conditions. For $0 \le q < k$ let

$$C^{k,q}[0,1] = \{f\epsilon C^k[0,1]|f^j(1) = 0, j = q,\dots,k-1\}$$

and given $T\epsilon D_n$ define a corresponding spline space

$$S_T^{k,q} = \{s\epsilon S_T^k|s^j(1) = 0, j = q,\dots,k-1\} .$$

When q=k we make the identification $C^{k,k}[0,1] = C^k[0,1]$ and $S_T^{k,k} = S_T^k$. A variable knot spline approximation problem in this setting would entail finding a best approximation to f from $\bigcup_{T\epsilon D_n} S_T^{k,q}$. As it is not clear that such a best approximation exists we circumvent this difficulty by instead considering the best approximation to f from the closure of this set $S_n^{k,q} = \overline{\bigcup_{T\in D_n} S_T^{k,q}}$. So the objective, in this case, is to find $s_{q,n}^* \in S_n^{k,q}$ so that

$$(3.2) \qquad ||f-s_{q,n}^*|| = \inf_s\{||f-s|| \mid s \in S_n^{k,q}\} .$$

The spline $s_{q,n}^*$ may have knot multiplicities and, consequently, may not have the maximum number of continuity constraints.

We now present certain results regarding the asymptotic behaviour of $s_{q,n}^*$ and $S_T^{k,q}f$, where $S_T^{k,q}$ is the orthogonal projector for $S_T^{k,q}$. The proofs of the following theorems are a consequence of results pertaining to the asymptotic properties of optimal quadrature formulas given by Barrow and Smith (1979) and are therefore omitted.

<u>Theorem 3.1.</u> Let $f \in C^{k,q}[0,1]$. Then, if $\{T_n\}$ is RS(h)

$$(3.3) \qquad \lim_{n\to\infty} n^k||f-S_{T_n}^{k,q}f|| = C_k\{\int_0^1 \frac{(f^{(k)}(x))^2}{h^{2k}(x)}dx\}^{1/2}$$

where $C_k = \left(|B_{2k}|/2k!\right)^{1/2}$ and B_{2k} is the 2k-th Bernoulli number.

Theorem 3.1 characterizes the asymptotic behaviour of the error in approximating f from $S_T^{k,q}$ in terms of the density defining the knot sequence $\{T_n\}$. The next theorem is concerned with the free knot case and provides an optimal density for knot selection.

<u>Theorem 3.2.</u> Let $f \in C^{k,q}[0,1]$ and define the density

$$(3.4) \qquad h(x) = \frac{|f^{(k)}(x)|^{2/2k+1}}{\int_0^1 |f^{(k)}(t)|^{2/2k+1}dt} .$$

Then, if T_n is RS(h)

$$(3.5)\quad \lim_{n\to\infty} n^k ||f-S_{T_n}^{k,q}f|| = \lim_{n\to\infty} n^k \inf_{T\varepsilon D_n} ||f - S_T^{k,q}f|| = C_k \left(\int_0^1 |f^{(k)}(x)|^{2/2k+1} dx\right)^{2k+1/2}.$$

Theorem 3.2 has the implication that the sequence of fixed knot approximations to f, $\{S_T^{k,q}f\}$, generated by h in (3.4) works as well, asymptotically, as the corresponding sequence of best free knot approximations. This suggests the use of $S_{T_n}^{k,q}f$ as an approximate solution to problem (3.2). It is worthwhile to note that both Theorems 3.1 and 3.2 deal exclusively with boundary constraints which arise as a result of the particular design problem being considered and are readily extended to other situations.

4. Proof of theorems. In this section we explore the connection between the results of Section 3 and the optimal design problem discussed in Section 2. In particular, proofs are given for Theorems 2.1 and 2.2.

Fix k and q and let R have the form (2.8). It is readily shown that $(-1)^k R(s,t)$ is the Green's function for the boundary value problem

$$D^{2k}g = h$$

$$(4.1)\qquad g^{(j)}(0) = 0 \qquad j = 0,\dots,k-1$$

$$g^{(j)}(1) = 0 \qquad j = k-q,\dots,k-1,k+q,\dots 2k-1.$$

Thus it follows from this property (or directly from (2.9)) that $r_t(s):=R(s,t)$ when considered as a function of s for fixed t is a spline of order 2k and continuity class C^{2k-2} with a knot at t. Consequently, R_n is, in this case, a nonlinear spline manifold and the optimal design problem (2.5) is in fact a variable knot spline approximation problem.

The RKHS generated by R is seen to consist of functions with k-1 absolutely continuous derivatives and is given by

$$(4.2)\quad H(R) = \begin{cases} \{f \mid f^{(j)}(0)=0\ j=0,\dots,k-1, f^{(j)}(1)=0, j=k-q,\dots k-1, f^{(k)} \varepsilon L^2[0,1]\}, & 1 \le q \le k \\ \{f \mid f^{(j)}(0)=0,\ j=0,\dots,k-1,\ f^{(k)} \varepsilon L^2[0,1]\} & , q = 0. \end{cases}$$

The inner product for f ang g in H(R) is given by

$$(4.3)\quad \langle f,g \rangle = \int_0^1 f^{(k)}(x) g^{(k)}(x) dx \quad .$$

For $T \varepsilon D_n$ define the $L^2[0,1]$ subspace

$$(4.4)\quad R_T^k := \operatorname{span}\{r_t^{(k)} \mid r_t \varepsilon R_T\} .$$

In view of (4.3) we have

$$(4.5)\quad ||f-R_T f||_R = ||f^{(k)} - (R_T f)^{(k)}|| = ||f^{(k)} - \mathcal{R}_T^k f^{(k)}||$$

where $\mathcal{R}_T^k$ is the $L^2[0,1]$ orthogonal projector for R_T^k. Consequently, the optimal design problem is equivalent to finding (when it exists) the best $L^2[0,1]$ approximation to $f^{(k)}$ from the splines in $\bigcup_{T \varepsilon D_n} R_T^k$.

To prove Theorems 2.1 and 2.2 we will show that the optimal design problem is nearly the equivalent of a variable knot spline approximation problem for $S_n^{k,q}$. This might be expected as it is apparent from the boundary conditions (4.1) that $r_t^{(k)} \varepsilon S_T^{k,q}$ for $t \varepsilon T\backslash\{0,1\}$, i.e., for $t \varepsilon \{t_1,\dots,t_{n-1}\}$. Hence, we are in a situation similar to that considered in Section 3. The problem preventing the immediate proof of Theorems 2.1 and 2.2 is the disparity between the dimensions and elements of the sets $S_T^{k,q}$ and R_T^k. It is readily seen that $S_T^{k,q}$ has dimension n + q - 1 and it will be shown subsequently that R_T^k has dimension n with $r_1^{(k)} \not\in S_T^{k,q}$ (for $q < k$). To clarify and resolve these difficulties it will be helpful to further analyze the properties of $r_t^{(k)}$.

Now consider the form of R given in (2.9). For $q > 0$ let

$$\nu(s) = (\nu_o(s), \ldots, \nu_{q-1}(s))'$$

where

$$\nu_i(s) = K^{(0,k-q+i)}(s,1)$$

for $i = 0,\ldots,q-1$. Using (2.6) we have that

$$\frac{\partial^k K(s,t)}{\partial s^k} = \frac{(t-s)_+^{k-1}}{(k-1)!} =: K_t^{(k)}(s) \tag{4.6}$$

and

$$\nu_i^{(k)}(s) = \frac{(1-s)^{q-i-1}}{(q-i-1)!} \quad . \tag{4.7}$$

In view of (4.3), $r_t^{(k)}(s)$ is now recognized as the error function resulting from the $L^2[0,1]$ approximation of $K_t^{(k)}(s)$ from the subspace

$$P_q := \text{span}\{1, (1-s), \ldots, \frac{(1-s)^{q-1}}{(q-1)!}\} \ .$$

This fact has several consequences which are of interest. First it is seen that $R_T^k \perp P_q$ (in L^2 norm). It also follows that, for $q < k$, $r_1^{(k)}(s)$ is a polynomial of degree k-1 which satisfies $r_1^{(2k-1)}(s) = 1$ and, therefore, $r_1^{(k)}$ is not contained in $S_T^{k,q}$. When $q = k$, $K_1^{(k)}(s) \in P_k$ so $r_1^{(k)}(s)$ will vanish identically. Finally, as $K_o^{(k)}(s)$ is zero on $[0,1]$, $r_o^{(k)}$ must vanish as well which implies that, for $q < k$, the dimension of R_T^k is n (the vanishing of $r_o^{(k)}$ entails that observations taken at zero provide no information about β a fact deduced more directly from (2.1) and (4.2)).

For $T \in D_n$ define the "design" T' by deleting the point $t_n = 1$ from T. For designs of this form we have $R^k_{T'} \subset S^{k,q}_T$. Approximations from $R^k_{T'}$ and $S^{k,q}_T$ are related by the following lemma.

Lemma. If $f \in H(R)$ and $T \in D_n$ then $R^k_{T'} f^{(k)}$ is the best $L^2[0,1]$ approximation to $f^{(k)}$ from $S^{k,q}_T$.

Proof. For $0 < q \le k$ the dimension of $R^k_{T'}$ is n-1. Since, for $q > 0$, $R^k_{T'} \perp P_q$ in L^2 norm) and $R^k_{T'} \subset S^{k,q}_T$, we have

$$S^{k,q}_T = P_q \oplus R^k_{T'} .$$

In addition, through use of the boundary conditions in (4.2) and integration by parts, it is seen that if $g \in H(R)$ and $p \in P_q$ then

$$\int_0^1 g^{(k)}(x)p(x)dx = 0 .$$

As $f - R_{T'} f$ is in H(R) it follows that $f^{(k)} - R^k_{T'} f^{(k)}$ is orthogonal to P_q and $R^k_{T'}$. Consequently, $f^{(k)} - R^k_{T'} f^{(k)}$ is orthogonal to $S^{k,q}_T$ for $0 < q \le k$. For $q = 0$ the lemma is verified through noting that in this case $R^k_{T'} = S^{k,0}_T$.

To complete the proof of Theorems 2.1 and 2.2 the difficulty associated with an observation taken at 1 must be resolved. This is accomplished through noting that

$$(4.8) \quad ||f - R_{T'} f||_R = ||f^{(k)} - S^{k,q}_T f^{(k)}|| \ge ||f^{(k)} - R^k_T f^{(k)}|| \ge ||f^{(k)} - S^k_T f^{(k)}|| .$$

Theorems 2.1 and 2.2 now follow, through use of the lemma and equation (4.8), from Theorems 3.1 and 3.2 respectively. It is important to note that in the free knot case our results necessarily pertain to the best H(R) approximation to f from the closed set $\bar{R}_n$ rather than R_n and, hence, derivative information may be required to obtain the bound $\inf_{T \in D_n} \mathrm{Var}(\hat{\beta}_T)$ for any particular n. However, when it is not possible to sample derivatives of the Y process an approximate solution may be obtained through sampling according to the density (2.11).

5. The case of many parameters. We now consider the general case of model (1.1) with $J > 1$. Assume that ψ, the criterion function, is a continuous real-valued function on the non-negative matrices which satisfies $\psi(0) = 0$ and $\psi(B) \geq \psi(C)$ if $B-C$ is a non-negative matrix. Given a particular criterion ψ we wish to construct asymptotically $\psi 1$ or $\psi 2$ optimum design sequences. To do so it will suffice to prove an analog of Sacks' and Ylvisaker's (1968) Theorem 3.2 for covariance kernels of the form (2.8)

Theorem 5.1. Let $f_j \varepsilon C^{2k}[0,1] \cap H(R)$, $j=1,\ldots,J$, and let $a_1,\ldots,a_J$ be a set of positive constants. Then, for any design sequence $\{T_n\}$

$$(5.1)\quad \liminf_{n\to\infty} n^{2k}\sum_{j=1}^{J} a_j ||f_j - R_{T_n} f_j||_R^2 \geq C_k^{\,2}\left(\int_0^1 \left(\sum_{j=1}^{J} a_j\left(f_j^{(2k)}(x)\right)^2\right)^{1/2k+1} dx\right)^{2k+1}$$

$$= \lambda \text{ (say)}.$$

If $\{T_n\}$ is RS(h) where

$$(5.2)\qquad h(x) = \frac{\left(\sum_{j=1}^{J} a_j\left(f_j^{(2k)}(x)\right)^2\right)^{1/2k+1}}{\int_0^1\left(\sum_{j=1}^{J} a_j\left(f_j^{(2k)}(s)\right)^2\right)^{1/2k+1} ds}$$

then

$$(5.3)\qquad \lim_{n\to\infty} n^{2k}\sum_{j=1}^{J} a_j ||f_j - R_{T_n} f_j||_R^2 = \lambda .$$

Proof. First note that since (5.3) is an immediate consequence of Theorem 2.1 it will suffice to prove (5.1). The latter result can be obtained through modification of work by Barrow and Smith (1978). We highlight the differences here and refer the reader to their paper for more details.

Through use of the lemma in Section 4 and the fact that, for $g \in L^2[0,1]$, $||g-S_T^{k,q}g|| \geq ||g-S_T^k g||$ it is seen that (5.1) will be proven when it can be shown that for $g_i \in C^k[0,1]$, $j = 1,\ldots,J$,

$$(5.4)\quad \sum_{j=1}^{J} a_j||g_j-S_{T_n}^k g_j||^2 \geq C_k^2\left(\int_0^1 \left(\sum_{j=1}^{J} a_j (g_j^{(k)}(x))^2\right)^{1/2k+1} dx\right)^{2k+1} .$$

The proof of (5.4) now proceeds in a step-wise manner similar to the proof of Theorem 2 of Barrow and Smith (1978). The inequality (5.4) is first shown to hold when $g_i(t) = c_i t^k/k!$, where c_i is a constant, and then when $g_i \in C^k[0,1]$ with $\sum_{j=1}^{J} a_j g_j^{(k)}(t)^2 \geq \delta > 0$ before finally considering the general case of $g_i \in C^k[0,1]$. The details involved in the vertification of each of these cases may be deduced from Barrow and Smith (1978).

Theorem 5.1 provides the crucial result for obtaining asymptotically optimal design sequences in the multi-parameter case. It is now possible to obtain analogs of the theorems given in Section 4 of Sacks and Ylvisaker (1968) using similar methods of proof. Examples of these results are provided by the following two theorems which deal with the trace (tr) and determinant (det) criteria.

Theorem 5.2 Let $\psi(B) = tr(BM)$ where M is a non-negative J x J matrix and define the density

$$(5.5)\qquad h(x) := \frac{[\phi(x)'M\phi(x)]^{1/2k+1}}{\int_0^1 [\phi(t)'M\phi(t)]^{1/2k+1}\, dt}$$

where

(5.6) $\phi(x) := (f_1^{(2k)}(x), \ldots, f_J^{(2k)}(x))'$.

Then the design sequence $\{T_n\}$ which is RS(h) is asymptotically ψ2-optimum.

Theorem 5.3. Let $\psi(B) = \det(B)$ and define the density

(5.7) $$h(x) := \frac{[\phi'(x)A^{-1}\phi(t)]^{1/2k+1}}{\int_0^1 [\phi(t)A^{-1}\phi(t)]^{1/2k+1}\,dt} .$$

Then the design sequence $\{T_n\}$ which is RS(h) is asymptotically ψ2-optimum.

Randall Eubank
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Patricia L. Smith
Department of Math Sciences
Old Dominion University
Norfolk, Virginia 23508

Philip W. Smith
Department of Math Science
Old Dominion University
Norfolk, Virginia 23508

REFERENCES

Barrow, D. L. and Smith, P. W. (1978). Asymptotic properties of best $L^2[0,1]$ approximation by splines with variable knots. Quarterly of Applied Math. 36, 293-304.

Barrow, D. L. and Smith, P. W. (1979). Asymptotic properties of optimal quadrature formulas. Separatum aus: Numerische Integration, 54-66.

Eubank, R. L., Smith, P.L. and Smith, P.W. (1981). Uniqueness and eventual uniqueness of optimal designs in some time series models. To appear Ann. Statist.

Parzen, Emanuel (1961a). An approach to time series analysis. Ann. Math. Statist. 32, 951-989.

Parzen, Emanuel (1961b). Regression analysis for continuous parameter time series. Proc. 4th Berkeley Sympos. Math., Statist. and Prob., Vol. I, Univ. California Press: Berkeley, Calif., 469-489.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. Ann. Math. Statist. 37, 66-89.

Sacks, J. and Ylvisaker, D. (1968). Designs for regression problems with correlated errors: many parameters. Ann. Math. Statist. 39, 49-69.

Sacks, J. and Ylvisaker, D. (1970). Designs for regression problems with correlated errors, III. Ann. Math. Statist. 41, 2057-2074.

Wahba, G. (1971). On the regression design problem of Sacks and Ylvisaker. Ann. Math. Statist. 42, 1035-1053.

Wahba, G. (1974). Regression design for some equivalence classes of kernels. Ann. Statist. 2, 925-934.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 151	2. GOVT ACCESSION NO.: AD-A109 546
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Asymptotically Optimal Designs for Some Time Series Models	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER: 151
7. AUTHOR(s): Randall L. Eubank, Patricia L. Smith and Philip W. Smith	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Southern Methodist University, Dallas, Texas 75275	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, VA 22217	12. REPORT DATE: November 1981
	13. NUMBER OF PAGES: 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

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

18. SUPPLEMENTARY NOTES

Short Title: Asymptotically Optimal Designs

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

The limiting behaviour (as the sample size increases) of the BLUE of the regression coefficients is investigated in the case that derivative information is not available. Using the results of Barrow and Smith (1979) on the asymptotic properties of optimal quadrature formulas several results obtained by Eubank, Smith and Smith (1981) are extended to a multiparameter setting and a wider class of processes which

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

includes multiple integrals of Brownian motion. The asymptotic behaviour of the variance of the BLUE is characterized in terms of the density defining the designs and densities which generate asymptotically optimal design sequences are provided for several optimality criteria.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

EN

DATE
FILME

2-8

DTI