J. R. Statist. Soc. B (1983),
45, *No.* 1, *pp.* 133–150

Bayesian "Confidence Intervals" for the Cross-validated Smoothing Spline

By GRACE WAHBA

University of Wisconsin, USA

[Received August 1981. Revised August 1982]

SUMMARY

We consider the model $Y(t_i) = g(t_i) + \epsilon_i$, $i = 1, 2, \ldots, n$, where $g(t)$, $t \in [0, 1]$ is a smooth function and the $\{\epsilon_i\}$ are independent $N(0, \sigma^2)$ errors with σ^2 unknown. The cross-validated smoothing spline can be used to estimate g non-parametrically from observations on $Y(t_i)$, $i = 1, 2, \ldots, n$, and the purpose of this paper is to study confidence intervals for this estimate. Properties of smoothing splines as Bayes estimates are used to derive confidence intervals based on the posterior covariance function of the estimate. A small Monte Carlo study with the cubic smoothing spline is carried out to suggest by example to what extent the resulting 95 per cent confidence intervals can be expected to cover about 95 per cent of the true (but in practice unknown) values of $g(t_i)$, $i = 1, 2, \ldots, n$. The method was also applied to one example of a two-dimensional thin plate smoothing spline. An asymptotic theoretical argument is presented to explain why the method can be expected to work on fixed smooth functions (like those tried), which are "smoother" than the sample functions from the prior distributions on which the confidence interval theory is based.

Keywords: SPLINE SMOOTHING; CROSS-VALIDATION; CONFIDENCE INTERVALS

1. INTRODUCTION

Consider the model

$$Y(t_i) = g(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad t_i \in [0, 1], \tag{1.1}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim N(0, \sigma^2 I_{n \times n})$, σ^2 is unknown and $g(\cdot)$ is a fixed but unknown function with $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$. The smoothing spline estimate of g given $Y(t_i) = y_i$, $i = 1, 2, \ldots, n$, which we will call $g_{n,\lambda}$, is the minimizer of

$$n^{-1} \sum_{i=1}^{n} (g(t_i) - y_i)^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt$$

and is a polynomial spline of degree $2m-1$. The parameter λ controls the tradeoff between the infidelity $n^{-1} \Sigma (g_{n,\lambda}(t_i) - y_i)^2$ (summing over $i = 1, \ldots, n$) and the roughness $\int_0^1 (g_{n,\lambda}^{(m)}(t))^2 dt$ of the solution. For $m = 2$, $g_{n,\hat{\lambda}}$, where $\hat{\lambda}$ is the generalized cross-validation estimate of λ, is an apparently popular non-parametric estimate of g. (Code appears in IMSL, 1980.) $\hat{\lambda}$ estimates λ^*, which is the minimizer of the expected predictive mean square error $ER(\lambda)$, where $R(\lambda) = n^{-1} \Sigma [g(t_i) - g_{n,\lambda}(t_i)]^2$. λ^* depends on g, n and σ^2 (see Craven and Wahba, 1979; Utreras, 1979a).

The smoothing spline $g_{n,\lambda}$ is also a Bayes estimate of g if g is assumed to be a sample function from a certain zero mean Gaussian prior. This property of smoothing spline estimates was discussed in some detail in Wahba (1978) as well as in Kimeldorf and Wahba (1970, 1971). However, it is known (and will be illustrated in Section 2) that, if g is a sample function from the relevant prior, then $E\int_0^1 (g^{(m)}(t))^2 dt = \infty$.

In this paper, which is a sequel to Wahba (1978), we first use the properties of $g_{n,\lambda}$ as a Bayes

Present address: Professor Grace Wahba, Department of Statistics, University of Wisconsin, 1210 W. Dayton St., Madison, WI 53706, USA.

 0035–9246/83/45133 $2.00

estimate to derive confidence intervals, based on the posterior variances of the $g_{n,\lambda}(t_i)$. This derivation appears in Section 2. Similar results in somewhat different forms have previously been obtained by Gamber (1979a), Lucas (1978) and Wecker and Ansley (1980).

The interesting question concerning these confidence intervals is: Suppose *not* that g is a sample function from the relevant prior, but that g is some fixed function with $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$. Suppose further that λ is taken as $\hat{\lambda}$, the generalized cross-validation (GCV) estimate of λ. Then, is there any reason to believe that the resulting 95 per cent "confidence intervals" will cover the true $g(t_i)$ about 95 per cent of the time?

The remainder of this paper is devoted to presenting evidence that, for large n, the answer to this question is a qualified yes. The evidence we present is of two forms. First, in Section 3 we present a summary of the results of a modest Monte Carlo study with $m = 2$, with three different smooth g, three values of n (32, 64 and 128), five values of σ, and equally spaced data points.

The analytical evidence we present here goes as follows. Let $s_{ii}(\lambda)$ be the posterior variance of $g_{n,\lambda}(t_i)$ derived using the prior distribution in Section 2. We present an argument that, if g is any function with $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$, then for n large, and $\lambda = \lambda^*$ (the minimizer of $ER(\lambda^*)$), we have in Section 6

$$ER(\lambda^*) = \alpha \cdot n^{-1} \sum_{i=1}^{n} s_{ii}(\lambda^*)\,(1 + o(1)), \tag{1.2}$$

where $o(1) \to 0$ as $n \to \infty$, and α is some number between $[1 + (1/4m)]\ [1 - (1/2m)]$ and 1 (depending on g). This expression says, that provided $\lambda \approx \lambda^*$, then the average of the posterior variances which are used in the construction of the confidence intervals, asymptotically comes close to the average square bias plus variance of g_{n,λ^*}.

The theoretical confidence interval results here extend immediately to the generalized splines discussed in Section 3 of Wahba (1978). For applications to splines on the plane, in Euclidean 3-space and the sphere see Wahba (1981a, 1982), Wahba and Wendelberger (1980) and Wendelberger (1981, 1982). In Section 3, we give an example of the confidence intervals computed for a thin plate smoothing spline estimate of a two-dimensional surface with $n = 169$. In this first (and only) two dimensional example tried, the confidence intervals covered 162 or 95.8 per cent of the true functional values.

Our results may be used to obtain confidence intervals centred around quantities like $\hat{\beta} = \int_0^1 w(t) g_{n,\hat{\lambda}}(t) dt$, which is an estimate of $\beta = \int_0^1 w(t) g(t) dt$.

A number of theoretical results, Monte Carlo experiments and applications concerning smoothing splines with GCV are available, for example see Merz (1978), Nogues and Sielken (1980), Utreras (1979a), Wegman and Wright (1980) and references cited there. If g has $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$, then the expected predictive mean square error with optimum λ converges rapidly, that is

$$ER(\lambda^*) = O(n^{-2m/(2m+1)}),$$

these rates agree with the best achievable rates in Stone (1980).

Knafl *et al.*, (1982) has also recently proposed confidence intervals for some non-parametric estimates which are sometimes related to smoothing splines. For more on non-parametric regression, see, for example, Agarwal and Studden (1980), Gasser and Rosenblatt (1979) and Rice and Rosenblatt (1981). Our philosophy is in the spirit of one suggested by Berger (1980), which is, derive confidence intervals based on some prior distribution, then forget the prior and see how well the intervals can be expected to perform on cases of interest.

2. THE POSTERIOR COVARIANCE OF $g_{n,\lambda}(t)$ IN THE BAYES MODEL

To establish the notation we repeat from Wahba (1978) the prior distribution on $g(t)$, $t \in [0, 1]$ for which $g_{n,\lambda}(t)$, $t \in [0, 1]$ is the posterior mean. It is: $g(t)$, $t \in [0, 1]$ has the same distribution as

$$X_\xi(t) = \sum_{j=1}^{m} \theta_j \phi_j(t) + b^{\frac{1}{2}} Z(t), \quad t \in [0, 1], \tag{2.1}$$

where $\theta = (\theta_1, \ldots, \theta_m)' \sim N(0, \xi I_{m \times m})$, $\phi_j(t) = t^{j-1}/(j-1)!$, $j = 1, \ldots, m$, $b = \sigma^2/n\lambda$ and $Z(\cdot)$ is the m-fold integrated Wiener process,

$$Z(t) = \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} \, dW(u),$$

$W(u)$ being the Wiener process, and $\xi \to \infty$. Then (from Wahba, 1978)

$$g_{n,\lambda}(t) = \lim_{\xi \to \infty} E_\xi\{g(t) \mid Y = y\},$$

where $Y = (Y(t_1), \ldots, Y(t_n))'$, $y = (y_1, \ldots, y_n)'$ and E_ξ is expectation over the posterior distribution of $g(t)$ with the prior (2.1) ($\xi \to \infty$ corresponds to a "partially improper" prior).

Before proceeding with the derivation of the posterior covariance of $g_{n,\lambda}(t)$ we observe that, in the $m = 1$ case, if g is continuous and $\int_0^1 (g'(t))^2 dt < \infty$, then we must have

$$\lim_{k \to \infty} k^{-1} \sum_{i=1}^{k} \left[k\left(g\left(\frac{i}{k}\right) - g\left(\frac{i-1}{k}\right)\right)\right]^2 = \int_0^1 (g'(t))^2 dt, \tag{2.2}$$

but if g is distributed according to the prior (2.1) then $g(t)$ is distributed as $\theta_1 + b^{\frac{1}{2}} W(t)$ and

$$\lim_{k \to \infty} E\, k^{-1} \sum_{i=1}^{k} \left[k\left(g\left(\frac{i}{k}\right) - g\left(\frac{i-1}{k}\right)\right)\right]^2 = \lim_{k \to \infty} kbE \sum_{i=1}^{k} \left[W\left(\frac{i}{k}\right) - W\left(\frac{i-1}{k}\right)\right]^2$$

$$= \lim_{k \to \infty} kb = \infty.$$

This argument can be repeated for $m = 2, 3, \ldots$ using mth divided differences instead of first differences. (More generally, where $\int_0^1 (g^{(m)}(t))^2 dt$ is replaced by $\| P_Q g \|_K^2$ in Section 3 of Wahba (1978) it can be shown that $E \| P_Q g \|_K^2 = \infty$.)

We now proceed to the derivation of the posterior covariance. Let

$$Q(s, t) = \int_0^1 \frac{(s-u)_+^{m-1}}{(m-1)!} \frac{(t-u)_+^{m-1}}{(m-1)!} \, du = EZ(s)Z(t),$$

let T be the $n \times m$ matrix with $j\nu$th entry $\phi_\nu(t_j)$ and let Q_n be the $n \times n$ matrix with jkth entry $Q(t_j, t_k)$. We always assume that the matrix T is of rank m; for this it is sufficient that there be at least m distinct t_i's. It will be convenient to use the so-called influence matrix $A(\lambda)$ defined by

$$\mathbf{g}_{n,\lambda} = A(\lambda) y,$$

where $\mathbf{g}_{n,\lambda} = (g_{n,\lambda}(t_1), \ldots, g_{n,\lambda}(t_n))'$. In the Bayes model discussed here we will substitute $\sigma^2/n\lambda$ for b, until further notice. A rather involved formula for $A(\lambda)$ is given in Wahba (1978, p. 367) but it can easily be seen, by substituting (4.2) of Wahba (1978) into (4.4) of Wahba (1978) that $A(\lambda)$ has the representation

$$A(\lambda) = I - n\lambda B'(BQ_nB' + n\lambda I)^{-1} B,$$

where B is any $n - m \times n$ dimensional matrix whose $n - m$ rows are orthonormal, and orthogonal to the columns of T. For later use we note that the ijth entry $a_{ij}(\lambda)$ of $A(\lambda)$ satisfies

$$a_{ij}(\lambda) = \frac{\partial g_{n,\lambda}(t_i)}{\partial y_j} \tag{2.3}$$

which is the source of the terminology "influence matrix".

Theorem 1. The posterior covariance matrix of $g_{n,\lambda}$ is

$$\operatorname{cov}(\mathbf{g}_{n,\lambda} \mid Y(t_1), \ldots, Y(t_n)) = \sigma^2 A(\lambda). \tag{2.4}$$

It is of some interest to obtain the complete posterior covariance function, call it $Q_{n,\lambda}(s, t)$, of $g_{n,\lambda}(s)$, where

$$Q_{n,\lambda}(s, t) = \operatorname{cov}\{g_{n,\lambda}(s), g_{n,\lambda}(t) \mid Y(t_1), \ldots, Y(t_n)\}.$$

This is because the posterior variance of a linear functional β of g can be estimated as the corresponding linear functional of $g_{n,\lambda}$, it will be the Bayes estimate, and furthermore its posterior variance can be determined by applying the functional to $Q_{n,\lambda}$ separately in each of its arguments. For example, let $w(t)$ be some continuous function and let $\beta = \int_0^1 w(t)g(t)dt$, $\hat{\beta} = \int_0^1 w(t)g_{n,\lambda}(t)dt$. Then it can be shown that $\hat{\beta} = E(\beta \mid y)$ and var $(\hat{\beta} \mid y) = \int_0^1\int_0^1 w(s)w(t)Q_{n,\lambda}(s, t)ds\,dt$. It is possible to extend this result to all linear functionals of g which are continuous in the appropriate reproducing kernel Hilbert space discussed in Section 3 of Wahba (1978).

Theorem 2. Let $0 < t_1, < \ldots, < t_n$ (so that Q_n is invertible). The posterior covariance function $Q_{n,\lambda}(s, t)$ of $g_{n,\lambda}(t)$, $t \in [0, 1]$ given $(Y(t_1, \ldots, Y(t_n))$ is given by

$$\begin{aligned} Q_{n,\lambda}(s, t) = {} & \operatorname{cov}\{(g(s), g(t)) \mid g(t_1), \ldots, g(t_n)\} \\ & + \sigma^2\{(\phi_1(s), \ldots, \phi_m(s))\theta^{-1}T'Q_n^{-1} + (Q(s, t_1), \ldots, Q(s, t_n))P_n\} \\ & \times A(\lambda)\{Q_n^{-1}T\theta^{-1}(\phi_1(t), \ldots, \phi_m(t))' + P_n(Q(t, t_1), \ldots, Q(t, t_n))'\}, \end{aligned}$$

where

$$\theta = T'Q_n^{-1}T, \quad P_n = Q_n^{-1} - Q_n^{-1}T\theta^{-1}T'Q_n^{-1}$$

and

$$\begin{aligned} \operatorname{cov}\{(g(s), g(t)) \mid g(t_1), \ldots, g(t_n)\} = b\Bigg\{ & (\phi_1(s), \ldots, \phi_m(s))\theta^{-1}\begin{pmatrix}\phi_1(t)\\ \vdots \\ \phi_m(t)\end{pmatrix} + Q(s, t) \\ & - (Q(s, t_1), \ldots, Q(s, t_n))Q_n^{-1}T\theta^{-1}\begin{pmatrix}\phi_1(t)\\ \vdots \\ \phi_m(t)\end{pmatrix} \\ & - (Q(t, t_1), \ldots, Q(t, t_n))Q_n^{-1}T\theta^{-1}\begin{pmatrix}\phi_1(s)\\ \vdots \\ \phi_m(s)\end{pmatrix} \\ & - (Q(s, t_1), \ldots, Q(s, t_n))P_n\begin{pmatrix}Q(t, t_1)\\ \vdots \\ Q(t, t_n)\end{pmatrix}\Bigg\}. \end{aligned}$$

We remark that b enters only if both s and t are not one of $(t_1, \ldots, t_n)$.

We now give the proofs of Theorems 1 and 2. The proofs are a direct application of Lemma 1 below.

Lemma 1. Let y, g and ϵ be zero mean Gaussian n-vectors and h a zero mean Gaussian l vector (all column vectors) with

$$y = g + \epsilon,$$

$Eee' = \sigma^2 I$, $Egg' = b\Sigma_{gg}$, $Egh' = b\Sigma_{gh}$, $Ehh' = b\Sigma_{hh}$, $E\epsilon h' = 0$, $Eeg' = 0$. Let $n\lambda = \sigma^2/b$ and $A(\lambda) = \Sigma_{gg}(\Sigma_{gg} + n\lambda I)^{-1}$, and suppose that Σ_{gg} is strictly positive definite. Then

$$E(h \mid y) = \Sigma_{hg}(\Sigma_{gg} + n\lambda I)^{-1} y, \tag{2.5}$$

$$\text{cov}(hh' \mid y) = b(\Sigma_{hh} - \Sigma_{hg}\Sigma_{gg}^{-1}\Sigma_{gh}) + \sigma^2\Sigma_{hg}\Sigma_{gg}^{-1}A(\lambda)\Sigma_{gg}^{-1}\Sigma_{gh}. \tag{2.6}$$

In particular, setting $h = g$ gives

$$E(g \mid y) = A(\lambda)y, \tag{2.7}$$

$$\text{cov}(g \mid y) = \sigma^2 A(\lambda). \tag{2.8}$$

The proof follows by application of Anderson (1958, p. 28), and tedious but straightforward algebra.

To prove Theorems 1 and 2, set $g = (g(t_1), \ldots, g(t_n))'$, $l = 1$, $h = g(s)$ and $\eta = \xi/b$. Then Σ_{hh}, Σ_{hg} and Σ_{gg} are determined, respectively by

$$Eg(s)g(t) = b\left[\eta \sum_{\nu=1}^{m} \phi_\nu(s)\phi_\nu(t) + Q(s,t)\right], \quad s, t \in [0, 1], \tag{2.9}$$

$$Eg(s)g = b[\eta T(\phi_1(s), \ldots, \phi_m(s))' + (Q(s,t_1), \ldots, Q(s,t_n))'], \quad s \in [0, 1], \tag{2.10}$$

$$Egg' = b[\eta T'T + Q_n]. \tag{2.11}$$

To complete the proof of Theorem 2, Σ_{hh}, Σ_{hg} and Σ_{gg} based on (2.9)–(2.11) are substituted into (2.5)–(2.8) and the limits taken as $\eta \to \infty$. These limits may be found using the following results (2.12)–(2.14) all found in Wahba (1978, equations (2.7)–(2.9)),

$$(\eta TT' + Q_n)^{-1} \equiv Q_n^{-1} - Q_n^{-1}T\theta^{-1}\{I + \eta^{-1}\theta^{-1}\}^{-1}T'Q_n^{-1}, \tag{2.12}$$

$$\lim_{\eta \to \infty} \eta T'(\eta TT' + Q_n)^{-1} = \theta T'Q_n^{-1}, \tag{2.13}$$

$$\lim_{\eta \to \infty} (\eta TT' + Q_n)^{-1} = P_n \tag{2.14}$$

and also

$$\lim_{\eta \to \infty} \{\eta I_{m\times m} - \eta T'(Q_n + \eta TT')^{-1}T'\eta\} = \theta^{-1}, \tag{2.15}$$

which can be obtained, after some manipulation by substituting (2.12) into the left-hand side, expanding powers of η and taking the limit. The author is grateful to Dr B. W. Silverman who provided (2.12)–(2.14).

The GCV estimate $\hat{\lambda}$ of λ used in the experiments below is the minimizer of the GCV function $V(\lambda)$ defined by

$$V(\lambda) = \frac{n^{-1}\,||(I - A(\lambda))y||^2}{(n^{-1}\text{Tr}(I - A(\lambda)))^2}, \tag{2.16}$$

for further details, see Craven and Wahba (1979).

3. THE MONTE CARLO EXPERIMENTS

The main Monte Carlo experiment consisted of a detailed study of the three cases of functions given below.

Case 1 $g(t) = \frac{1}{3}\beta_{10,5}(t) + \frac{1}{3}\beta_{7,7}(t) + \frac{1}{3}\beta_{5,10}(t)$,

Case 2 $g(t) = \frac{6}{10}\beta_{30,17}(t) + \frac{4}{10}\beta_{3,11}(t)$,

Case 3 $g(t) = \frac{1}{3}\beta_{20,5}(t) + \frac{1}{3}\beta_{12,12}(t) + \frac{1}{3}\beta_{7,30}(t)$,

where

$$\beta_{p,q}(t) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1}(1-t)^{q-1}, \quad 0 \leqslant t \leqslant 1.$$

All three functions satisfy g, g' continuous, $\int_0^1 (g''(t))^2 dt < \infty$, and are non-negative functions integrating to one. In all the experiments in this section we set $m = 2$. The choice of m is discussed further in Section 5. In addition, in an attempt to defeat the method Case 4 was chosen with a discontinuity in its first derivative and Case 5 was chosen to be discontinuous.

$$\text{Case 4} \quad g(t) = \begin{cases} 0, & 0 \leqslant t \leqslant \frac{1}{3}, \\ 36(t - \frac{1}{3}), & \frac{1}{3} \leqslant t \leqslant \frac{1}{2}, \\ 36(\frac{2}{3} - t), & \frac{1}{2} \leqslant t \leqslant \frac{2}{3}, \\ 0, & \frac{2}{3} \leqslant t \leqslant 1, \end{cases}$$

$$\text{Case 5} \quad g(t) = \begin{cases} 0, & 0 \leqslant t \leqslant \frac{1}{3}, \\ 72(t - \frac{1}{3}), & \frac{1}{3} \leqslant t \leqslant \frac{1}{2}, \\ 0, & \frac{1}{2} < t \leqslant 1. \end{cases}$$

In all cases, $t_i = i/n$.

To simplify the computer programming and economize in computer time, a periodic version of the smoothing spline estimate was actually implemented. The general case can be handled by the program developed by Wendelberger (1981). The test functions were deliberately chosen to satisfy the periodic boundary conditions $g^{(\nu)}(0) = g^{(\nu)}(1)$, $\nu = 0, 1, 2, 3$ so no new source of error is being introduced. The results can be expected to be similar to the general case provided g satisfies the Neumann boundary conditions $g''(0) = g'''(0) = g''(1) = g'''(1)$ which are always satisfied by the smoothing spline with $m = 2$. Let n be even and let F_n be the n-dimensional space of functions spanned by the sine and cosine functions

$$\{1, \cos 2\pi\nu t, \nu = 1, 2, \ldots, n/2, \sin 2\pi\nu t, \nu = 1, 2, \ldots, n/2 - 1\}.$$

It can be demonstrated that the minimizer in F_n of

$$n^{-1} \sum_{i=1}^{n} \left(g\left(\frac{i}{n}\right) - y_i\right)^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt \tag{3.1}$$

is

$$g_{n,\lambda}(t) = a_0 + 2 \sum_{\nu=1}^{n/2-1} \frac{a_\nu \cos 2\pi\nu t + b_\nu \sin 2\pi\nu t}{[1 + \lambda(2\pi\nu)^{2m}]} + \frac{a_{n/2} \cos \pi n t}{[1 + \lambda(\pi n)^{2m}]}, \tag{3.2}$$

where

$$a_\nu = n^{-1} \sum_{j=1}^{n} \left(\cos 2\pi\nu \frac{j}{n} \right) y_j, \quad \nu = 0, 1, \ldots, n/2; \quad b_\nu = n^{-1} \sum_{j=1}^{n} \left(\sin 2\pi\nu \frac{j}{n} \right) y_i,$$

$$\nu = 1, 2, \ldots, n/2 - 1. \quad (3.3)$$

$g_{n,\lambda}$ of (3.2) is the estimate of $g(t)$ that we will be using. The demonstration of (3.2) can be carried out by noting that if the discrete Fourier coefficients $\{a_\nu, b_\nu\}$ of (any) vector y are defined by (3.3) then

$$y_i = a_0 + 2 \sum_{\nu=1}^{n/2-1} \left(a_\nu \cos 2\pi\nu \left(\frac{i}{n} \right) + b_\nu \sin 2\pi\nu \left(\frac{i}{n} \right) \right) + a_{n/2} \cos 2\pi(n/2) \left(\frac{i}{n} \right),$$

and vice versa, and by using the second derivatives of the sine and cosine functions, and by substituting into (3.1).

It is not hard to show that

$$\| (I - A(\lambda)) y \|^2 \equiv \text{RSS}(\lambda) = 2 \sum_{\nu=1}^{n/2-1} \left(\frac{\lambda}{\lambda_\nu + \lambda} \right)^2 (a_\nu^2 + b_\nu^2) + \left(\frac{\lambda}{\lambda_{n/2} + \lambda} \right)^2 a_{n/2}^2, \tag{3.4}$$

$$a_{ii}(\lambda) \equiv a(\lambda) = n^{-1} + \frac{2}{n} \sum_{\nu=1}^{n/2-1} \frac{\lambda_\nu}{\lambda_\nu + \lambda} + n^{-1} \frac{\lambda_{n/2}}{\lambda_{n/2} + \lambda}$$

where $\lambda_\nu = (2\pi\nu)^{-2m}$. $V(\lambda)$ of (2.16) is computed from $V(\lambda) = \text{RSS}(\lambda)/n(1 - a(\lambda))^2$, and $\hat{\lambda}$ is its minimizer. $\hat{\lambda}$ is found by a global search based on equal increments in $\log_{10}\lambda$. $\hat{\sigma}^2(\lambda) = \text{RSS}(\lambda)/n(1 - a(\lambda))$, where $n(1 - a(\lambda)) = \text{EDF}(\lambda)$, the equivalent degrees of freedom for error when λ is used. ($na(\lambda) = \text{Tr}A(\lambda)$ is the EDF(λ) for signal, by analogy with regression.) The estimated 95 per cent confidence intervals (CI) are given by $g_{n,\hat{\lambda}}(i/n) \pm 1.96\hat{\sigma}(\hat{\lambda})\sqrt{(a(\hat{\lambda}))}$. There is a conceptual question whether 1.96 or the 0.025 point of the t distribution with EDF($\hat{\lambda}$) degrees of freedom should be used. When $n = 128$ or 64, the EDF($\hat{\lambda}$) was typically greater than 30 and $t_{0.025}(\text{EDF}(\hat{\lambda})) \approx 1.96$. For $n = 32$, the use of $t_{0.025}(\text{EDF}(\hat{\lambda}))$ instead of 1.96 would most likely have improved the confidence intervals obtained here somewhat. We also examined properties of the 95 per cent pseudo confidence intervals (PCI) given by $g_{n,\hat{\lambda}}(i/n) \pm 1.96\sigma\sqrt{(a(\hat{\lambda}))}$. Here σ is the standard deviation used in the generation of the $\{\epsilon_i\}$.

The main Monte Carlo experiment consisted of studying all $3 \times 3 \times 5 = 45$ combinations of Cases 1, 2,3, $n = 32, 64, 128$ and $\sigma = 0.0125, 0.025, 0.05, 0.1$ and 0.2. Data were generated for 10 replications of each of these 45 combinations of Cases, σ's and n's. To evaluate the method, for each replicate we computed the inefficiency ISUBV = $R(\hat{\lambda})/\min_\lambda R(\lambda)$, which measures how well $\hat{\lambda}$ estimates the minimizer of $R(\lambda)$, the VRATIO = $\hat{\sigma}^2(\hat{\lambda})/n^{-1} \Sigma \epsilon_i^2$, summing over i from 1 to n, which indicates how good $\hat{\sigma}^2(\hat{\lambda})$ is, and CI 95 and PCI 95. CI 95 is the percentage of the true $g(i/n)$ which were covered by the CI's and PCI 95 is the percentage of the true $g(i/n)$ covered by the PCI's. We first show a few examples. Fig. 1a gives a plot of $g(t)$, simulated data and $g_{n,\hat{\lambda}}(t)$ for one replicate of Case 1 with $\sigma = 0.1, n = 128$. Fig. 1b gives $g(t)$ and the simulated data from Fig. 1a along with a confidence "band". For visual effect the two dashed lines which are the curves $g_{n,\hat{\lambda}}(t) \pm 1.96\hat{\sigma}(\hat{\lambda})\sqrt{(a(\hat{\lambda}))}$ have been plotted. Strictly speaking these curves only have meaning at $t = i/n$, where they are the endpoints of the confidence intervals. All 128 of the CI's covered the true $g(i/n)$, but "just barely". Fig. 1c gives $V(\lambda)$ and $R(\lambda)$ plotted against $\log \lambda$. We had $\log \hat{\lambda} = -5.778$, the log of the minimizer of $R(\lambda)$ was -6.000, ISUBV = 1.078 and VRATIO was 1.04. Figs 2 and 3 are analogous to Fig. 1b, with one replicate each of Case 2, $\sigma = 0.1$, $n = 64$ and Case 3, $\sigma = 0.1$, $n = 32$ given. Table 1 gives ISUBV, VRATIO, PCI 95 and CI 95 for each of the

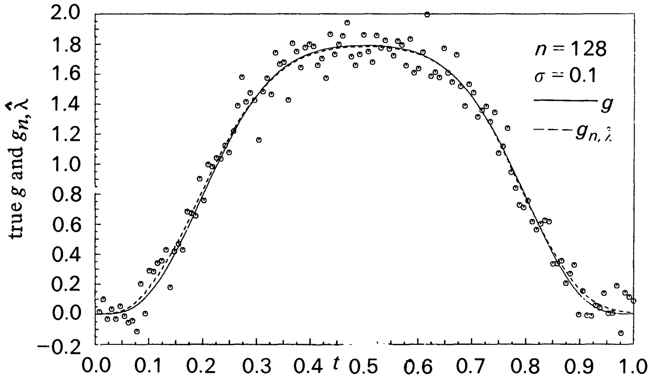

Fig. 1a. $g(t)$, simulated data, and $g_{n,\hat{\lambda}}(t)$ for an example of Case 1.

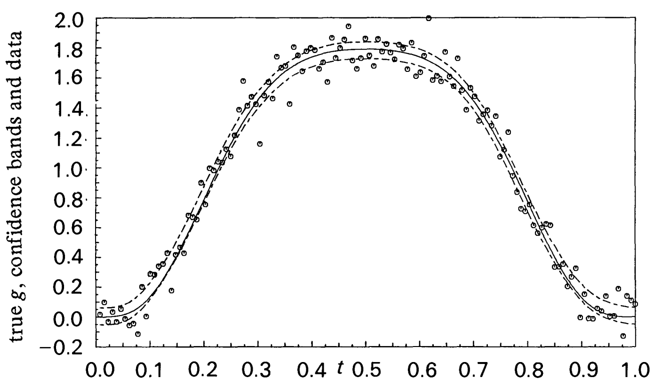

Fig. 1b. $g(t)$ and data of Figure 1 with 95 per cent confidence bands.

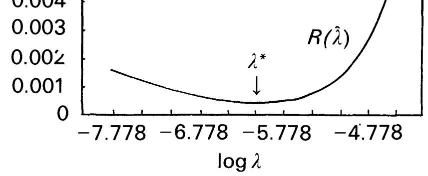

Fig. 1c. $R(\lambda)$ and $V(\lambda)$.

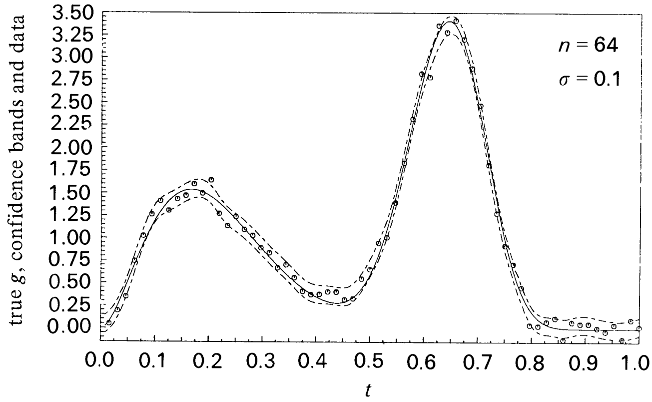

Fig. 2. $g(t)$, data and confidence bands. Case 2.

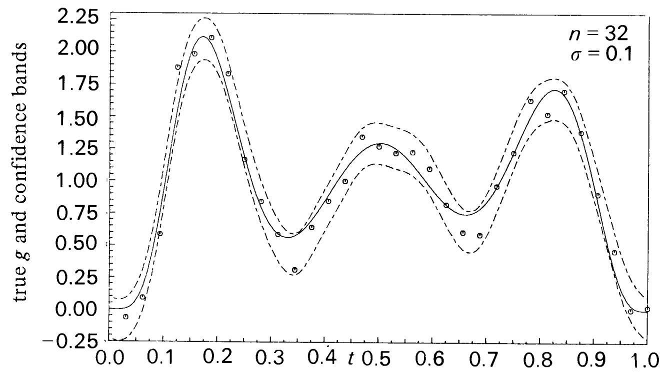

Fig. 3. $g(t)$, data and confidence bands, Case 3.

10 replicates for the combination Case 2, $n = 64$, $\sigma = 0.1$, as well as the sample means and standard deviations of the columns. This table and the corresponding tables for the other 44 combinations tried appears in Wahba (1981b). Due to lack of space these results will only be summarized here.

TABLE 1
Summary data for 10 *replications of Case* 2, $n = 64$, $\sigma = 0.1$

REPL	ISBUV	VRATIO	PCI 95	CI 95
1	1.023	0.90	96.88	96.88
2	1.134	0.85	95.31	95.31
3	1.000	1.02	95.31	95.31
4	1.070	0.96	100.00	100.00
5	1.001	0.81	90.63	89.06
6	1.012	1.03	96.88	100.00
7	1.024	1.14	100.00	100.00
8	1.089	0.95	100.00	100.00
9	1.057	1.00	96.88	96.88
10	1.081	1.05	95.31	96.88
Sample mean	1.049	0.97	96.72	97.03
Sample S.D.	0.042	0.09	2.75	3.24

The excellent results in Table 1 were typical of all of the $n = 128$ combinations, all but two of the $n = 64$ combinations and some of the $n = 32$ combinations. That is, the efficiency ISUBV of the GCV estimate of is very close to 1, good estimates of σ^2 are obtained (VRATIO ≈ 1), and the CI 95's (as well as the PCI 95's) are quite close to 95. (Visual inspection of the replicates with CI 95 = 100 showed that the CI's were close to having CI 95 < 100.) Table 2a gives the sample mean of the 10 values of CI 95 for each of the 45 combinations of Cases, σ's and n's. All entries less than 85 have been circled. For comparison, Table 2b gives the sample mean of the 10 values of PCI 95 for each of the 30 combinations with $n = 64$ or 32. These numbers are seen to range from 91.56 to 99.06, so that the poor performance of the CI's in the circled combinations is directly attributable to poor estimates of σ^2.

TABLE 2

	(a) CI 95			(b) PCI 95	
	$n = 128$	$n = 64$	$n = 32$	$n = 64$	$n = 32$
$\sigma = 0.0125$					
Case 1	97.42	95.94	86.87	96.09	95.31
Case 2	96.88	80.31	31.56	96.56	92.50
Case 3	96.09	91.25	12.19	96.09	94.37
$\sigma = 0.025$					
Case 1	97.11	96.41	85.94	97.34	94.94
Case 2	97.42	72.66	67.19	95.47	95.31
Case 3	96.72	92.97	74.06	95.78	95.00
$\sigma = 0.05$					
Case 1	96.64	95.16	74.06	95.78	92.19
Case 2	96.56	94.06	82.81	97.66	93.44
Case 3	95.16	92.66	65.63	96.72	92.19
$\sigma = 0.1$					
Case 1	94.92	92.97	96.56	93.59	98.44
Case 2	95.55	97.03	92.19	96.72	95.44
Case 3	97.27	90.31	87.81	94.53	96.88
$\sigma = 0.2$					
Case 1	94.14	97.03	91.25	97.97	91.56
Case 2	95.86	98.91	90.00	99.06	93.75
Case 3	96.56	93.12	82.19	97.34	97.19

In the Case 2, $\sigma = 0.0125$, $n = 64$ combination, where the mean of ten CI 95 values was 80.31, two of the ten replications were identified as having a VRATIO $\leqslant 0.02$. These two poor estimates of σ^2 could be easily identified by an experimenter with even a crude knowledge of σ^2. Upon eliminating these two replications the average CI 95 over the remaining eight cases was 96.48. Similarly in the Case 2, $\sigma = 0.025, n = 64$ combination there were two replications with manifestly bad estimates of σ^2. (VRATIO < 0.01.) Upon eliminating these two cases the average CI 95 over the remaining cases was 90.82. For the $n = 32$ combinations, there were several cases of wildly erroneous CI 95's; however, they were almost uniformly detectable given a knowledge of σ^2 to a factor of 10. In many of these cases, EDF($\hat{\lambda}$) was small. Then one cannot really expect the method of estimation of σ^2 to work well, although it is nice to know that we can usually detect cases where it is awful. It can be seen from Table 2b that the PCI's are working well. A few examples of $n = 16$ were tried, and the results unsatisfactory. Neither GCV nor these CI's are recommended for $n = 16$.

Figs 4 and 5 are analogous to Fig. 1a, for one replicate of Case 4, $n = 64$, $\sigma = 0.1$, and Case 5, $n = 64$, $\sigma = 0.1$. Neither of these cases satisfy g, g' continuous, $\int (g''(t))^2 dt < \infty$ and hence, none of the theory developed here is necessarily applicable. For these two cases, only the four examples of all combinations of $n = 64$, $n = 128$, $\sigma = 0.05$ and $\sigma = 0.1$ were tried. For Case 4, the mean CI 95

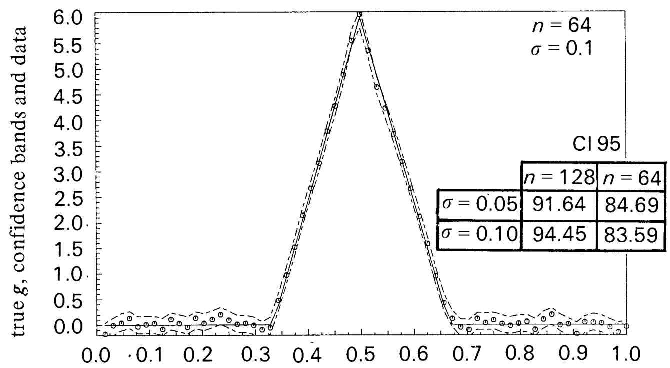

Fig. 4. $g(t)$, data and confidence bands. Case 4.

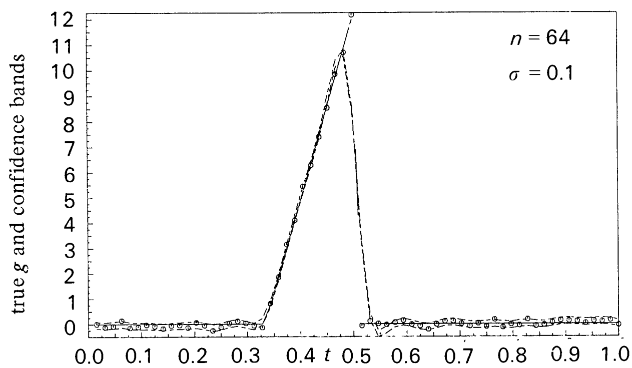

Fig. 5. $g(t)$, data and confidence bands. Case 5.

for each of the ten replications appears as an inset in Fig. 4. The average VRATIOS were about the same as for the Cases 1–3, but the variability is larger. The results are better than we expected but suggest caution in using the method on functions with discontinuous first derivative. The asymptotic result of Section 4 below does not apply to this case, and results could be worse for larger n.

Case 5 represents an attempt to defeat the method soundly. Ordinarily one would not attempt to estimate a function with a jump by a cubic spline, which has a continuous second derivative. Case 5 has a jump of 12 at $t = 0.5$. We ran the same four examples of n, σ combinations as we did for Example 4. Here σ was overestimated by a factor between around 30 and 300. Overshoot (or "Gibbs effect") in $g_{n,\hat{\lambda}}$ near $t = 0.55$ is clearly visible in Fig. 5. The two confidence intervals adjacent to the jump failed to cover the true value. In the $n = 64$ examples exactly 62 (= 96.88 per cent) and in the $n = 128$ case examples exactly 126 (98.44 per cent) true values of g were covered by the confidence intervals in each replication.

4. A BIVARIATE EXAMPLE

Finally, we give a bivariate example, kindly provided by J. Wendelberger, using the computer program developed in his thesis (Wendelberger, 1981, 1982). Fig. 6 depicts Franke's "Principal test function"

$$f(x, y) = 0.75 \exp\left[-\frac{(9x-2)^2 + (9y-2)^2}{4}\right]$$

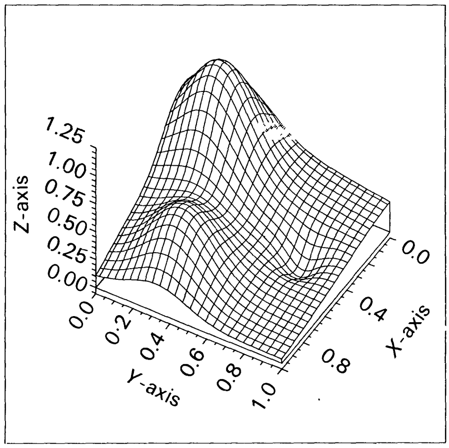

Fig. 6. Franke's principal test function.

$$+0.75\exp\left[-\frac{(9x+1)^2}{49}-\frac{9y+1}{10}\right]$$

$$+0.5\exp\left[-\frac{(9x-7)^2+(9y-3)^2}{4}\right]$$

$$-0.2\exp\left[-(9x-4)^2-(9y-7)^2\right].$$

which Franke (1979) used in an extensive comparison of different interpolation methods. Data were generated by the model

$$z_{ij}=f(x_i,y_j)+\epsilon_{ij}$$

with

$$x_i=\frac{2i+1}{2N},\quad y_j=\frac{2j+1}{2N},\quad i,j=1,2,\ldots,N,$$

with $N=13$, giving $n=N^2=169$ data points. The peak height of f was approximately 1.2 and σ was taken as 0.03. Here, f was estimated as the so called "thin plate smoothing spline" which is the minimizer (in an appropriate space) of

$$n^{-1}\sum_{i=1}^{n}(z_{ij}-f(x_i,y_j))^2+\lambda\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(f_{xx}^2+2f_{xy}^2+f_{yy}^2)\,dx\,dy.$$

It is not required in this method that a regular grid (x_i,y_j) be chosen. A regular grid was selected here so that we could plot cross-sections easily. Details of the theory, the cross-validation estimate of λ and a computational scheme are given in Wahba and Wendelberger (1980). An improved computational algorithm and further numerical results are given in Wendelberger (1981). Fig. 7 gives the resulting thin plate smoothing spline estimate of f. Fig. 8 gives four selected cross-sections for 4 fixed values of x, $x=(2i+1)/N$, for $i=7, 9, 11, 13$. In each cross-section is plotted $f((2i+1)/N,y)$, $0\leqslant y\leqslant 1$ (solid line), $f_{n,\hat{\lambda}}((2i+1)/N,y)$ $0\leqslant y\leqslant 1$, where $f_{n,\hat{\lambda}}$ is the thin plate smoothing spline (dashed line), the data z_{ij}, $j=1,2,\ldots,13$, for i fixed, and confidence bars, which extend between

$$f_{n,\hat{\lambda}}((2i+1)/N,y_j)\pm 1.96\hat{\sigma}(\hat{\lambda})\sqrt{(a_{ij,ij}(\hat{\lambda}))},$$

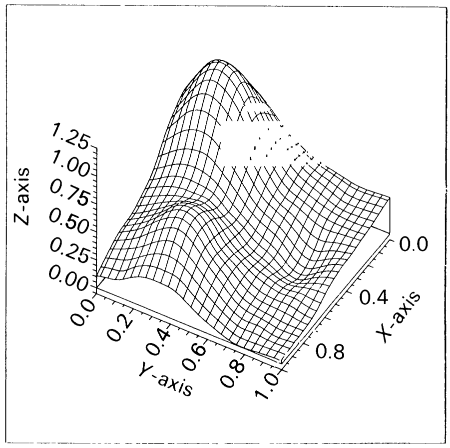

Fig. 7. Spline fit 169 points with sigma = 0.03.

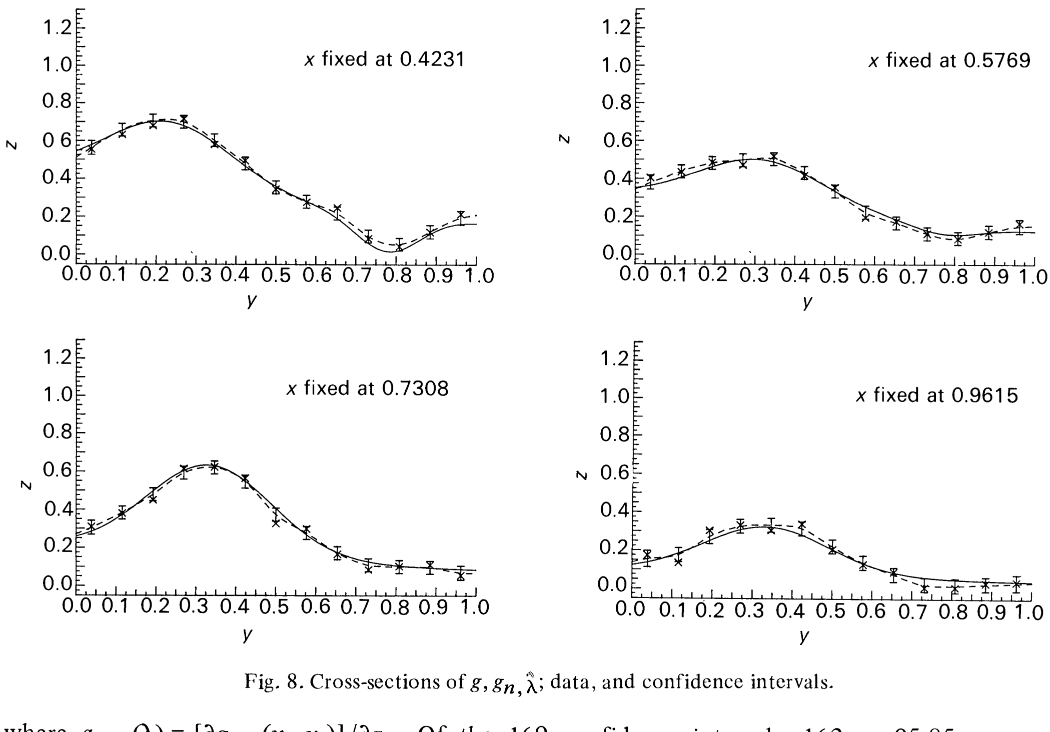

Fig. 8. Cross-sections of $g, g_{n,\hat{\lambda}}$; data, and confidence intervals.

where $a_{ij,ij}(\lambda) = [\partial g_{n,\lambda}(x_i, y_j)]/\partial z_{ij}$. Of the 169 confidence intervals, 162 or 95.85 per cent covered the corresponding true value of $f(x_i, y_j)$. This example was not "cooked" but was in fact the only example run by J. Wendelberger.

5. ON THE CHOICE OF m AND THE ESTIMATION OF b

So far we have been considering m fixed and given. Then $m = 2$ corresponds in some sense to a prior belief that $\int_0^1 (g''(t))^2 dt$ is small, and since this corresponds to mean square curvature, it may be thought of as corresponding to visual smoothness, or lack of visual "wiggliness". This might perhaps explain its popularity. Cubic splines tend to be aesthetically pleasing. From the point of view of optimizing predictive mean square error, m may be estimated by generalized cross-validation, see Gamber (1979b) and Wahba and Wendelberger (1980). In the latter paper, in some two-dimensional examples based on $f(x_i, y_i)$ = simulated 500 millibar height, where the $\{x_i, y_i\}$ are locations of the North American radiosonde network, an m of 5 was found to be optimum. In some meteorological applications on the sphere the so-called "penalty functional" $\int_0^1 (g^{(m)}(t))^2 dt$ can be replaced by a penalty functional defined on the sphere, which can be determined from prior (available) meteorological data on certain sample Fourier–Bessel coefficients. Due to lack of space we omit discussion of this approach here but details may be found in Wahba (1982). We remark, however, that in the case of periodic functions (on the circle or sphere) the highest derivative being assumed square integrable governs a bound on the rate of decay of the Fourier coefficients.

In order to use Theorem 2 to obtain the posterior variance of $\hat{\beta}$ and similar quantities, it is necessary to "know" or estimate b. In the case $\int_0^1 (g^{(m)}(t))^2 dt < \infty$ the appropriate meaning of b is not completely obvious. We conjecture that it should be $b = n^{-1} \int_0^1 (g^{(m)}(t))^2 dt$, where n is the sample size. We are presently studying the estimate $\hat{b}$ defined by

$$\hat{b}(\lambda) = \int_0^1 (g_{n,\lambda}^{(m)}(t))^2 dt / \mathrm{Tr} A(\lambda), \quad \hat{b} = \hat{b}(\hat{\lambda}).$$

If g is a random function with $b = \sigma^2/n\lambda$, it can be shown that $E\hat{b}(\lambda) = b$. Further work is needed, especially for use with the interesting application $\hat{\beta} = g'(t_0)$.

6. ON THE RELATION BETWEEN $ER(\lambda^*)$ AND $\frac{\sigma^2}{n} \sum_{i=1}^{n} a_{ii}(\lambda^*)$

The following theorem describes a relationship between $ER(\lambda^*)$ and $(\sigma^2/n) \Sigma_{i=1}^n a_{ii}(\lambda^*)$.

Theorem 3. Let g have $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$ and let $g_{n,\lambda}$ and $R(\lambda)$ be as in Section 1. Let $t_i = i/n$, $i = 1, 2, \ldots, n$. Let λ^* be the minimizer of $ER(\lambda)$. Then, as $n \to \infty$,

$$ER(\lambda^*) = \alpha \frac{\sigma^2}{n} \sum_{i=1}^{n} a_{ii}(\lambda^*)(1 + o(1)) \tag{6.1}$$

for some $\alpha \in [(1 + \frac{1}{4m})(1 - \frac{1}{2m}), 1]$.

Argument:

We must consider the case $g(\cdot) \in \pi_{m-1}$ (polynomials of degree $\leqslant m-1$), and $g \notin \pi_{m-1}$ separately. If $g \in \pi_{m-1}$, then $\lambda^* = \infty$, $A(\infty)$ is the orthogonal projection operator onto the discretized polynomials (columns of T defined in Section 2). $A(\infty) = A^2(\infty)$ and $ER(\infty) \equiv (\sigma^2/n) \mathrm{Tr} A(\infty)$, so (6.1) holds with $\alpha = 1$. Suppose $g \notin \pi_{m-1}$. We have

$$ER(\lambda) = b^2(\lambda) + \frac{\sigma^2}{n} \mathrm{Tr} A^2(\lambda),$$

where

$$b^2(\lambda) = n^{-1} \| (I - A(\lambda)) g \|^2.$$

Let $\{\gamma_\nu, nb_\nu\}$, $\nu = 1, 2, \ldots, n-m$, be the eigenvectors and eigenvalues of BQ_nB'. These are independent of the choice of B, but do depend on n, that is $\gamma_\nu \equiv \gamma_{\nu n}$, $b_\nu \equiv b_{\nu n}$. Let $g_\nu = (g, \gamma_\nu)$. Then

$$b^2(\lambda) = \sum_{\nu=1}^{n-m} \frac{\lambda^2 g_\nu^2}{(b_\nu + \lambda)^2},$$

$$n^{-1} \operatorname{Tr} A^2(\lambda) = n^{-1} \sum_{\nu=1}^{n-m} \left(\frac{b_\nu}{\lambda + b_\nu} \right)^2 = n^{-1} \sum_{\nu=1}^{n-m} \frac{1}{(1 + \lambda/b_\nu)^2},$$

$$n^{-1} \sum_{i=1}^{n} a_{ii}(\lambda) = n^{-1} \operatorname{Tr} A(\lambda) = n^{-1} \sum_{\nu=1}^{n-m} \frac{1}{(1 + \lambda/b_\nu)}.$$

From Craven and Wahba (1979), and Utreras (1979b, 1980, 1981), it follows, (for $t_i = i/n$), that there exists some C such that

$$n^{-1} \operatorname{Tr} A^2(\lambda) \approx \frac{C l_m}{n\lambda^{1/2m}}, \quad n^{-1} \operatorname{Tr} A(\lambda) \approx \frac{C \tilde{l}_m}{n\lambda^{1/2m}},$$

provided $n\lambda^{1/2m} \to \infty$, where

$$l_m = \int_0^\infty \frac{dx}{(1 + x^{2m})^2}, \quad \tilde{l}_m = \int_0^\infty \frac{dx}{(1 + x^{2m})}.$$

It is shown in Craven and Wahba (1979) that

$$b^2(\lambda) \leqslant \lambda J_m(g), \tag{6.2}$$

where

$$J_m(g) = \int_0^1 (g^{(m)}(t))^2 \, dt.$$

(Inequality (6.2) actually holds for all the seminorms considered in Section 3 of Wahba (1978), whether or not the $\{t_i\}$ are equally spaced.) More generally, if

$$\sum_{\nu=1}^{n-m} \frac{g_\nu^2}{b_\nu^2} \leqslant J_{mp}(g),$$

where J_{mp} is independent of n, and $1 \leqslant p \leqslant 2$, then clearly

$$b^2(\lambda) \leqslant \lambda^p J_{mp}(g).$$

For $p = 2$, if $g \notin \pi_{m-1}$ then it can be seen that

$$\sum_{\nu=1}^{n-m} \frac{g_\nu^2}{b_\nu^2} \leqslant J_{2m}(g) \tag{6.3}$$

entails that

$$b^2(\lambda) = \lambda^2 J_{2m}(g) \, (1 + o(1))$$

as $\lambda \to 0$, $n \to \infty$. It appears that if the $\{t_i\}$ are approximately equally spaced, then it is sufficient for (6.3) that g has a representation of the form

$$g(t) = \int_0^1 Q(t,s)\rho(s)ds + \sum_{\nu=1}^{m-1} \theta_\nu \phi_\nu(s),$$

where $\rho(= g^{(2m)})$ is some sufficiently regular function. See Wahba (1979) for a heuristic argument in the thin plate spline case, also Wahba (1977a).

Letting $J_{mp}(g) = g_p$, we have

$$R(\lambda) \leqslant \left(\lambda^p g_p + \frac{\sigma^2 Cl_m}{n\lambda^{1/2m}} \right) (1 + o(1)) \tag{6.4}$$

as $\lambda \to 0$, $n\lambda^{1/2m} \to \infty$, for each $p \in [1, 2]$ for which g_p is finite, with equality for $p = 2$, if $g < \infty$. The minimizer of the right-hand side of (6.4) is

$$\lambda_p^* = \left(\frac{\sigma^2}{g_p} \frac{Cl_m}{2mp} n^{-1} \right)^{2m/(2mp+1)}$$

and, letting $\theta = 2mp/(2mp + 1)$ gives

$$R(\lambda_p^*) \leqslant \left(\frac{\sigma^2 Cl_m}{2mpn} \right)^\theta g_p^{1-\theta} (2mp + 1) \tag{6.5}$$

$$\frac{\sigma^2}{n} \operatorname{Tr} A(\lambda_p^*) \simeq \left(\frac{\sigma^2 Cl_m}{2mpn} \right)^\theta g_p^{(1-\theta)} (2mp) \left(\frac{2m}{2m-1} \right),$$

where we have used $\tilde{l}_m / l_m = (2m)/(2m - 1)$. Arguing heuristically that equality in (6.5) must hold for some p between 1 and 2 gives

$$\frac{R(\lambda_p^*)}{\frac{\sigma^2}{n} \operatorname{Tr} A(\lambda_p^*)} \simeq \left(\frac{2mp+1}{2mp} \frac{2m-1}{2m} \right) (1 + o(1))$$

for some $p \in [1, 2]$. Since this quantity is between

$$\left(1 + \frac{1}{4m}\right)\left(1 - \frac{1}{2m}\right) \quad \text{and} \quad \left(1 + \frac{1}{2m}\right)\left(1 - \frac{1}{2m}\right),$$

the result follows.

We remark that this argument can be repeated in the general context of Wahba (1978) whenever the rate of decay of the eigenvalues $\{b_\nu\}$ and the generalized fourier coefficients $\{g_\nu\}$ are known and some summability conditions on the $\{b_{\nu n}\}$ are satisfied. Thus by using the conjectures concerning the $\{b_\nu\}$ in Wahba (1979), our arguments can no doubt extend to the thin plate spline. See Wahba (1977b).

ACKNOWLEDGEMENTS

We wish to thank Bill Wecker for generously providing an early version of the Wecker and Ansley manuscript and rekindling our interest in this topic following a talk he gave at Madison in autumn 1980. We thank Jim Wendelberger for providing the thin plate spline example of Section 3, and, last but not least, we thank Chris Sheridan, an undergraduate at Madison who very capably wrote the computer program for the Monte Carlo study.

This work was supported by the ONR under Contract No. N00014-77-C-0675.

REFERENCES

Agarwal, G. G. and Studden, W. J. (1980) Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* 8, 1307–1325.

Anderson, T. W. (1958) *An Introduction to Multivariate Statistical Analysis*, p. 28. New York: Wiley.

Berger, J. (1980) A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.*, 8, 716–761.

Craven P. and Wahba, G. (1979) Smoothing noisy data with spline functions. *Numer. Math.*, **31**, 377–403.

Franke, R. (1979) A critical comparison of some methods for interpolation of scattered data, Naval Postgraduate School Report No. NPS-53-79-003, March 1979.

Gamber, H. A. (1979a) Confidence regions for periodic functions. *Commun. Statist.* **A8**, 1437–1446.

——— (1979b) Choice of an optimal shape parameter when smoothing noisy data. *Commun. Statist.*, **A8**, 1425–1436.

Gasser, T. and Rosenblatt, M. (1979) (eds) *Smoothing Techniques for Curve Estimation*, (Lecture Notes in Mathematics, No. 757). Springer-Verlag, Berlin.

IMSL (1980) International Mathematical and Statistical Libraries, Inc. Edition 8, Subroutine ICSSCV, Houston, TX.

Kimeldorf, G. and Wahba, G. (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, **41**, 495–502.

——— (1971) Some results on Tchebycheffian spline functions. *J. Math. Anal. Applic.*, **33**, 82–95.

Knafl, G., Sacks, J. and Ylvisaker, D. (1982) Model robust confidence intervals, II. in "Statistical Decision Theory and Related Topics", S. S. Gupta and J. O. Berger, eds. Academic Press, New York, 87–102.

Lucas, H. A. (1978) Estimation of smoothing parameters to smooth noisy data and confidence regions for the underlying function, Ph.D. Thesis, University of Wisconsin, Madison. (H. A. Lucas is now H. A. Gamber.)

Merz, P. M. (1978) Spline smoothing by generalized cross-validation, a technique for data smoothing, Chevron Research Co., Richmond, CA.

Nogues, A. and Sielken, R. L. Jr (1980) A comparative study of spline regression. Texas A & M University Project Themes TR No. 67.

Rice J. and Rosenblatt, M. (1981) Integrated mean square error of a smoothing spline. *J. Approx. Theory* **33**, 353–369.

Stone, C. (1980) Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8, 1348–1360.

Utreras, F. (1979a) Cross-validation techniques for smoothing spline functions in one or two dimensions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds). (Lecture Notes in Mathematics No. 757.) Springer-Verlag, Berlin.

——— (1979b) Natural spline functions, its associated eigenvalue problem. Departmento de Matematicas, Universidad de Chile, IT No. MA-80-B-208.

——— (1980) On the eigenvalue problem associated with cubic splines: The arbitrarily spaced knots case. Universidad de Chile, Departamento de Mathematicas manuscript.

——— (1981) Optimal smoothing of noisy data using spline functions. *SIAM J. Sci. Stat. Comput.* **2**, **3**, 349–362.

Wahba, G. (1977a) Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numerical Analysis*, **14**, 651–667.

——— (1977b) Invited discussion to "Consistent nonparametric regression by C. J. Stone. *Ann. Statist.*, **5**, 637–645.

——— (1978) Improper priors, spline smoothing, and the problem of guarding against model errors in regression. *J. R. Statist. Soc.* B, **49**, 364–372.

——— (1979) Convergence rates of "Thin Plate" smoothing splines when the data are noisy. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds). Lecture Notes in Mathematics No. 757. Springer-Verlag.

——— (1981a) Spline interpolation and smoothing on the sphere. *SIAM J. Scientific and Statistical Computing* **2**, 5–16.

——— (1981b) Some new techniques for variational objective analysis using splines, Hough functions, and sample spectral data. Preprint volume of "Seventh Conference on Probability and Statistics in Atmospheric Sciences, Nov. 2–6, 1981". Boston: American Meteorological Society.

——— (1981b) Bayesian confidence intervals for the cross validated smoothing spline. University of Wisconsin–Madison, Department of Statistics, TR 645.

——— (1982). Vector splines on the sphere, with application to the theory estimation of vorticity and divergence from discrete, noisy data. in "Multivariate Approximation Theory II, W. Schempp and K. Zeller, eds. Birkhauser Verlag, Basel, 407–429.

Wahba, G. and Wendelberger, J. (1980) Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Rev.*, **108**, 36–57.

Wecker, W. P. and Ansley, C. F. (1980) Linear and nonlinear regression reviewed as a signal extraction problem. The University of Chicago Graduate School of Business, manuscript.

Wegman, E. and Wright, I. (1980) Splines in statistics, to appear. *J. Amer. Statist. Assoc.*

Wendelberger, J. (1981) The computation of Laplacian smoothing splines with examples. University of Wisconsin, Madison, Department of Statistics, TR 648.
——— (1982) Smoothing Noisy Data With Multidimensional Splines and Generalized Cross Validation. Ph. D. Thesis, Department of Statistics, University of Wisconsin, Madison.