The Annals of Statistics
1977, Vol. 5, No. 4, 709–721

PREDICTION FUNCTIONS AND MEAN-ESTIMATION FUNCTIONS FOR A TIME SERIES

BY LAWRENCE PEELE[1] AND GEORGE KIMELDORF[2]

Florida State University

Let $T \subseteq I$ be sets of real numbers. Let $\{Y(t): t \in I\}$ be a real time series whose mean is an unknown element of a known class of functions on I and whose covariance kernel k is assumed known. For each fixed $s \in I$, $Y(s)$ is predicted by a minimum mean square error unbiased linear predictor $\hat{Y}(s)$ based on $\{Y(t): t \in T\}$. If $\hat{y}(s)$ is the evaluation of $\hat{Y}(s)$ given a set of observations $\{Y(t) = g(t): t \in T\}$, then the function $\hat{y}$ is called a prediction function. Mean-estimation functions are defined similarly. For certain prediction and estimation problems, characterizations are derived for these functions in terms of the covariance structure of the process. Also, relationships between prediction functions and spline functions are obtained that extend earlier results of Kimeldorf and Wahba (*Sankhyā Ser. A* **32** 173–180).

1. Introduction. Let I denote a set of real numbers and let $\{Y(t): t \in I\}$ be a real time series of the form

$$Y(t) = m_0(t) + X(t)\,, \tag{1.1}$$

where $\{X(t)\}$ has mean 0 and known positive definite covariance kernel $k(s, t) = E[X(s)X(t)]$. The mean function m_0 for $\{Y(t)\}$ is an unknown member of a known class M of functions on I. Given any subset T of I, let $L[Y(t): t \in T]$ denote the vector space of finite linear combinations of elements of $\{Y(t): t \in T\}$ with inner product given by $\langle U, V\rangle = \text{Cov}(U, V)$. Denote the completion of this inner-product space by $L^2[Y(t): t \in T]$, so that $L^2[Y(t): t \in T]$ is the Hilbert space generated by $\{Y(t): t \in T\}$ with inner product determined by $\langle Y(s), Y(t)\rangle = k(s, t)$. For each $s \in I$ let $Y(s)$ be predicted according to some optimality criterion by an element $\hat{Y}(s) \in L^2[Y(t): t \in T]$. Let g be any particular real-valued function and suppose for each $t \in T$ it is observed that $Y(t) = g(t)$. We call $\hat{y}(s)$ the *evaluation* of $\hat{Y}(s)$ given $\{Y(t) = g(t): t \in T\}$ if $\hat{Y}(s) = \sum c_i(s)Y(t_i)$ implies $\hat{y}(s) = \sum c_i(s)g(t_i)$; the evaluation $\hat{y}(s)$ for infinite T is defined in Section 5. The function $\hat{y}$ whose value at each $s \in I$ is $\hat{y}(s)$ is called a *prediction function*. Similarly, for each $s \in I$, let $m_0(s)$ be estimated according to some optimality criterion by an element $\hat{Z}(s) \in L^2[Y(t): t \in T]$. If for each $s \in I$, $\hat{z}(s)$ is the evaluation of $\hat{Z}(s)$

Received February 1975; revised December 1976.

[1] Now at The University of Florida. Research supported by the United States Army Research Office, Durham, under Grant No. DAHC04-74-G-0188.

[2] Now at the University of Texas at Dallas. Research supported by the Air Force Office of Scientific Research under Grant AFOSR-74-2581A.

AMS 1970 *subject classifications*. Primary 62M10, 41A15.

Key words and phrases. Prediction functions, mean-estimation functions, spline functions, time series.

given the observations $\{Y(t) = g(t): t \in T\}$, the function $\hat{z}$ is called a *mean-estimation function*.

The goal of this paper is to examine prediction functions and mean-estimation functions for a time series with a known covariance and an unknown mean function. Under various assumptions on M and the set $\{g(t): t \in T\}$ of observations of the random variables $\{Y(t): t \in T\}$, prediction functions and mean-estimation functions are characterized in terms of the covariance structure. Examples are given where prediction functions are certain types of spline functions.

For fixed $s \in I$, a random variable $\hat{Y}(s)$ is called a *minimum mean square error unbiased linear* (MEUL) *predictor* for $Y(s)$ if among random variables W satisfying the conditions

$$\text{(1.2)} \qquad \text{unbiasedness:} \quad E_m(W) = m(s) \quad \text{for all} \quad m \in M;$$

$$\text{(1.3)} \qquad \text{linearity:} \quad W \in L^2[Y(t): t \in T],$$

the minimum of $E[W - Y(s)]^2$ occurs when $W = \hat{Y}(s)$. Similarly, a random variable $\hat{Z}(s)$ is called a *minimum variance unbiased linear* (MVUL) *estimator* for $m_0(s)$ if $\hat{Z}(s)$ is a random variable W which has minimal variance among random variables W satisfying (1.2) and (1.3).

Characterizations for prediction functions and mean-estimation functions will be based on the important concept of a reproducing kernel Hilbert space, developed by Aronszajn (1950) and used extensively with stochastic processes by Parzen (1959), (1961), (1970), and others. A Hilbert space H of functions on a set I of real numbers is called a *reproducing kernel Hilbert space* (RKHS) with reproducing kernel k if k is a function on $I \times I$ such that for each $t \in I$, $k(\cdot, t) \in H$ and $\langle f, k(\cdot, t)\rangle = f(t)$ for all $f \in H$, where $k(\cdot, t)$ denotes the function on I whose value at $s \in I$ is $k(s, t)$. Each nonnegative definite symmetric kernel k on $I \times I$ generates a unique RKHS of functions on I with reproducing kernel k. This RKHS is denoted by $H(k)$. It can be easily seen that the Hilbert space $L^2[k(\cdot, t): t \in I]$ generated by $\{k(\cdot, t): t \in I\}$ with $\langle k(\cdot, s), k(\cdot, t)\rangle = k(s, t)$ is the RKHS $H(k)$. For $T \subseteqq I$, let $L^2[k(\cdot, t): t \in T]$ denote the closed subspace of $H(k)$ generated by $\{k(\cdot, t): t \in T\}$. If k is the covariance kernel for a process $\{Y(t): t \in I\}$, then the isometry (1–1 onto inner-product preserving map) that takes $Y(t)$ to $k(\cdot, t)$ for each t maps $L^2[Y(t): t \in I]$ onto $H(k)$ and $L^2[Y(t): t \in T]$ onto $L^2[k(\cdot, t): t \in T]$.

Section 3 characterizes prediction functions and mean-estimation functions in the case in which T is finite and M is a finite-dimensional vector space of functions. The case in which M is a certain convex set of functions is considered in Section 4. Section 5 extends previous results to the case in which T is an infinite set and M is a vector space of functions, not necessarily finite dimentional.

2. Background material. MEUL prediction functions were first studied by Kimeldorf and Wahba (1970a), (1970b). Their latter paper considered differential operators of the form

$$\text{(2.1)} \qquad \mathscr{L} = \textstyle\sum_{i=0}^{q} a_i \mathscr{D}^i,$$

where $\mathscr{D}$ is the differentiation operator, the functions a_i have q continuous derivatives on (0, 1), and $a_q \equiv 1$. They showed that in certain cases an MEUL prediction function $\hat{y}$ is an $\mathscr{L}$-spline of interpolation to a finite set of points $\{(t_j, \lambda_j): j = 1, 2, \cdots, n\}$. That is, $\hat{y}$ is the unique function $y \in H_b = \{f: \mathscr{D}^{q-1}f$ is absolutely continuous and $\mathscr{L}f \in L^2[0, 1]\}$ which minimizes

$$\int_0^1 [(\mathscr{L}y)(t)]^2 \, dt$$

among functions satisfying $y(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$.

The following lemma of Kimeldorf and Wahba (1970b) is used in the proofs of Theorems 3.1, 3.2, and 4.1 below. A generalization—Lemma 5.1— of this lemma is used in the proofs of Theorem 5.1 and Theorem 5.2. We will use P to denote an orthogonal projection operator in Hilbert space.

LEMMA 2.1. *Let $H = H_1 + H_2$ be the direct orthogonal sum of a finite-dimensional real Hilbert space H_1 and any real Hilbert space H_2. Let P_i be the projection operator onto H_i and J be a finite-dimensional subspace of H such that $P_1(J) = H_1$. Then for any given elements, $\tilde{w}, \tilde{u} \in H$:*

(a) *there exists a unique element $w = \hat{w} \in J$ which minimizes $\|\tilde{w} - w\|^2$ subject to the constraint $P_1(\tilde{w} - w) = 0$;*

(b) *there exists a unique element $u = \hat{u} \in H$ for which $\|P_2 u\|^2$ is minimized among elements u satisfying $\tilde{u} - u \in J^\perp$;*

(c) $\langle \tilde{u}, \hat{w} \rangle = \langle \hat{u}, \tilde{w} \rangle$.

3. A finite-dimensional space of possible mean functions. In this section, $\{Y(t): t \in I\}$ is assumed to have the model (1.1) where M is a finite-dimensional space of functions on I. Let $M_2 = M \cap H(k)$. Let M_1 be a subspace of M complementary to M_2. That is, M_1 and M_2 satisfy

$$M_2 \subseteqq H(k), \tag{3.1}$$

$$M_1 \cap H(k) = \{0\}, \tag{3.2}$$

and

$$M_1 + M_2 = M. \tag{3.3}$$

Let $\{f_i: i = 1, 2, \cdots, q\}$ be a basis for M_1. For $s, t \in I$, let $k_1(s, t) = \sum_{i=1}^q f_i(s) f_i(t)$. With the inner product defined by requiring $\langle f_i, f_j \rangle = \delta_{ij}$, M_1 becomes a Hilbert space; in fact, $M_1 = H(k_1)$, Let $k_0 = k_1 + k$. It can be easily seen from the definition of an RKHS that the RKHS $H(k_0)$ is the orthogonal sum of the RKHS's $H(k_1)$ and $H(k)$. Since M is a finite-dimensional subspace of $H(k_0)$, M is a closed subspace of $H(k_0)$.

Prediction and mean estimation in this section are based on a finite number of observations; that is, T is assumed to consist of finitely many elements $t_1, t_2, \cdots, t_n \in I$. The estimators or predictors considered are those of the form

$$\sum_{j=1}^n c_j Y(t_j) \tag{3.4}$$

(that is, elements of $L[Y(t_j): j = 1, 2, \cdots, n]$). Let $J_0 \doteq L[k_0(\cdot, t_j): j = 1, 2, \cdots, n]$. For $m \in M$, $E_m(Y(s)) = m(s) = \langle m, k_0(\cdot, s)\rangle_{H(k_0)}$, and $E_m(\sum_{j=1}^n c_j Y(t_j)) = \langle m, \sum_{j=1}^n c_j k_0(\cdot, t_j)\rangle_{H(k_0)}$. Hence, $\sum_{j=1}^n c_j Y(t_j)$ is an unbiased predictor for $Y(s)$ if and only if $P_M(\sum_{j=1}^n c_j k_0(\cdot, t_j)) = P_M(k_0(\cdot, s))$. It follows that the condition

$$P_M(J_0) = M \tag{3.5}$$

is equivalent to the condition that, for each $s \in I$, there exists an unbiased linear predictor (that is, a random variable satisfying (1.2) and (1.3)) for $Y(s)$. The following theorem extends a result of Kimeldorf and Wahba (1970b), and the proof is similar to their proof.

THEOREM 3.1. *Let* $\{Y(t): t \in I\}$ *have the model* (1.1) *where* M *is a finite-dimensional space of functions on* I. *Suppose that* (3.5) *is satisfied, and for each* $t \in I$, *let* $\hat{Y}(t) = \sum_{j=1}^n \hat{c}_j(t) Y(t_j)$ *be the* MEUL *predictor for* $Y(t)$. *Then the* MEUL *prediction function* $\hat{y}$, *given particular observations* $\{Y(t_j) = \lambda_j : j = 1, 2, \cdots, n\}$, *is the unique function minimizing*

$$||P_{M^\perp}(y)||^2_{H(k_0)}$$

among functions satisfying

(a) $$y \in H(k_0)$$

and

(b) $$y(t_j) = \lambda_j \qquad \text{for} \quad j = 1, 2, \cdots, n\,.$$

PROOF. Let $H = H(k_0)$, $H_1 = M$, $H_2 = M^\perp$, $J = J_0$, and $\tilde{w} = k_0(\cdot, t)$. Let $\tilde{u}$ be any function satisfying (a) and (b) of the theorem. (Since k is positive definite, the matrix $[k(t_i, t_j)]_{n\times n}$ is nonsingular and there exists such an element $\tilde{u}$.) Now apply Lemma 2.1 as follows. A random variable $\sum_{j=1}^n c_j Y(t_j)$ is an unbiased predictor for $Y(t)$ if and only if $P_M(\sum_{j=1}^n c_j k_0(\cdot, t_j)) = P_M(k_0(\cdot, t))$. Since $H(k_1)$ is a closed subspace of M, it follows that if $P_M(\sum_{j=1}^n c_j k_0(\cdot, t_j)) = P_M(k_0(\cdot, t))$, then $P_{H(k_1)}(\sum_{j=1}^n c_j k_0(\cdot, t_j)) = P_{H(k_1)}(k_0(\cdot, t))$ and $||\sum_{j=1}^n c_j k_0(\cdot, t_j) - k_0(\cdot, t)||^2_{H(k_0)} = ||\sum_{j=1}^n c_j k(\cdot, t_j) - k(\cdot, t)||^2_{H(k)} = E[\sum_{j=1}^n c_j Y(t_j) - Y(t)]^2$. If $\sum_{j=1}^n \hat{c}_j(t) k_0(\cdot, t_j)$ is the unique element $w = \hat{w} \in J$ minimizing $||w - \tilde{w}||^2_{H(k_0)}$ subject to the constraint $P_M(w) = P_M(\tilde{w})$, then $\sum_{j=1}^n \hat{c}_j(t) Y(t_j)$ is the MEUL predictor for $Y(t)$. Since $\tilde{u} - u \in J^\perp$ if and only if $\tilde{u}$ and u agree at $t_1, t_2, \cdots, t_n$, it remains only to show that the prediction function $\hat{y}$ is the function $\hat{u}$ of Lemma 2.1. For fixed $t \in I$,

$$\begin{aligned} \hat{y}(t) &= \textstyle\sum_{j=1}^n \hat{c}_j(t)\lambda_j \\ &= \textstyle\sum_{j=1}^n \hat{c}_j(t)\langle \tilde{u}, k_0(\cdot, t_j)\rangle \\ &= \langle \tilde{u}, \textstyle\sum_{j=1}^n \hat{c}_j(t) k_0(\cdot, t_j)\rangle \\ &= \langle \tilde{u}, \hat{w}\rangle \\ &= \langle \hat{u}, \tilde{w}\rangle \\ &= \langle \hat{u}, k_0(\cdot, t)\rangle = \hat{u}(t)\,. \end{aligned}$$

REMARK 3.0. Note that condition (3.5) is independent of the choice of the space M_1 satisfying (3.2) and (3.3) and of the choice of basis for M_1. To see this, it is sufficient to verify that condition (3.5) is equivalent to the nonexistence of a nonzero element $m \in M$ for which $m(t_j) = 0$ for $j = 1, 2, \cdots, n$. Also, if M^* is the orthogonal complement in $H(k)$ of $M \cap H(k)$ so that each function $y \in M + H(k)$ has a unique representation $y = y_1 + y_2$ where $y_1 \in M$ and $y_2 \in M^*$, then the MEUL prediction function $\hat{y}$ of Theorem 3.1 can also be characterized as the unique function $y = y_1 + y_2 \in M + H(k)$ which minimizes $||y_2||^2_{H(k)}$ subject to the constraints $y(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$. Therefore, the MEUL prediction function $\hat{y}$ is also independent of the choice of M_1 and of the basis for M_1.

REMARK 3.1. Let $A = [f_i(t_j)]_{q \times n}$. If $M_2 = \{0\}$, then condition (3.5) is equivalent to the condition that rank $(A) = q$. If $\{m_1^*, m_2^*, \cdots, m_{q^*}^*\}$ is an orthonormal basis for M_2, then condition (3.5) is equivalent to the condition that

$$\operatorname{rank}\begin{bmatrix} A \\ \cdots \\ B \end{bmatrix} = q + q^*$$

where $B = [m_i^*(t_j)]_{q^* \times n}$.

REMARK 3.2. Theorem 2.1 of Kimeldorf and Wahba (1970b) is the special case of Theorem 3.1 when $M_2 = \{0\}$ and a differential operator $\mathscr{L}$ has M_1 for a null space and maps $H(k)$ isometrically onto $L^2[0, 1]$. The model of Kimeldorf and Wahba (1970b) led to an MEUL prediction function which was an $\mathscr{L}$-spline, If that model were altered by letting the covariance kernel for $\{Y(t): t \in [0, 1]\}$ be k_0 rather than k, then the MEUL prediction function would be the same $\mathscr{L}$-spline. If I is an interval of infinite length and $H(k_0) = \{f: \mathscr{D}^{q-1}f$ is absolutely continuous on compact subintervals of I and $\mathscr{L}f \in L^2(I)\}$, then the differential operator theory of Dunford and Schwartz (1963) can be used to derive a result similar to Theorem 2.1 of Kimeldorf and Wahba.

The next theorem characterizes an MVUL mean-estimation function for the same model as that of Theorem 3.1.

THEOREM 3.2. *Let $\{Y(t): t \in I\}$ have the model* (1.1) *where M is a finite-dimensional space of functions on I. Suppose that* (3.5) *is satisfied, and for each $t \in I$, let $\hat{Z}(t) = \sum_{j=1}^n \hat{d}_j(t) Y(t_j)$ be the* MVUL *estimator for $m_0(t)$. Then the* MVUL *mean-estimation function $\hat{z}$, given particular observations $\{Y(t_j) = \lambda_j: j = 1, 2, \cdots, n\}$, is $P_M(\hat{y})$ where $\hat{y}$ is the* MEUL *prediction function.*

PROOF. Let $t \in I$. As in the proof for Theorem 3.1, $\sum_{j=1}^n d_j Y(t_j)$ is an unbiased linear estimator for $m_0(t)$ if and only if $P_M(\sum_{j=1}^n d_j k_0(\cdot, t_j)) = P_M(k_0(\cdot, t))$. Minimizing Var $(\sum_{j=1}^n \hat{d}_j Y(t_j))$ subject to this constraint is equivalent to minimizing $||\sum_{j=1}^n d_j k_0(\cdot, t_j) - P_M(k_0(\cdot, t))||^2$ subject to the same constraint since

$$\begin{aligned}\operatorname{Var}(\textstyle\sum_{j=1}^n d_j Y(t_j)) &= ||\textstyle\sum_{j=1}^n d_j k(\cdot, t_j)||^2_{H(k)} \\ &= ||\textstyle\sum_{j=1}^n d_j k_0(\cdot, t_j) - P_M(\sum_{j=1}^n d_j k_0(\cdot, t_j))||^2_{H(k_0)} \\ &\quad + ||P_{M_2}(\textstyle\sum_{j=1}^n d_j k_0(\cdot, t_j))||^2_{H(k_0)} \\ &= ||\textstyle\sum_{j=1}^n d_j k_0(\cdot, t_j) - P_M(k_0(\cdot, t))||^2_{H(k_0)} + ||P_{M_2}(k_0(\cdot, t))||^2_{H(k_0)} .\end{aligned}$$

Now apply Lemma 2.1 with $H = H(k_0)$, $H_1 = M$, $H_2 = M^\perp$, $J = J_0$, $\tilde{w} = P_M(k_0(\cdot, t))$, and $\tilde{u}$ any element of $H(k_0)$ such that $\tilde{u}(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$. It follows from the proof of Theorem 3.1 that the element denoted $\hat{u}$ in Lemma 2.1 is the MEUL prediction function $\hat{y}$. For fixed $t \in I$, if $\hat{w} = \sum_{j=1}^n \hat{d}_j(t)k_0(\cdot, t_j)$ then $\sum_{j=1}^n \hat{d}_j(t)Y(t_j)$ is the MVUL estimator for $m_0(t)$. The MVUL mean-estimation function $\hat{z}$ is defined for $t \in I$ by

$$\begin{aligned}
\hat{z}(t) &= \textstyle\sum_{j=1}^n \hat{d}_j(t)\lambda_j \\
&= \textstyle\sum_{j=1}^n \hat{d}_j(t)\langle \tilde{u}, k_0(\cdot, t_j)\rangle \\
&= \langle \tilde{u}, \textstyle\sum_{j=1}^n \hat{d}_j(t)k_0(\cdot, t_j)\rangle \\
&= \langle \tilde{u}, \hat{w}\rangle \\
&= \langle \hat{u}, \tilde{w}\rangle \\
&= \langle \hat{y}, P_M(k_0(\cdot, t))\rangle \\
&= \langle P_M(\hat{y}), k_0(\cdot, t)\rangle \\
&= [P_M(\hat{y})](t) .
\end{aligned}$$

REMARK 3.3. Suppose that the hypotheses of Theorem 3.1 are satisfied. Let $m_i = f_i$ for $i = 1, 2, \cdots, q$, and let $\{m_i : i = q+1, q+2, \cdots, q+q^*\}$ be a basis for M_2. Then m_0 has a unique (but unknown) representation $\sum_{i=1}^{q+q^*} \beta_i m_i$. If $\hat{B}_i$ denotes the MVUL estimator for β_i, then for each $t \in I$, $\sum_{i=1}^{q+q^*} \hat{B}_i m_i(t) = \hat{Z}(t)$, the MVUL estimator for $m_0(t)$. A similar equivalence holds if M_1 is $\{0\}$ and M_2 is a closed (infinite-dimensional) subspace of $H(k)$—see Theorem 5.3 and Remark 5.4.

4. A minimax mean-estimation result. The main result of this section differs from the results of the previous section in that the set M is no longer a linear space and the method of estimation of $m_0(t)$ is minimax rather than MVUL. Suppose that $\{Y(t) : t \in I\}$ is of the form (1.1), and the mean function m_0 is known to be bounded by the variance in the sense that

$$\sup\{|E_{m_0}(V)| \cdot [\mathrm{Var}(V)]^{-\frac{1}{2}} : V \neq 0 \text{ and } V \in L[Y(t) : t \in I]\} \leqq \alpha , \tag{4.1}$$

where α is a known positive number. If $m \in H(k)$, then

$$||m|| = \sup\{|\langle m, v\rangle| \cdot ||v||^{-1} : 0 \neq v \in H(k)\} . \tag{4.2}$$

For $V \in L[Y(t) : t \in I]$, let v denote the isometric image of V in $H(k)$ under the isometry taking $Y(t)$ to $k(\cdot, t)$ for $t \in I$. Then for $m \in H(k)$

$$E_m(V) = \langle m, v\rangle . \tag{4.3}$$

In view of (4.2) and (4.3), it can be seen that assumption (4.1) is equivalent to the assumption that $m_0 \in M$ where

$$M = \{m \in H(k) : ||m|| \leqq \alpha\} . \tag{4.4}$$

For each $t \in I$, $m_0(t)$ is estimated by an element $V = \sum_{j=1}^n \hat{b}_j(t)Y(t_j) \in L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$ which minimizes

$$\sup\{E_m[V - m(t)]^2 : m \in M\} . \tag{4.5}$$

The following two remarks can be seen from (4.3) and the fact that

$$\mathrm{Var}\,(V) = \|v\|^2_{H(k)} = \sup\,\{\langle v, f\rangle^2 : f \in H(k), \|f\| = 1\}\,.$$

REMARK 4.1. For $V \in L^2[Y(t)\colon t \in I]$ and fixed $t \in I$, the quantity (4.5) is equal to

$$\mathrm{Var}\,(V) + \alpha^2 \sup\,\{[E_m(V) - m(t)]^2 \cdot \|m\|^{-2} : 0 \neq m \in H(k)\}\,. \tag{4.6}$$

REMARK 4.2. For fixed $t \in I$, the quantity (4.5) is also equal to

$$\mathrm{Var}\,(V) + \alpha^2\, \mathrm{Var}\,(V - Y(t))\,. \tag{4.7}$$

THEOREM 4.1. *Let* $\{Y(t) : t \in I\}$ *have the model* (1.1) *where, for a known positive number* α, *M is given by* (4.4). *For fixed* $t \in I$, *let* $\hat{V}(t) = \sum_{j=1}^n \hat{b}_j(t)Y(t_j)$ *minimize* (4.5) *for* $V \in L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$. *Given particular observations* $\{Y(t_j) = \lambda_j : j = 1, 2, \cdots, n\}$, *the minimax mean-estimation function* $\hat{v}$, *where for* $t \in I$

$$\hat{v}(t) = \textstyle\sum_{j=1}^n \hat{b}_j(t)\lambda_j\,,$$

is

$$\frac{\alpha^2}{1 + \alpha^2}\hat{f}\,,$$

where $\hat{f}$ *is the unique function* $f \in H(k)$ *that minimizes* $\|f\|$ *among functions satisfying* $f(t_j) = \lambda_j$ *for* $j = 1, 2, \cdots, n$.

PROOF. For $V \in L^2[Y(t)\colon t \in I]$ let v denote the isometric image of V under the usual isometry from $L^2[Y(t)\colon t \in I]$ onto $H(k)$. Let J denote the isometric image of $L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$. It follows from Remark 4.2 that the quantity to be minimized for $v \in J$ is

$$\|v\|^2 + \alpha^2\|v - k(\cdot, t)\|^2\,. \tag{4.8}$$

Since $v \in J$, it follows that for some $\mathbf{c} = [c_1, c_2, \cdots, c_n]'$,

$$v = \textstyle\sum_{j=1}^n c_j k(\cdot, t_j)\,.$$

Let K be the $n \times n$ matrix whose ijth element is $k(t_i, t_j)$. Then (4.8) becomes

$$\mathbf{c}'K\mathbf{c} + \alpha^2(\mathbf{c}'K\mathbf{c} - 2\mathbf{c}'\mathbf{k}_t + k(t, t)) \tag{4.9}$$

where $\mathbf{k}_t' = [k(t_1, t), k(t_2, t), \cdots, k(t_n, t)]$. Expression (4.9) is minimized by

$$\hat{\mathbf{c}} = \frac{\alpha^2}{1 + \alpha^2} K^{-1}\mathbf{k}_t$$

so that $\hat{V}(t)$ is given by

$$\hat{V}(t) = \frac{\alpha^2}{1 + \alpha^2} \mathbf{k}_t'K^{-1}\mathbf{Y} \tag{4.10}$$

where $\mathbf{Y}' = [Y(t_1), Y(t_2), \cdots, Y(t_n)]'$. It follows that the minimax mean-estimation function $\hat{v}$ satisfies

$$\hat{v}(t) = \frac{\alpha^2}{1 + \alpha^2} \mathbf{k}_t'K^{-1}\boldsymbol{\lambda}$$

where $\boldsymbol{\lambda}' = [\lambda_1, \lambda_2, \cdots, \lambda_n]$, and consequently

$$\hat{v} = \frac{\alpha^2}{1+\alpha^2}\hat{f}\,.$$

5. Extensions. Under certain conditions, Theorem 3.1 and Theorem 3.2 can be extended to the case when T has infinitely many elements and M is an infinite-dimensional linear space of functions on I. In an application (see Example 5.1) of Theorem 5.1, for a certain process $\{Y(t): t \in [0, 1]\}$ and a differential operator $\mathscr{L}$ of the form (2.1), an MEUL prediction function $\hat{y}$, given (infinitely many) particular observations $\{Y(t) = g(t): t \in T\}$, is an $\mathscr{L}$-spline of interpolation to the (infinitely many) points $\{(t, g(t)): t \in T\}$. A function $\hat{y}$ is called an *$\mathscr{L}$-spline of interpolation* to $\{(t, g(t)): t \in T\}$ if $\hat{y}$ minimizes

$$\int_0^1 [(\mathscr{L}y)(t)]^2\, dt$$

among functions $y \in H_b$ satisfying

$$y(t) = g(t) \qquad \text{for} \quad t \in T\,. \tag{5.1}$$

(The set H_b is defined in Section 2.)

Just as Lemma 2.1 led to the results of Section 3, Lemma 5.1, which is an extension of Lemma 2.1, leads to the results of Section 5. The main difference between the proof of Lemma 5.1 and the proof by Kimeldorf and Wahba (1970b) of Lemma 2.1 is the need in proving Lemma 5.1 to show that the sum of two particular closed subspaces of a Hilbert space is closed. In general, if either of two closed subspaces is finite-dimensional, then their sum is closed; an example by Halmos (1951) shows that if neither closed subspace is finite-dimensional, then their sum is not necessarily closed.

LEMMA 5.1. *Let $H = H_1 + H_2$ be the direct orthogonal sum of real Hilbert spaces H_1 and H_2. Let P_i be the projection operator onto H_i and J be a closed subspace of H such that $P_1(J) = H_1$. Then for any given elements, $\tilde{w}, \tilde{u} \in H$:*

(a) *there exists a unique element $w = \hat{w} \in J$ which minimizes $||\tilde{w} - w||^2$ subject to the constraint $P_1(\tilde{w} - w) = 0$;*

(b) *there exists a unique element $u = \hat{u} \in H$ for which $||P_2(u)||^2$ is minimized among elements u satisfying $\tilde{u} - u \in J^\perp$;*

(c) $\langle \tilde{u}, \hat{w}\rangle = \langle \hat{u}, \tilde{w}\rangle$.

PROOF. For Hilbert spaces A, B and C, the notation $A = B \oplus C$ means $A = B + C$ and $B \cap C = \{0\}$. That is, each $a \in A$ has a unique representation $a = b + c$ with $b \in B$ and $c \in C$. The notation $A = B \oplus^\perp C$ means $A = B \oplus C$ and B is perpendicular to C.

If $h_1, h_2 \in H$ and D is a closed subspace of H, $h_2 + D$ is called a flat, and the unique element $\hat{h} \in h_2 + D$ minimizing $||h_1 - h||^2$ among $h \in h_2 + D$ is called the projection of h_1 onto the flat $h_2 + D$.

It will be shown that

$$(5.2) \qquad P_2(J^\perp) \text{ is closed and } \quad H_2 = P_2(J^\perp) \oplus^\perp (J \cap H_2)\,.$$

Observe that $H_2 = (J \cap H_2) \oplus^\perp [H_2 \cap (J \cap H_2)^\perp] = (J \cap H_2) \oplus^\perp [H_2 \cap \overline{(H_1 + J^\perp)}] = (J \cap H_2) \oplus^\perp \overline{P_2(J^\perp)}$ since $\overline{H_1 + J^\perp} = H_1 + \overline{P_2(J^\perp)}$. The proof that $P_2(J^\perp)$ is closed is presented at the conclusion of the proof of Lemma 5.1.

In view of (5.2) and the fact that $H = H_1 \oplus^\perp H_2$ it follows that

$$(5.3) \qquad H = H_1 \oplus^\perp P_2(J^\perp) \oplus^\perp (J \cap H_2)$$

where H_1, $P_2(J^\perp)$, and $(J \cap H_2)$ are mutually perpendicular closed subspaces of H.

It is easily seen from (5.3) that

$$(5.4) \qquad \text{for } \quad x, y \in J \qquad P_1(x) = P_1(y) \Rightarrow P_{P_2(J^\perp)}(x) = P_{P_2(J^\perp)}(y)\,.$$

Since $P_1(J) = H_1$, there exists $w_0 \in J$ such that $P_1(w_0) = P_1(\tilde{w})$. From (5.4) it follows that $\hat{w}$ is the projection of $\tilde{w}$ onto the flat $w_0 + (J \cap H_2)$ and that

$$(5.5) \qquad P_1(\hat{w}) = P_1(\tilde{w})\,, \qquad P_{P_2(J^\perp)}(\hat{w}) = P_{P_2(J^\perp)}(w_0)\,, \qquad \text{and}$$
$$P_{J \cap H_2}(\hat{w}) = P_{J \cap H_2}(\tilde{w})\,.$$

The element $\hat{u}$ of part (b) of Lemma 5.1 is the unique element x (if it exists) of the flat $\tilde{u} + J^\perp$ which minimizes $||P_2(x)||^2$. For $u \in J^\perp$, it follows from (5.3) that

$$(5.6) \qquad \tilde{u} + u = P_1(\tilde{u} + u) + P_{P_2(J^\perp)}(\tilde{u} + u) + P_{(J \cap H_2)}(\tilde{u})$$

and that

$$(5.7) \qquad ||P_2(\tilde{u} + u)||^2 = ||P_{P_2(J^\perp)}(\tilde{u} + u)||^2 + ||P_{J \cap H_2}(\tilde{u})||^2\,.$$

It follows from the assumption $P_1(J) = H_1$ that

$$(5.8) \qquad \text{for } \quad x, y \in J^\perp \qquad P_2(x) = P_2(y) \Rightarrow P_1(x) = P_1(y)\,.$$

In view of (5.7) and (5.8), it can be seen that

$$(5.9) \qquad \hat{u} = \tilde{u} + u_0$$

where u_0 is the unique element $u \in J^\perp$ satisfying $P_{P_2(J^\perp)}(u) = -P_{P_2(J^\perp)}(\tilde{u})$. Hence it trivially follows that

$$(5.10) \qquad P_{P_2(J^\perp)}(\hat{u}) = 0 \qquad \text{and} \qquad P_{J \cap H_2}(\hat{u}) = P_{J \cap H_2}(\tilde{u})\,.$$

Since $\hat{w} \in J$ and $u_0 \in J^\perp$, (5.9) implies that

$$(5.11) \qquad \langle \hat{w}, \tilde{u} \rangle = \langle \hat{w}, \hat{u} \rangle\,.$$

It follows from (5.5) and (5.10) that

$$(5.12) \qquad \langle \hat{u}, \hat{w} \rangle = \langle \hat{u}, \tilde{w} \rangle\,.$$

Part (c) is an immediate consequence of (5.11) and (5.12).

The proof that $P_2(J^\perp)$ is closed uses the fact that if A and B are closed subspaces of a Hilbert space then the map P_A from B to A is the adjoint of the map P_B from A to B. Hence the map $P_{J^\perp}$ from H_2 to $J^\perp$ is the adjoint of the map P_2

from $J^\perp$ to H_2. Let $j^\perp \in J^\perp$. Since $P_1(J) = H_1$, there exists $j \in J$ such that $P_1(j) = P_1(j^\perp)$. Since $j^\perp - j \in H_2$ and $P_{J^\perp}(j^\perp - j) = j^\perp$ it follows that $P_{J^\perp}(H_2) = J^\perp$. Hence the adjoint of the map P_2 has a closed range and consequently, by VI.6.4 of Dunford and Schwartz (1958), $P_2(J^\perp)$ is closed.

REMARK 5.1. Let H_1 and k_0 correspond to a differential operator $\mathscr{L}$ as in Kimeldorf and Wahba (1970b). Suppose that $P_{H_1}(J_T{}^0) = H_1$ where $T \subseteqq [0, 1]$ and $J_T{}^0 = L^2[k_0(\cdot, t) : t \in T]$. It follows from part (b) of Lemma 5.1 that a collection of points $\{(t, g(t)) : t \in T\}$ can be interpolated by an $\mathscr{L}$-spline if and only if there exists $y \in H_b$ such that y satisfies (5.1). There will always exist such an element y if T contains only finitely many elements. The $\mathscr{L}$-spline, if it exists, is unique.

Let $\{Y(t) : t \in I\}$ have the model (1.1) where M is a closed subspace of $H(k)$ and M need not be finite-dimensional. Suppose that g is a function on $T \subseteqq I$. Then observing that $Y(t) = g(t)$ for all $t \in T$ is equivalent to observing for all random variables $V = \sum c_j Y(t_j) \in L[Y(t) : t \in T]$ that $V = \sum c_j g(t_j)$. However, if T contains infinitely many elements it is not obvious what should be considered to be the observed value of V if V is known only to belong to $L^2[Y(t) : t \in T]$. Suppose, however, that there exists $f \in H(k)$ such that $f(t) = g(t)$ for each $t \in T$. If $V \in L^2[Y(t) : t \in T]$ and $\{V_n\}_{n=1}^\infty$ is any sequence in $L[Y(t) : t \in T]$ such that $\lim_{n\to\infty} V_n = V$, where the limit is in $L^2[Y(t) : t \in I]$, then the observed value $\langle f, v_n \rangle$ of V_n converges as $n \to \infty$ to $\langle f, v \rangle$, where v_n and v are the isometric images in $H(k)$ of V_n and V under the isometry taking $Y(t)$ to $k(\cdot, t)$. Since the value $\langle f, v \rangle$ is independent of the approximating sequence $\{V_n\}_{n=1}^\infty$, it is reasonable to take $\langle f, v \rangle$ as the observed value for V, and this procedure is followed in Theorem 5.1 and Theorem 5.2.

THEOREM 5.1. *Let* $\{Y(t) : t \in I\}$ *have the model* (1.1) *where* M *is a closed subspace of* $H(k)$. *Suppose that* $P_M(J_T) = M$ *where* $J_T = L^2[k(\cdot, t) : t \in T]$. *For each* $t \in T$, *let* $\hat{Y}(t)$ *be the* MEUL *predictor for* $Y(t)$ *among elements of* $L^2[Y(t) : t \in T]$. *If for some function* $f \in H(k)$, $Y(t) = f(t)$ *is observed for each* $t \in T$, *then the* MEUL *prediction function* $\hat{y}$ *is the unique function minimizing*

$$||P_{M^\perp}(y)||^2$$

among functions satisfying

(a) $y \in H(k)$

and

(b) $y(t) = f(t)$ *for* $t \in T$.

PROOF. This proof is similar to the proof of Theorem 3.1. For fixed $t \in I$, let $\tilde{w}$ denote $k(\cdot, t)$ and $\tilde{u}$ denote f. For each $V \in L^2[Y(t) : t \in T]$ let v denote the RKHS isometric image of V under the usual isometry. If Lemma 5.1 is applied to $\tilde{w}$ and $\tilde{u}$ above with $H = H(k)$, $H_1 = M$, $H_2 = M^\perp$, and $J = J_T$, then if $\hat{v}_t$ denotes $\hat{w}$, it can be seen that the MEUL predictor $\hat{Y}(t)$ for $Y(t)$ is $\hat{V}_t$. For each

$t \in I$ the MEUL prediction function $\hat{y}$ satisfies $\hat{y}(t) = \langle f, \hat{v}_t \rangle = \langle \tilde{u}, \hat{w} \rangle = \langle \hat{u}, \tilde{w} \rangle = \langle \hat{u}, k(\cdot, t) \rangle = \hat{u}(t)$ where $\hat{u}$ is the unique function $u \in H(k)$ minimizing $||P_{M^\perp}(u)||^2$ subject to the condition that $u(t) = f(t)$ for each $t \in T$.

THEOREM 5.2. *Under the hypotheses of Theorem* 5.1, *let* $\hat{Z}(t)$ *denote the* MVUL *estimator for* $m_0(t)$. *If for each* $t \in T$, $Y(t) = f(t)$ *is observed for some function* $f \in H(k)$, *then the* MVUL *mean-estimation function* $\hat{z}$ *is* $P_M(\hat{y})$ *where* $\hat{y}$ *is the* MEUL *prediction function.*

PROOF. This proof is similar to the proof of Theorem 3.2. If for fixed $t \in I$, Lemma 5.1 is applied with $\tilde{w} = P_M(k(\cdot, t))$, $\tilde{u} = f$, $H = H(k)$, $H_1 = M$, $H_2 = M^\perp$, and $J = J_T$, then if $\hat{v}_t$ denotes $\hat{w}$, it can be seen that the MVUL estimator $\hat{Z}(t)$ is $\hat{V}_t$, the isometric image of $\hat{v}_t$. For each $t \in I$ the MVUL mean-estimation function $\hat{z}$ satisfies $\hat{z}(t) = \langle f, \hat{v}_t \rangle = \langle \tilde{u}, \hat{w} \rangle = \langle \hat{v}, \tilde{w} \rangle = \langle \hat{u}, P_M(k(\cdot, t)) \rangle = [P_M(\hat{u})](t)$. Since $\hat{u} = \hat{y}$, the proof is complete.

REMARK 5.2. Suppose that the model of Kimeldorf and Wahba (1970b) is altered in three ways. A subset T of $[0, 1]$ containing infinitely many elements replaces $\{t_1, t_2, \cdots, t_n\}$. The assumption $P_{H(k_1)}(J_T^0) = H(k_1)$, where $J_T^0 = L^2[k_0(\cdot, t) : t \in T]$, replaces the assumption that rank $(A) = q$ where $A = [f_i(t_j)]_{q \times n}$. The covariance kernel for $\{Y(t) : t \in [0, 1]\}$ is k_0 instead of k. (The set of mean functions is $H(k_1)$.) If, for some function $f \in H_b$, $Y(t) = f(t)$ is observed for $t \in T$, then the MEUL prediction function $\hat{y}$ is the unique $\mathscr{L}$-spline of interpolation to $\{(t, f(t)) : t \in T\}$.

The following example is an application of Theorem 5.1 and Remark 5.2 in which the MEUL prediction function is an $\mathscr{L}$-spline of interpolation to infinitely many points.

EXAMPLE 5.1. Let $\{N(t) : t \in [0, 1]\}$ be the Poisson process with intensity one. Let $N^*(t)$ denote $N(t) - t$. Then $E(N^*(t)) = 0$ for each $t \in [0, 1]$, and for $s, t \in [0, 1]$, $\text{Cov}(N^*(s), N^*(t)) = \min(s, t)$. Let the random variable U have mean zero and variance one and be independent of $\{N(t) : t \in [0, 1]\}$. If for each $t \in [0, 1]$, $X(t)$ is defined to be $U + N^*(t)$, then $\{X(t) : t \in [0, 1]\}$ has mean zero and covariance kernel $k(s, t) = 1 + \min(s, t)$, which is positive definite on $[0, 1]$. Now let the process $\{Y(t) : t \in [0, 1]\}$ have the model

$$Y(t) = \theta_0 + X(t)$$

where θ_0 is an unknown real number. Let $\{t_n\}_{n=1}^\infty$ be a sequence in $[0, 1]$ such that $t_n < t_{n+1}$ for all integers n. With $\mathscr{L} = \mathscr{D}$, it can be seen from Kimeldorf and Wahba (1970b) that $H(k) = \{f : f \text{ is absolutely continuous and } \mathscr{L}f \in L^2[0, 1]\}$ and if M is the subspace consisting of the constant functions then $\mathscr{L}$ maps $M^\perp$ isometrically onto $L^2[0, 1]$. With probability one, $\sup_n (N(t_n)) \leq N(1)$; hence, with probability one the sequence $\{N(t_n)\}_{n=1}^\infty$ will be a nondecreasing integer-valued sequence taking on finitely many values. Let $\{Y(t_n) = \lambda_n\}_{n=1}^\infty$ be a sequence of observations. If α denotes the unknown value taken on by U, then for

$n = 1, 2, \cdots$ the value taken on by $N(t_n)$ is $\lambda_n - \theta_0 - \alpha + t_n$. The constant function $\theta_0 + \alpha$ and the function g, where $g(t) = t$ for $t \in [0, 1]$, are both in $H(k)$. Unless an event of probability zero has occurred, $\{\lambda_n - \theta_0 - \alpha + t_n\}_{n=1}^{\infty}$ is a nondecreasing sequence with finitely many jumps, and $\{(t_n, \lambda_n - \theta_0 - \alpha + t_n)\}_{n=1}^{\infty}$ can be interpolated by a linearly segmented function $f^* \in H(k)$. The function $f = f^* + \theta_0 + \alpha - g$ is in $H(k)$ and interpolates the observed sequence $\{(t_n, \lambda_n)\}_{n=1}^{\infty}$. Thus, for the model of this example with $T = \{t_n\}_{n=1}^{\infty}$, with probability one the hypotheses of Theorem 5.1 are satisfied, and the MEUL prediction function $\hat{y}$ is the unique $\mathscr{L}$-spline of interpolation to $\{(t_n, \lambda_n)\}_{n=1}^{\infty}$.

REMARK 5.3. If the process $\{Y(t) : t \in I\}$ is Gaussian and if T contains infinitely many elements, then it can be shown that the probability of the existence of a function $f \in H(k)$ for which $Y(t) = f(t)$ for all $t \in T$ is zero.

The next theorem extends Remark 3.3 to the case when M is a closed subspace of $H(k)$ and $\{m_i\}_{i=1}^{\infty}$ is an orthonormal basis for M.

THEOREM 5.3. *Let* $\{Y(t) : t \in I\}$ *have the model* (1.1) *where* M *is a closed subspace of* $H(k)$ *and* $\{m_i\}_{i=1}^{\infty}$ *is an orthonormal basis for* M. *Hence* m_0 *has a unique but unknown representation* $\sum_{i=1}^{\infty} \theta_i m_i$. *If* $P_M(J_T) = M$ *where* $J_T = L^2[k(\cdot, t) : t \in T]$, *then there exists in* $L^2[Y(t) : t \in T]$ *an* MVUL *estimator* A_i *for* θ_i *for* $i = 1, 2, \cdots$. *For each* $t \in I$, *the sequence* $\{\sum_{i=1}^{n} A_i m_i(t)\}_{n=1}^{\infty}$ *converges in* $L^2[Y(t) : t \in I]$ *to* $\hat{Z}(t)$, *the* MVUL *estimator for* $m_0(t)$. *That is,* $\hat{Z}(t) = \sum_{i=1}^{\infty} A_i m_i(t)$ *for each* $t \in I$.

PROOF. Under the linear mapping that takes $k(\cdot, t)$ to $Y(t)$, $L^2(Y(t) : t \in I)$ is isometric to $H(k)$, and J_T is isometric to $L^2(Y(t) : t \in T)$. The assumption $P_M(J_T) = M$ implies that the mapping P_M from $J \cap (J \cap M^{\perp})^{\perp}$ to M is one-to-one and onto. From a consequence of the interior mapping principle (see II.2.2 of Dunford and Schwartz (1958)) it follows that the mapping P_M^{-1} from M onto $J \cap (J \cap M^{\perp})^{\perp}$ is continuous.

If $\hat{z}_t = P_M^{-1}(P_M(k(\cdot, t)))$, it can be seen that $\hat{Z}(t)$ is the isometric image of $\hat{z}_t$. If $a_i = P_M^{-1}(m_i)$, then it can be seen that the MVUL estimator for θ_i is the isometric image A_i of a_i. Note that $P_M(k(\cdot, t)) = \sum_{i=1}^{\infty} m_i(t) m_i = \lim_{n\to\infty} \sum_{i=1}^{n} m_i(t) m_i$. It now follows from the linearity and continuity of P_M^{-1} that $P_M^{-1}(P_M(k(\cdot, t))) = \sum_{i=1}^{\infty} a_i m_i(t)$, and it follows from the isometry that $\hat{Z}(t) = \sum_{i=1}^{\infty} A_i m_i(t)$.

REMARK 5.4. The previous theorem is still true if $\{m_i\}_{i=1}^{\infty}$ is assumed only to be a Schauder basis, rather than an orthonormal basis. A very similar result is true when M has an uncountable orthonormal basis $\{m_\alpha\}$ since, for any fixed $t \in I$, $m_\alpha(t) = 0$ for all but countably many of the basis elements.

REFERENCES

[1] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 377–404.

[2] DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators—Part I.* Interscience, New York.

[3] DUNFORD, N. and SCHWARTZ, J. T. (1963). *Linear Operators—Part II.* Interscience, New York.

[4] HALMOS, P. R. (1951). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity.* Chelsea, New York.

[5] KIMELDORF, G. S. and WAHBA, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

[6] KIMELDORF, G. S. and WAHBA, G. (1970b). Spline functions and stochastic processes. *Sankhyā Ser. A* **32** 173–180.

[7] PARZEN, E. (1959). Statistical inference on time series by Hilbert space methods. Stanford Univ. Statistics Department Technical Report #23.

[8] PARZEN, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[9] PARZEN, E. (1962). *Stochastic Processes.* Holden-Day, San Francisco.

[10] PARZEN, E. (1970). Statistical inference on time series by RKHS methods. In *Seminar of the Canadian Mathematical Congress on Time Series and Stochastic Processes* (R. Pyke, ed.), 1–37.

DEPARTMENT OF STATISTICS
NUCLEAR SCIENCES BUILDING
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

PROGRAMS IN MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
BOX 688
RICHARDSON, TEXAS 75080