WSD 0645

WSD 0645

DEPARTMENT OF STATISTICS

University of Wisconsin
1210 W. Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 645

July 1981

BAYESIAN CONFIDENCE INTERVALS
FOR THE CROSS VALIDATED
SMOOTHING SPLINE

by

Grace Wahba

This work was supported by the USARO under Grant No. DAAG29-80-K0042.

ABSTRACT

We consider the model $Y(t_i) = g(t_i) + \varepsilon_i$, $i = 1,2,\ldots,n$, where $g(t)$, $t\varepsilon[0,1]$ is a smooth function and the $\{\varepsilon_i\}$ are independent $N(0,\sigma^2)$ errors with σ^2 unknown. The cross validated smoothing spline will be used to estimate g nonparametrically 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. First, properties of smoothing splines as Bayes estimates are used to derive confidence intervals based on the posterior covariance function of the estimate. To compute the confidence intervals it is necessary to know or to estimate σ^2. We estimate σ^2 here by the residual sum of squares divided by the equivalent degrees of freedom, both of which are determined using the generalized cross validation estimate of the smoothing parameter. A Monte Carlo study is carried out to suggest by example to what extent the resulting 95% confidence intervals can be expected to cover about 95% of the true (but in practice unknown) values of $g(t_i)$, $i = 1,2,\ldots,n$. Three smooth example functions, 5 values of σ^2, and $n = 32$, 64 and 128 were tried. Confidence intervals based on known σ^2 were extremely reliable for all 3 n's, generally covering close to 95% of the true $\{g(t_i)\}$. Confidence intervals based on estimated σ^2's were also highly reliable for all $n = 128$ and most $n = 64$ examples tried. Degraded results were sometimes seen for $n = 32$. Failure of the method for small n appears to be accompanied by estimates of σ^2 off by orders of magnitude. which would frequently be evident to an experimenter. The method was also applied to one example of a two dimensional thin plate smoothing spline with $n = 169$, and 162 or 95.8% of the true values were covered by the 95% confidence intervals. 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.

Key words: Spline smoothing, cross-validation, confidence intervals

1. INTRODUCTION

Consider the model

$$Y(t_i) = g(t_i) + \varepsilon_i, \; i = 1,2,\ldots,n, \; t_i \varepsilon [0,1] \tag{1.1}$$

where $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)' \sim N(0,\sigma^2 I_{n\times n})$, σ^2 is unknown and $g(\cdot)$ is a fixed but unknown function with m-1 continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$. (Equivalently, g is in the Sobolev Hilbert space known as W_2^m, see Adams (1975).) 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 in W_2^m of

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

The parameter λ controls the tradeoff between the infidelity $\frac{1}{n}\sum_{i=1}^{n} (g_{n,\lambda}(t_i)-y_i)^2$ and the roughness $\int_0^1 (g_{n,\lambda}^{(m)}(t))^2 dt$ of the solution.

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 Gaussian prior (as opposed to assuming $g\varepsilon W_2^m$). This property of smoothing spline estimates was discussed in some detail in Wahba (1978), hereinafter referred to as W. It is the purpose of this paper, which is a sequel to W, to use the properties of $g_{n,\lambda}$ as a Bayes estimate to derive confidence intervals about the estimate, based on the posterior variances of the $g_{n,\lambda}(t_i)$, to provide a Monte Carlo study demonstrating the effectiveness of the resulting confidence intervals for several examples of g when g is a fixed function in W_2^m, and to provide a rough analytical argument why these confidence intervals should "work", as well as they appear to in the Monte Carlo study. Some type of analytical argument is needed, because (as we shall see) under the prior for which $g_{n,\lambda}$ is a posterior mean,

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

We note that Gamber (1979a, Lucas (1978)) and Wecker and Ansley (1980) have previously observed that the prior distribution for which the posterior mean is a smoothing spline, can be used to obtain confidence intervals via the posterior covariance function. The prior under discussion is: $g(t), t\varepsilon[0,1]$ has the same distribution as

$$X_\xi(t) = \sum_{j=1}^{m} \theta_j \phi_j(t) + b^{1/2} Z(t),\ t\varepsilon[0,1], \tag{1.2}$$

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

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

$W(u)$ being the Weiner process, and $\xi \to \infty$. It is proved in W (see also Kimeldorf and Wahba (1971)), that, upon setting $\lambda = \sigma^2/nb$, the smoothing spline satisfies

$$g_{n,\sigma^2/nb}(t) = \lim_{\xi\to\infty} E_\xi\{g(t)|Y=y\},$$

where $Y = (Y(t_1),\ldots,Y(t_n))'$, $y = (y_1,\ldots,y_n)'$, E_ξ is expectation over the posterior distribution of $g(t)$ with the prior (1.2) ($\xi\to\infty$ corresponds to a "partially improper" prior). We will, in this study, treat λ as an infidelity-roughness control parameter rather as the process parameter defined by $n\lambda = \sigma^2/b$.

That the fixed function and sample function models are different and that this affects the meaning of λ, can be quickly illustrated in the m = 1 case. In the fixed function case ($g\varepsilon W_2^1$)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [n(g(\frac{i+1}{n})-g(\frac{i}{n}))]^2 = \int_0^1 (g'(t))^2 dt \tag{1.3}$$

whereas in the random function (or Bayesian) case it is easy to show that

$$E \frac{1}{n} \sum_{i=1}^{n} [n(g(\frac{i+1}{n})-g(\frac{i}{n}))]^2 = nb \tag{1.4}$$

so that by any reasonable limiting argument we cannot have $E\int_0^1 (g'(t))^2 dt < \infty$.

The comparison of (1.3) and (1.4) can be shown to extend to m = 2,3,... by replacing first divided differences by m^{th} divided differences, and a similar comparison can be made in the general case of Section 3 of W.

In the random function case, λ can be estimated by the method of maximum likelihood, first suggested by Anderssen and Bloomfield (1974), see also Wahba (1977a). In the fixed function case, we estimate λ by estimating the minimizer of some suitable loss function, $R(\lambda)$, for example

$$R(\lambda) = \frac{1}{n} \sum_{i=1}^{n} (g(t_i)-g_{n,\lambda}(t_i))^2, \tag{1.5}$$

the predictive mean square error. In this paper we will be using the generalized cross validation (GCV) estimate of λ from Craven and Wahba (1979) for our studies. This estimate is an estimate of the minimizer of $R(\lambda)$, and does not require knowledge of σ^2. If σ^2 is known λ can be estimated by the method of unbiassed risk estimation, see Craven and Wahba (1979). Code is available in IMSL (1980) for computing $g_{n,\lambda}$ for

$m = 2$ and estimating λ by GCV. For $g \varepsilon W_2^m$, the predictive mean square error with optimum λ converges rapidly,

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

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

Our approach in this work is as follows:

1. Use the model (1.2) to derive the posterior covariance matrix for $\underset{\sim}{g}_{n,\lambda} = (g_{n,\lambda}(t_1),\ldots,g_{n,\lambda}(t_n))'$, for fixed λ and σ^2. Given λ and σ^2 this results in confidence intervals for each $g(t_i)$, centered at $g_{n,\lambda}(t_i)$.
2. Use the data to estimate the (optimum m.s.e.) λ by GCV, and σ^2 by $\hat{\sigma}^2(\hat{\lambda})$ where

   $$\hat{\sigma}^2(\hat{\lambda}) = RSS(\hat{\lambda})/EDF(\hat{\lambda}),$$

   RSS is the residual sum of squares, EDF is the equivalent degrees of freedom for error when $\hat{\lambda}$ is used, and $\hat{\lambda}$ is the GCV estimate of λ.
3. Run a set of Monte Carlo experiments with several fixed function models, several sample sizes and several values of σ to collect some evidence as to whether these confidence intervals can be expected to have useful properties over some range of practical situations.

4. Obtain an asymptotic argument why the 95% confidence intervals obtained as in 1. and 2. should cover about 95% of the true $\{g(t_i)\}$ values, when $g\varepsilon W_2^m$ and λ is a good estimate of the minimizer of $R(\lambda)$.

In the Monte Carlo experiments we have taken three different smooth functions with $\int_0^1 |g(t)|dt = 1$ and $\sigma = .0125$, $.025$, $.05$, $.1$ and $.2$. We considered only $t_i = i/n$, and $n = 128$, 64 and 32. Thus there were $3 \times 5 \times 3 = 45$ examples of all combinations of functions, σ's and n's. Data was generated for 10 replicates of each example via the model (1.1). Confidence intervals were determined as in 1. and 2. (details are given in Section 2) and the percentage of true $g(t_i)$ covered by the confidence intervals was recorded. To determine sensitivity to the estimate $\hat{\sigma}^2(\hat{\lambda})$ of σ^2, the confidence intervals derived in 1. with λ estimated by GCV but with $\hat{\sigma}^2(\hat{\lambda})$ replaced by the true σ^2, were determined, and the coverage percentages also recorded. We will call these "pseudo confidence intervals", since it is necessary to know σ^2 to compute them. For the 45 examples, the percent of true $g(t_i)$ covered by the pseudo confidence intervals in the 10 replications varied from a very satisfactory 92% to 98% with a small within group variance. Nearly identical favorable results were obtained using $\hat{\sigma}^2(\hat{\lambda})$ instead of σ^2 for all of the $n = 128$ examples, for nearly all of the $n = 64$ and some of the $n = 32$ examples. For the smaller n and $\sigma = .0125$ and $.025$, there were occasional unreliable estimates for σ^2, they tend to be too small as $g_{n,\lambda}$ came very close to interpolating the data. It is interesting to note in the results, however, that if one has "some idea" about the size of σ^2, then the exceptional (unreliable) cases can be readily spotted since the poor estimates for σ^2 tended to be around two orders of magnitude too low. That is, if $\hat{\sigma}^2(\hat{\lambda})$ was bad it was typically bad enough to be spotted as nonsense. Thus, for moderately large sample sizes, (>32) if g is known a priori to be "smooth" and some rough idea of the size of σ^2 is known, it appears that the proposed confidence intervals can be used in practice "with confidence".

The analytical result obtained in 4. is as follows. Let $s_{ii}(\lambda)$ be the posterior variance of $g_{n,\lambda}$, derived using the random function model. Then, if $g \varepsilon W_2^m$, n large and λ^* the minimizer of $ER(\lambda)$, we argue that, for large n

$$ER(\lambda^*) = \frac{1}{n}\alpha \sum_{i=1}^{n} s_{ii}(\lambda^*)(1+o(1)) \tag{1.6}$$

where α is some number between $(1 + \frac{1}{4m})(1 - \frac{1}{2m})$ and 1.

The expression (1.6) says, that if $g \varepsilon W_2^m$, then asymptotically, the average square bias plus variance is approximately equal to the expression for the average of the posterior variances which are used in the confidence intervals, provided λ is taken as the minimizer of $ER(\lambda)$.

We note that the theoretical confidence interval results here extend immediately to the generalized splines discussed in Section 3 of W. For applications to splines on the plane and the sphere see Wahba (1981) Wahba and Wendelberger (1980), and Wendelberger (1981). Following the Monte Carlo study just described, 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%

of the true functional values. This example was calculated using some improved numerical algorithms being developed for cross validated splines in several dimensions by J. Wendelberger (1981).

Nogues and Sielken (1980) have proposed a jacknife technique to obtain confidence intervals for the cross validated smoothing spline estimate. Numerical results were not presented by them but we believe that it would be very interesting to compare numerically the intervals they propose with the ones given here. Knafl, Sacks and Ylvisaker (1981) have also recently proposed confidence intervals for some nonparametric estimates which are sometimes related to smoothing splines. For more on nonparametric regression, see, for example, Agarwal and Studden (1980), Gasser and Rosenblatt (1979), Knafl, Sacks and Ylvisaker (1981), Nogues and Sielken (1980), Stone (1980), W, and references cited there. We remark that our philosophy is is the spirit of one suggested by Berger (1980), that 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

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 jk^{th} 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 influence matrix $A(\lambda)$ defined by

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

where $\underset{\sim}{g}_{n,\lambda} = (g_{n,\lambda}(t_1),\ldots,g_{n,\lambda}(t_n))'$. In the Bayes model discussed here we will substitute $n\lambda$ for σ^2/b, until further notice. A rather involved formula for $A(\lambda)$ is given in W (p. 367), but it can easily be seen, by substituting (4.2) of W into (4.4) of W 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 ij^{th} entry $a_{ij}(\lambda)$ of $A(\lambda)$ satisfies

$$a_{ij}(\lambda) = \frac{\partial g_{n,\lambda}(t_i)}{\partial y_j} \qquad (2.1)$$

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

$$\operatorname{cov}(\underset{\sim}{g}_{n,\lambda} | Y(t_1),\ldots,Y(t_n)) = \sigma^2 A(\lambda), (\lambda = \sigma^2/nb). \qquad (2.2)$$

The proof will be given later. The diagonal entries $\sigma^2 a_{ii}(\lambda)$, $i = 1,2,\ldots,n$, of the posterior covariance matrix $\sigma^2 A(\lambda)$ are used to define (individual) 95% (posterior) confidence intervals for $g(t_i)$ by

$$g_{n,\lambda}(t_i) \pm 1.96\sigma\sqrt{a_{ii}(\lambda)}. \tag{2.3}$$

These are the intervals that we will study, with $\lambda = \hat{\lambda}$, the GCV estimate of λ. We remark that the formulae (2.1 and 2.2) as well as the results below apply to the generalized splines in Section 3 of W, to thin plate splines on the plane (see Wahba and Wendelberger (1980)) and on the sphere (see Wahba (1981)). Formula (5) of Gamber (1979a) can be shown to be a special case of (2.2). For completeness we will give the complete posterior covariance function of $g_{n,\lambda}(t)$, $t\varepsilon[0,1]$, although we will not use it further in this paper.

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

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

where

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

and

$$\mathrm{Cov}\{(g(s),g(t))|g(t_1),\ldots,g(t_n)\} = b\Big\{(\phi_1(s),\ldots,\phi_m(s))\theta^{-1}\begin{pmatrix}\phi_1(t)\\ \vdots \\ \phi_m(t)\end{pmatrix} + Q(s,t)$$

$$- (Q(s,t_1),\ldots,Q(s,t_n))Q_n^{-1}T\theta^{-1}\begin{pmatrix}\phi_1(t)\\ \vdots \\ \phi_m(t)\end{pmatrix}$$

$$- (Q(t,t_1),\ldots,Q(t,t_n))Q_n^{-1}T\theta^{-1}\begin{pmatrix}\phi_1(s)\\ \vdots \\ \phi_m(s)\end{pmatrix}$$

$$- (Q(s,t_1),\ldots,Q(s,t_n))P_n\begin{pmatrix}Q(t,t_1)\\ \vdots \\ Q(t,t_n)\end{pmatrix}\Big\} \quad (2.4)$$

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

The proof of Theorems 1 and 2 proceed via

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

$$y = g + \varepsilon,$$

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

$$E(h|y) = \textstyle\sum_{hg}(\sum_{gg}+n\lambda I)^{-1}y \quad (2.5)$$

$$\mathrm{cov}(hh'|y) = b(\textstyle\sum_{hh}-\sum_{hg}\sum_{gg}^{-1}\sum_{gh}) + \sigma^2\sum_{hg}\sum_{gg}^{-1}A(\lambda)\sum_{gg}^{-1}\sum_{gh} \;. \quad (2.6)$$

In particular, setting h = g gives

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

$$\mathrm{cov}(g|y) = \sigma^2 A(\lambda) \tag{2.8}$$

Proof of Lemma. The proof follows by application of Anderson (1958) and tedious but straightforward algebra.

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

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

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

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

To complete the proof of Theorem 2, $\sum_{hh}$, $\sum_{hg}$ and $\sum_{gg}$ 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 W (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 has learned that the anonymous referee who provided (2.12)-(2.14) was B. Silverman.

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

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

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

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

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

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

where

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

All three cases are in W_2^m for $m = 2$.

In addition two cases were chosen with discontinuities in the function or its first derivative, and one two dimensional example is given.

An optimal m can be estimated from the data by cross validation. (See Gamber (1979b), Wahba and Wendelberger (1980)) but we chose for simplicity not to do that here. In all cases we fixed $m = 2$.

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 subspace of W_2^2 spanned by

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

In the Monte Carlo studies we let $t_i = i/n$, $i = 1,2,\ldots,n$, and generated data $y = (y_1,\ldots,y_n)'$ by

$$y_i = g(\tfrac{i}{n}) + \varepsilon_i \qquad i = 1,2,\ldots,n$$

where the ε_i were pseudorandom normal deviates with mean 0 and variance σ^2. Given y, the minimizer in F_n of

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

is

$$g_{n,\lambda}(t) = a_o + 2\sum_{\nu=1}^{n/2-1}\frac{a_\nu \cos 2\pi\nu t + b_\nu \sin 2\pi\nu t}{[1+\lambda(2\pi\nu)^{2m}]} + \frac{a_{n/2}\cos\pi nt}{[1+\lambda(\pi n)^{2m}]} \qquad (3.1)$$

where

$$a_o = \frac{1}{n}\sum_{j=1}^{n} y_j$$

$$a_\nu = \frac{1}{n}\sum_{j=1}^{n}(\cos 2\pi\nu\tfrac{j}{n})y_j, \qquad 1,\ldots,n/2$$

$$b_\nu = \frac{1}{n}\sum_{j=1}^{n}(\sin 2\pi\nu\tfrac{j}{n})y_j, \qquad \nu = 1,2,\ldots,n/2-1$$

This is the estimate of g that is being used.

It is not hard to show that

$$||(I-A(\lambda))y||^2 \equiv RSS(\lambda) = 2\sum_{\nu=1}^{n/2-1}\left(\frac{\lambda}{\lambda_\nu+\lambda}\right)^2(a_\nu^2+b_\nu^2) + \left(\frac{\lambda}{\lambda_{n/2}+\lambda}\right)^2 a_{n/2}^2, \tag{3.2}$$

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

where $\lambda_\nu = (2\pi\nu)^{-2m}$. $V(\lambda)$ of (2.16) is computed from

$$V(\lambda) = \frac{\frac{1}{n}RSS(\lambda)}{(1-a(\lambda))^2}, \tag{3.3}$$

and $\hat{\lambda}$ is the minimizer of (3.3).

$$\hat{\sigma}^2(\lambda) = \frac{RSS(\lambda)}{n-p(\lambda)}, \quad p(\lambda) = na(\lambda). \tag{3.4}$$

It is reasonable to consider $p(\lambda)$ as the equivalent degrees of freedom for signal and $n-p(\lambda)$ the EDF for error. The estimated 95% confidence intervals are given by

$$g_{n,\hat{\lambda}}\left(\frac{i}{n}\right) \pm 1.96\hat{\sigma}(\hat{\lambda})\sqrt{a(\hat{\lambda})}$$

$\hat{\lambda}$ is found by a global search based on equally spaced increments of $\log\lambda$. There is a conceptual question whether 1.96 or the .025 point of the t distribution with $EDF(\hat{\lambda})$ degrees of freedom should be used. For the $n = 128$ and $n = 64$ examples the $EDF(\hat{\lambda})$ was typically greater than 30 and $t_{.025}(EDF(\hat{\lambda})) \approx 1.96$. For $n = 32$, the use of $t_{.025}(EDF(\hat{\lambda}))$ instead of

1.96 would most likely have improved the confidence intervals obtained here somewhat. This point will be discussed further. We also examined properties of the 95% pseudo confidence intervals given by

$$g_{n,\hat{\lambda}}(\tfrac{i}{n}) \pm 1.96\sigma\sqrt{a(\hat{\lambda})} .$$

Here σ is the standard deviation of the $\{\varepsilon_i\}$ that is input to the Monte Carlo study.

We will first describe three examples, and then give a summary of the Monte Carlo experiment. Figure 1 gives a plot of g(t) for case 1, (continuous line), n = 128 simulated data points with $\sigma = .1$ (circles) and $g_{n,\hat{\lambda}}(t)$ (dashed line). Figure 2 gives a plot of g(t) and the same data as Figure 1. For visual effect confidence "bands" have been plotted in Figure 2, the upper and lower dashed lines are

$$g_{n,\hat{\lambda}}(t) + 1.96\hat{\sigma}(\hat{\lambda})\sqrt{a(\hat{\lambda})}$$

and

$$g_{n,\hat{\lambda}}(t) - 1.96\hat{\sigma}(\hat{\lambda})\sqrt{a(\hat{\lambda})}. \tag{3.5}$$

Strictly speaking in the absence of knowledge of b or a bound on $\int(g^{(m)}(t))^2dt$ these bands only have meaning at $t = \frac{i}{n}$, $i = 1,2,\ldots,n$. Considering the n = 128 intervals centered at $g_{n,\hat{\lambda}}(\frac{i}{n})$, in this example 100% of them covered the true values of $g(t_i)$. It can be seen that the bands are not unduly large, however. Figure 3 gives a plot of $V(\lambda)$ and the predictive mean square error $R(\lambda)$ defined in (1.5).

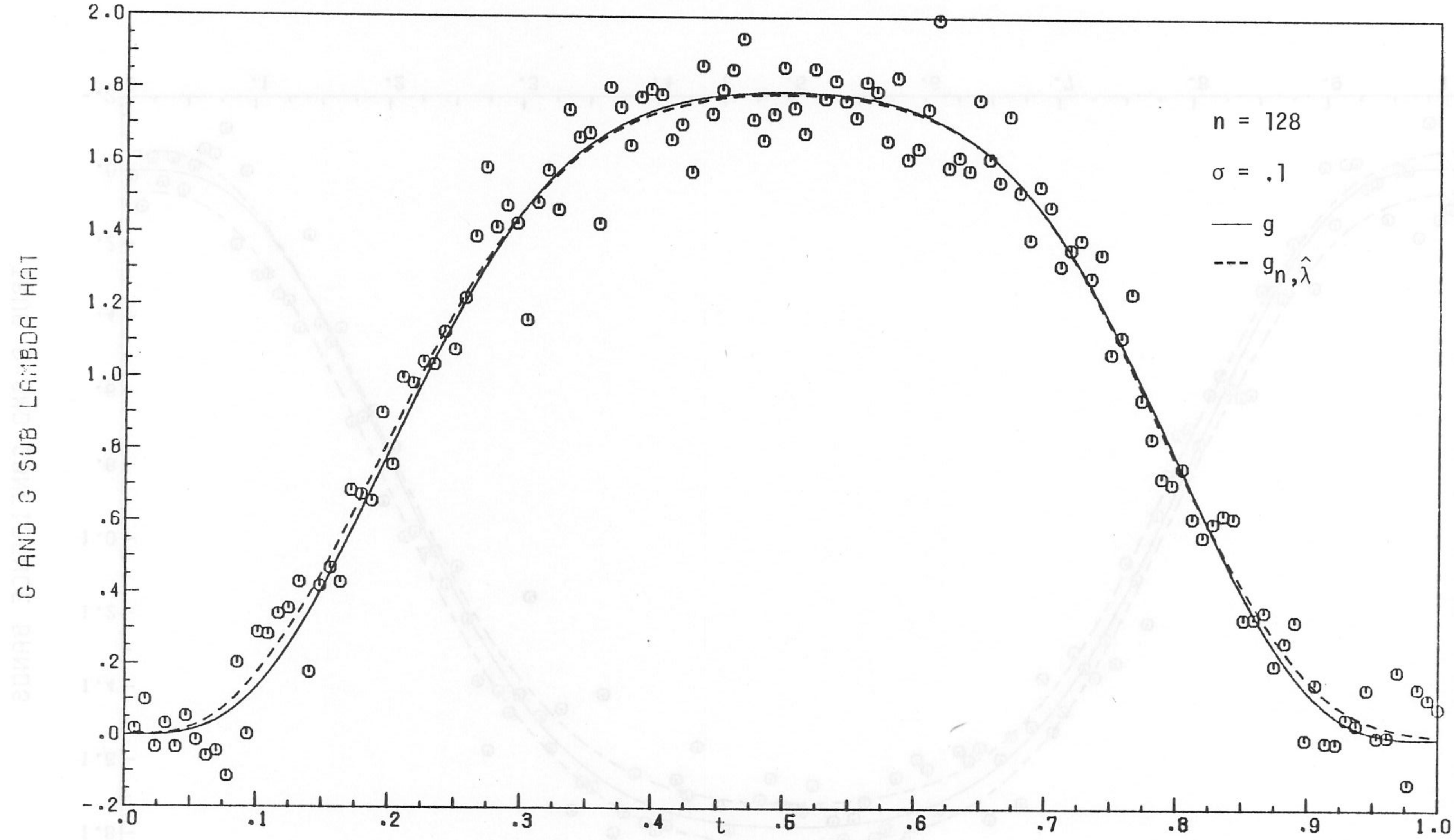

Figure 1. $g(t)$, simulated data, and $g_{n,\hat{\lambda}}(t)$ for an example of Case 1.

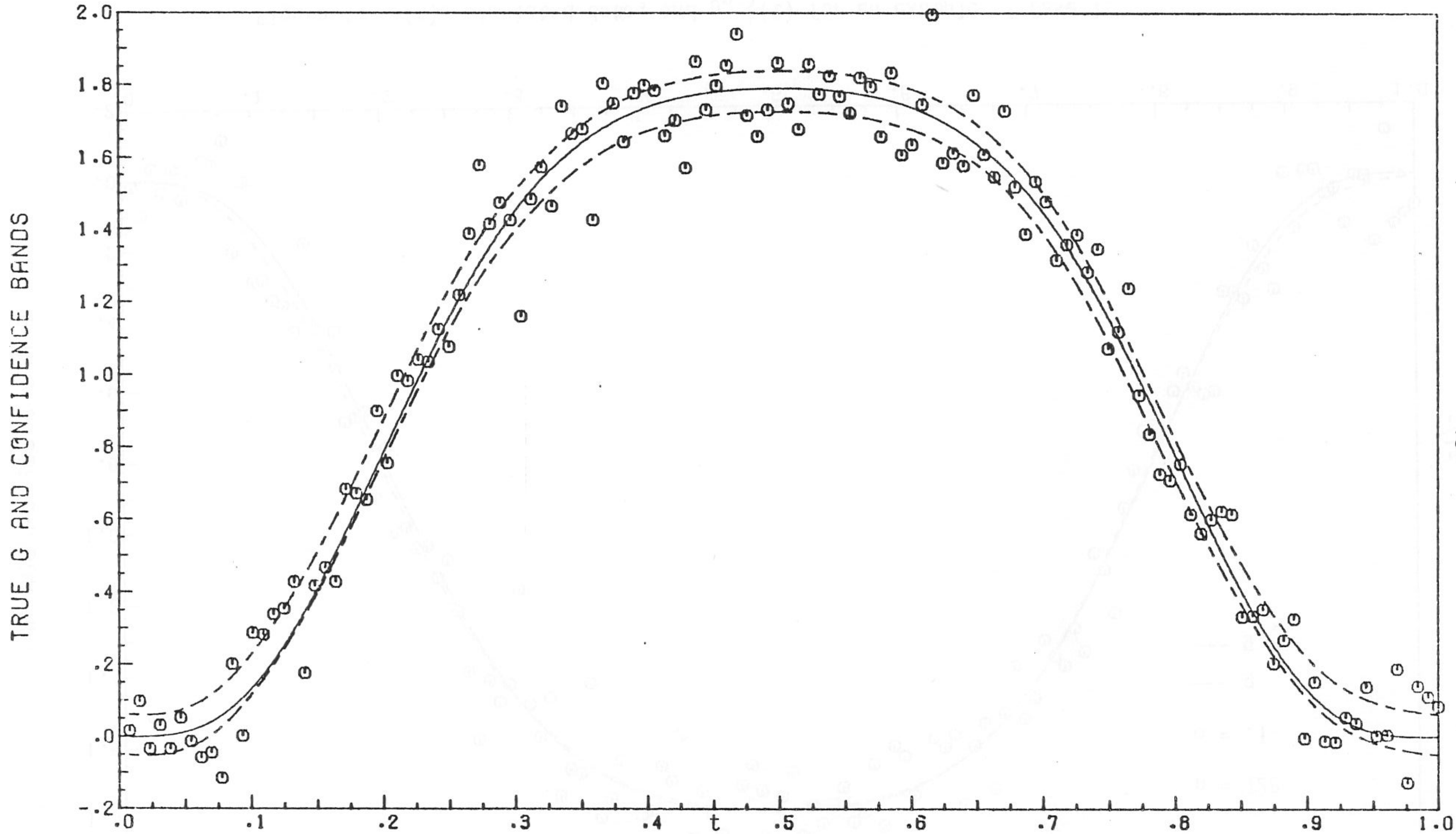

Figure 2. g(t) and data of Figure 1 with 95% confidence bands.

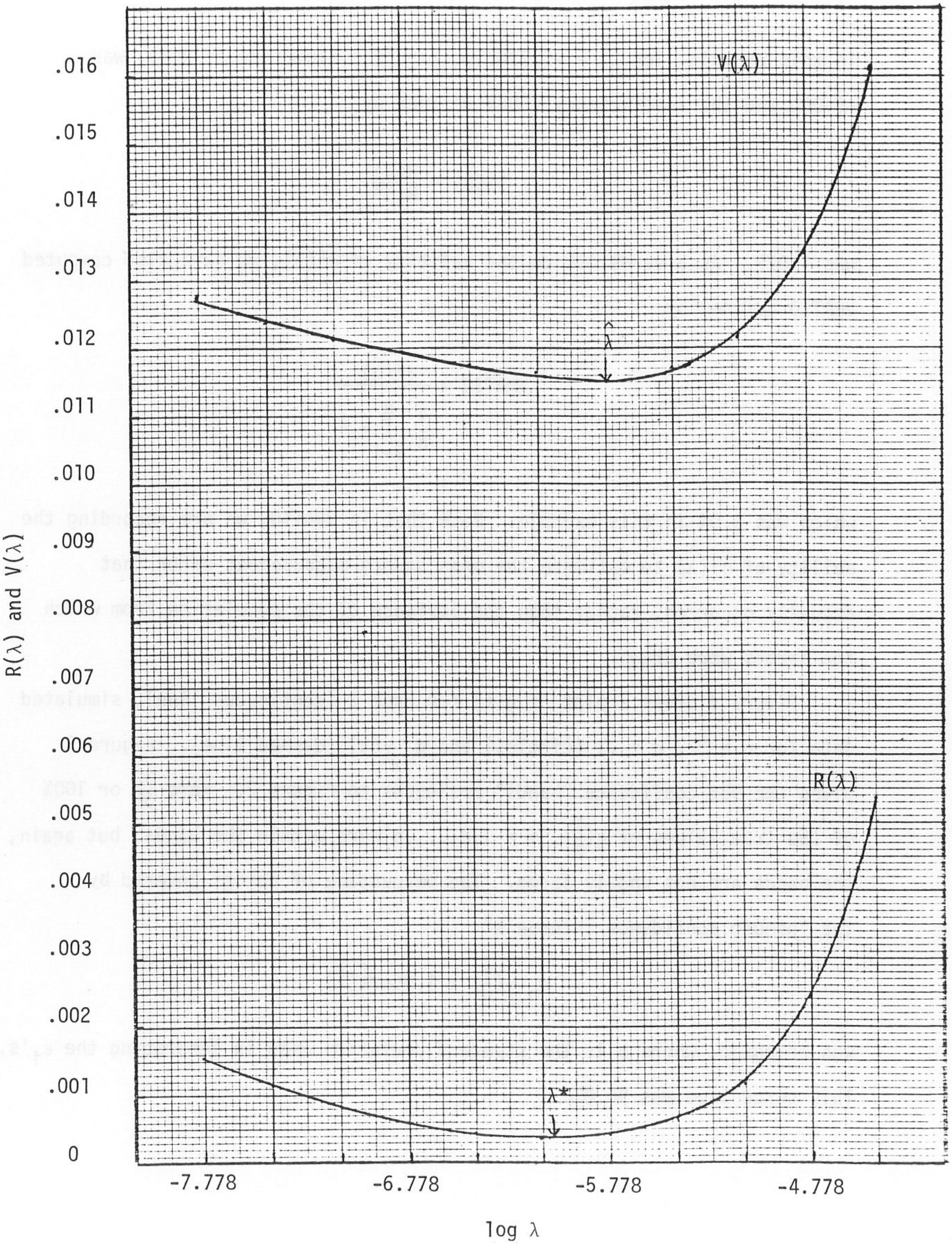

Figure 3. $R(\lambda)$ and $V(\lambda)$.

In this example, $\log_{10}\hat{\lambda} = -5.778$, $\log_{10}\lambda^*$ (the minimizer of $R(\lambda)$, was -6.000, and the inefficiency ISUBV defined by

$$\text{ISUBV} = \frac{R(\hat{\lambda})}{R(\lambda^*)}$$

was 1.078. As a yardstick on the validity of $\hat{\sigma}^2(\hat{\lambda})$, we have also computed VRATIO defined by

$$\text{VRATIO} = \frac{\hat{\sigma}^2(\hat{\lambda})}{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^{\,2}}$$

which was 1.04 in this example. Note that in VRATIO, we are recording the ability of $\hat{\sigma}^2(\hat{\lambda})$ to estimate the mean square measurement error that actually occurred, rather than the variance of the population from which the errors were drawn.

Figure 4 gives a plot of $g(t)$ for case 3 (continuous line), simulated data for $n = 32$, $\sigma = .1$ (circles) and $g_{n,\hat{\lambda}}(t)$ (dashed line). Figure 5 gives the 95% confidence "bands" analogous to Figure 2. Here $\frac{32}{32}$ or 100% of the true values of $g(\frac{i}{n})$, $i = 1,2,\ldots,n$ were within the bands, but again, the bands are not unduly large. The percentage of points covered by the "pseudo confidence intervals"

$$g_{n,\hat{\lambda}}(\tfrac{i}{n}) \pm 1.96\sigma\sqrt{a(\hat{\lambda})}$$

was computed, where σ is the standard deviation used in generating the ε_i's. This percentage was 96.8%.

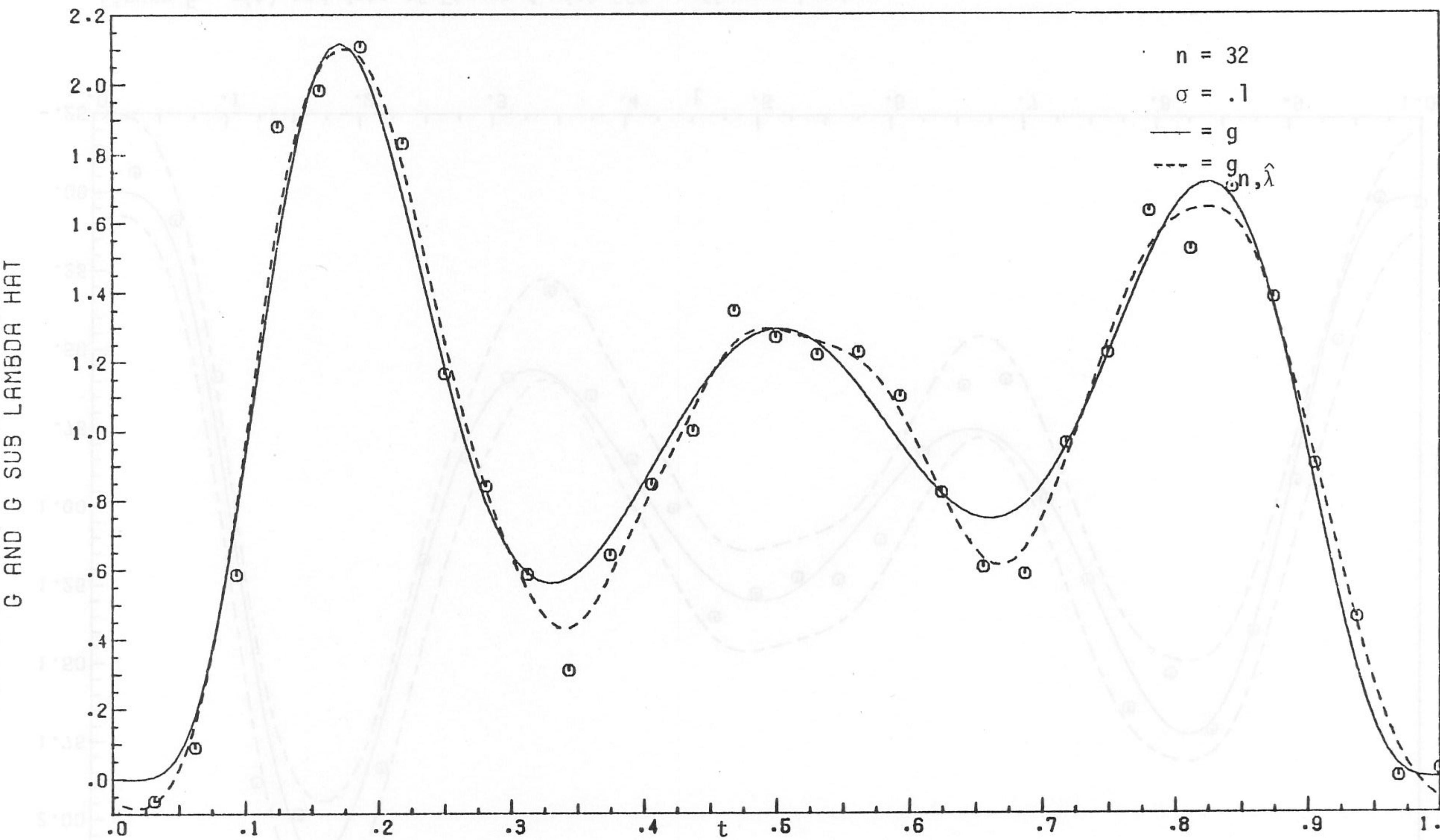

Figure 4. $g(t)$, simulated data, and $g_{n,\hat{\lambda}}(t)$ for an example of Case 3.

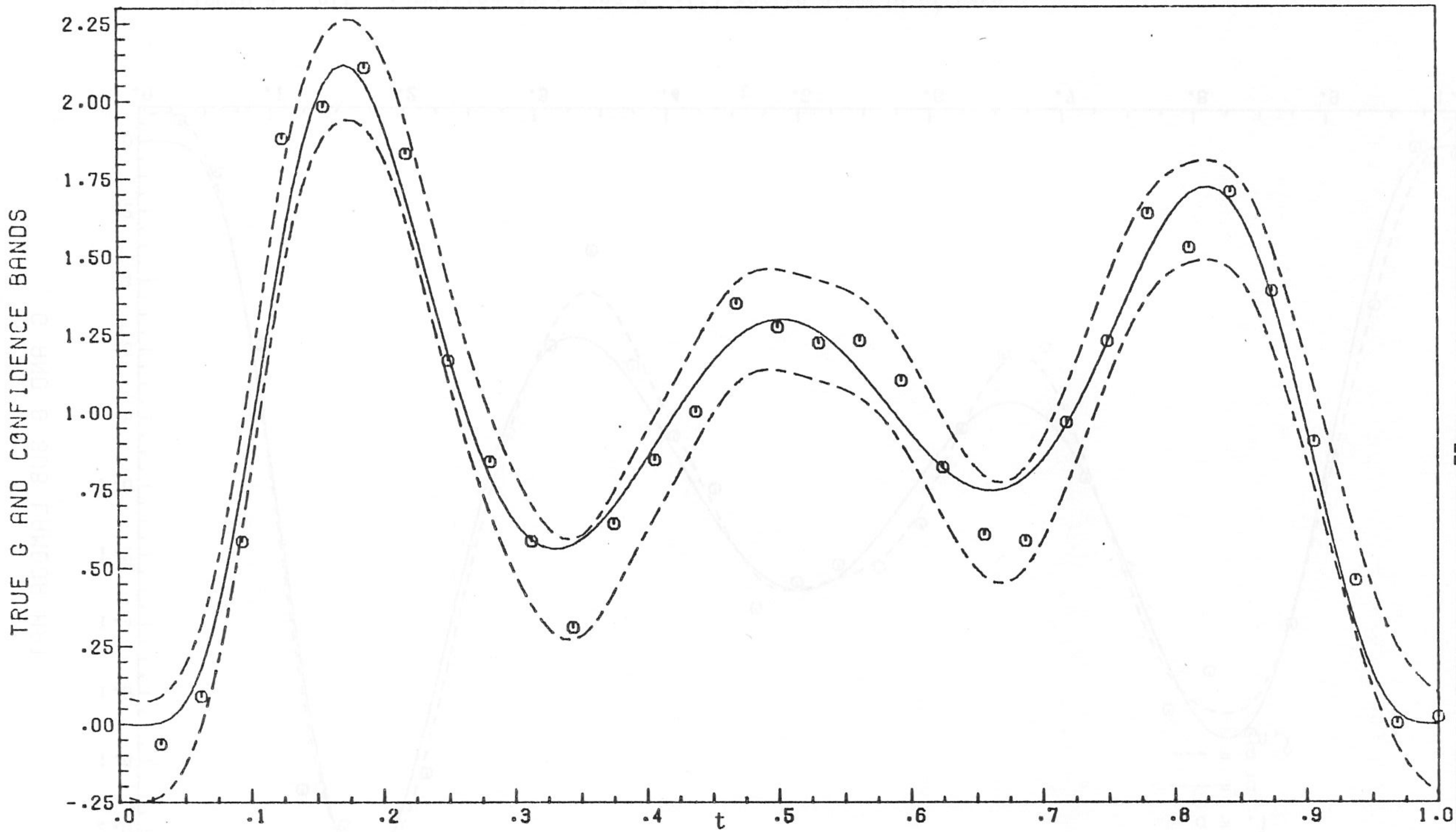

Figure 5. g(t) and data of Figure 4 with 95% confidence bands.

Figure 6-10 are analogous to figures 2 and 5 for Case 2, $n = 64$, with differing σ's. The data in the five figures 6-10 were simulated using $\sigma = .0125$, .025, .05, .1 and .2 respectively. The percentages of the true values of $g(\frac{i}{n})$ covered by the confidence bands in these 5 figures were 96.88, 95.31, 92.19, 100.00 and 96.88, respectively, with a grand average of 96.25%.

The main part of the Monte Carlo experiment consisted of a study of Cases 1, 2 and 3 for $n = 32$, 64 and 128, and $\sigma = .0125$, .025, .05, .01 and .2, giving $3 \times 3 \times 5 = 45$ "examples". Each example was replicated 10 times and for each replication we recorded the inefficiency ISUBV, the VRATIO, "CI 95" which is defined as the % of values of $\{g(\frac{i}{n})\}$ covered by the interval $g_{n,\hat{\lambda}}(\frac{i}{n}) \pm 1.96\hat{\sigma}(\hat{\lambda})\sqrt{a(\hat{\lambda})}$, and "PCI 95", defined as the % of values of $\{g(\frac{i}{n})\}$ covered by the interval $g_{n,\hat{\lambda}}(\frac{i}{n}) \pm 1.96\sigma\sqrt{a(\hat{\lambda})}$.

Table 1 gives ISUBV, VRATIO, PCI 95 and CI 95 for 10 replications of Case 2, with $n = 64$ and $\sigma = .1$. Over the 10 replications it can be seen that the 95% confidence intervals covered between 89.06% and 100% of the true values, for a grand average, over the 10 replications, of 97.03%, that the inefficiency ISUBV was close to 1.0 in all cases and that the estimate of $\hat{\sigma}^2$ was quite good (VRATIO≈1). The PCI's were slightly better than the CI's, as to be expected.

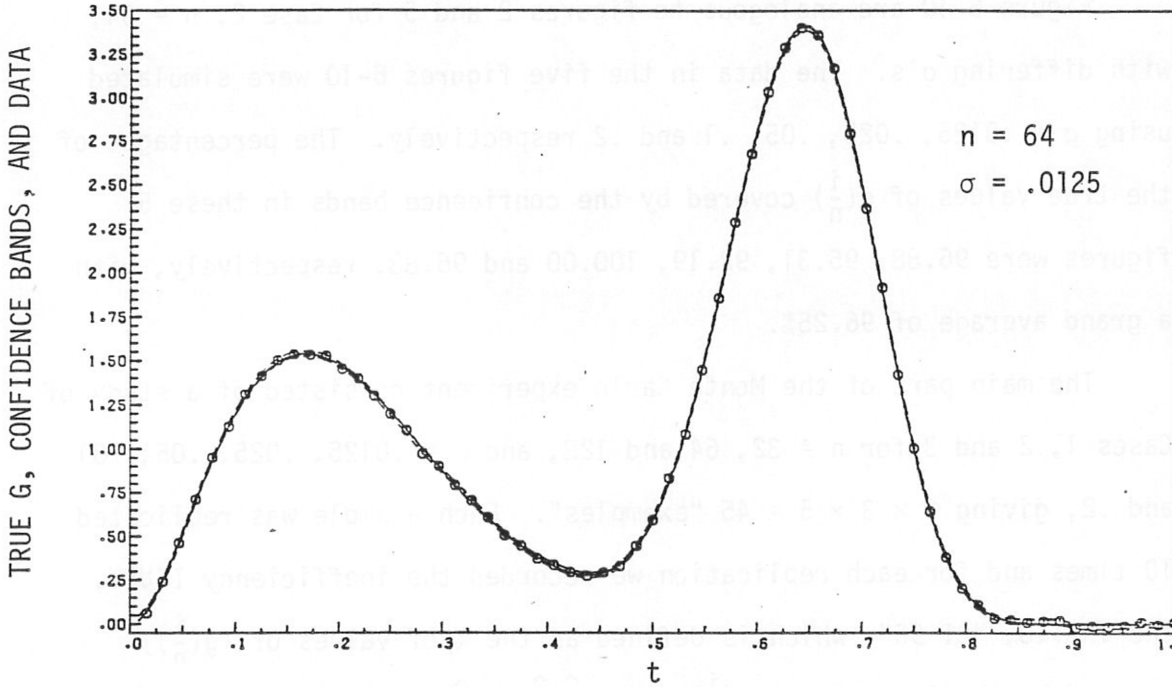

Figure 6. g(t), data, and confidence bands, case 2.

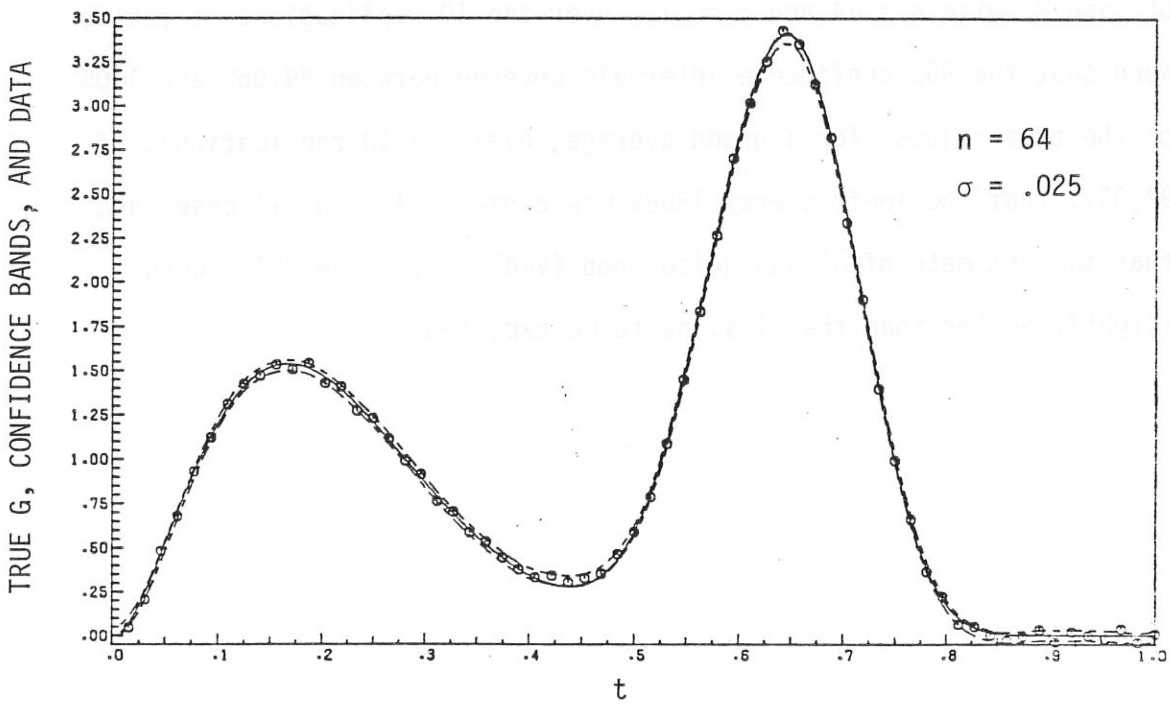

Figure 7. g(t), data, and confidence bands, case 2.

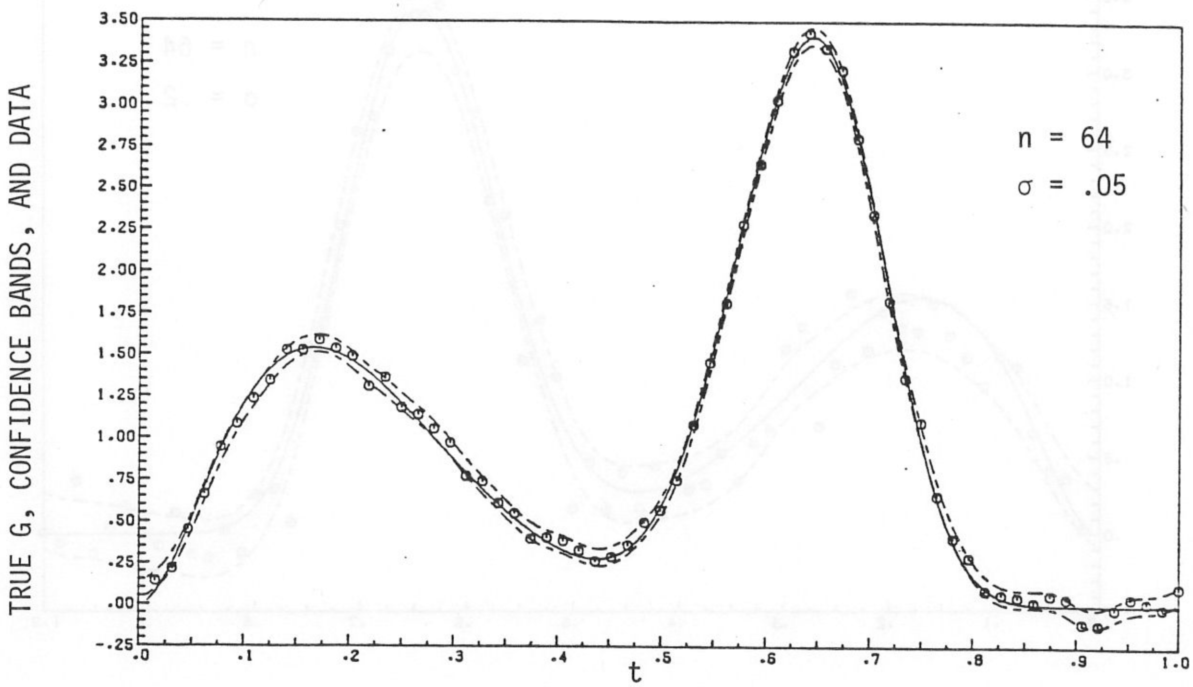

Figure 8. g(t), data, and confidence bands, Case 2.

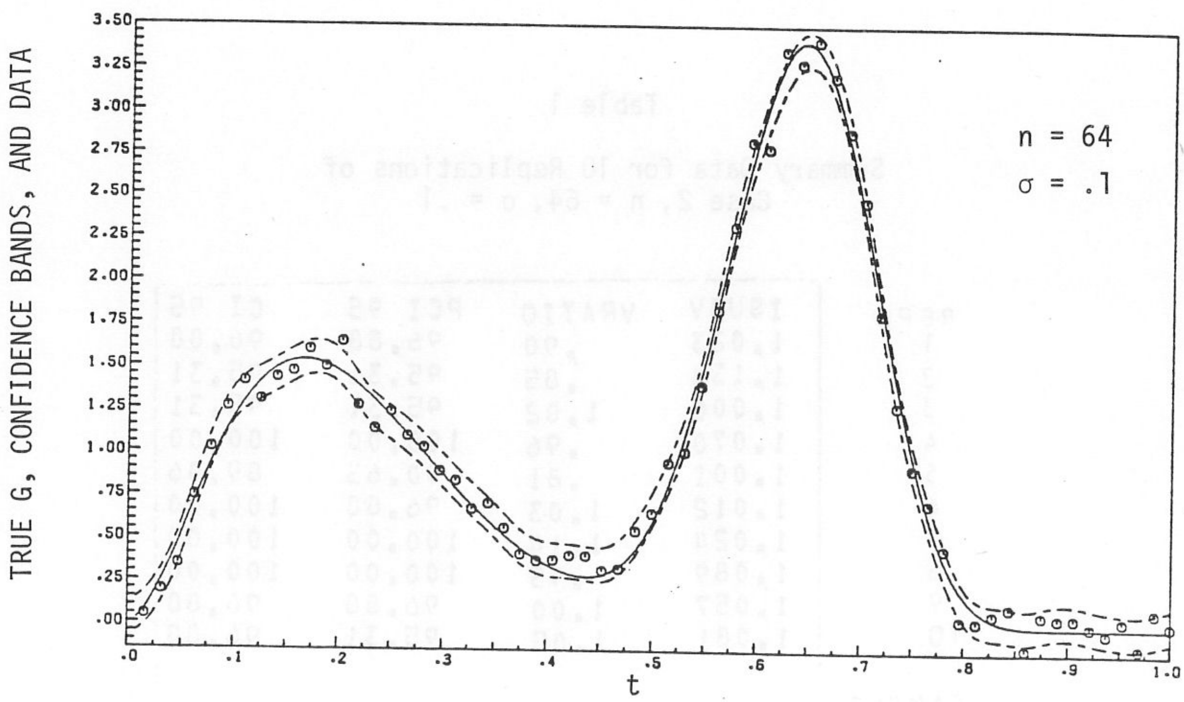

Figure 9. g(t), data and confidence bands, Case 2.

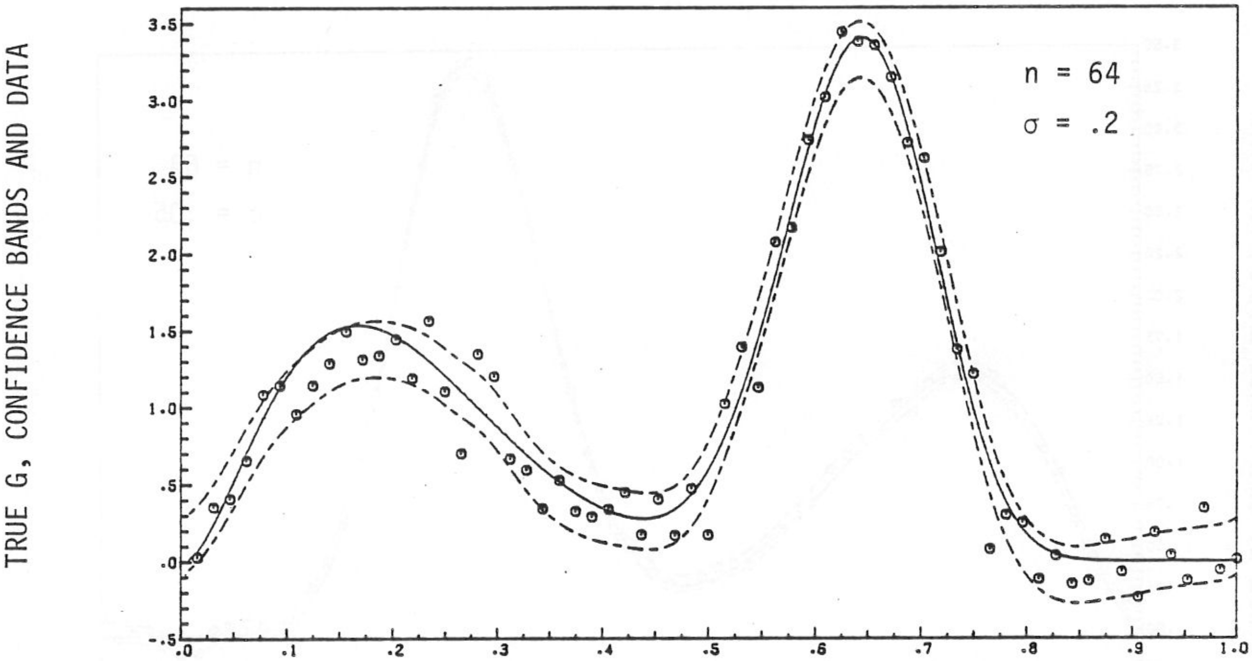

Figure 10. g(t), data, and confidence bands, Case 2.

Table 1

Summary Data for 10 Replications of
Case 2, n = 64, σ = .1

REPL	ISUBV	VRATIO	PCI 95	CI 95
1	1.023	.90	96.88	96.88
2	1.134	.85	95.31	95.31
3	1.000	1.02	95.31	95.31
4	1.070	.96	100.00	100.00
5	1.001	.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	.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	.97	96.72	97.03
SAMPLE S.D.	.042	.09	2.75	3.24

Table 2 gives the sample means and standard deviations of ISUBV, VRATIO, PCI 95 and CI 95 for n = 128, for σ = .0125, .025, .05, .1 and .2 for each of the 3 cases, based on 10 replications of each case. Appendix 1 gives the individual replicate data analogous to Table 1 for each of the 15 examples with n = 128. Both the CI's and the PCI's come remarkably close to covering 95% of the true values. $\hat{\sigma}^2(\hat{\lambda})$ comes remarkably close to estimating $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2$. A slight tendency to systematically underestimate $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2$ is evident in VRATIO. Unfortunately it will be seen that this tendency becomes more pronounced as n decreases.

Tables 3 and 4 give the corresponding summary data for n = 64 and n = 32. DFSIGNAL, which is n - EDF($\hat{\lambda}$), is also given for the n = 32 case. The pseudo confidence intervals continue to come remarkably close to covering 95% of the true values. The performance of the CI's appears to degrade for the smaller σ's with some rather poor results for n = 32, σ = .0125. It can be seen that the poor confidence intervals are associated with VRATIO's substantially less than 1.

Table 2

n = 128

	ISUBV MEAN	ISUBV S.D.	VRATIO MEAN	VRATIO S.D.	PCI 95 MEAN	PCI 95 S.D.	CI 95 MEAN	CI 95 S.D.
σ = .0125								
Case 1	1.063	.05	.99	.04	97.81	3.92	97.42	4.38
Case 2	1.036	.06	.99	.04	97.19	2.19	96.88	2.61
Case 3	1.086	.19	.92	.07	96.72	1.63	96.09	1.91
σ = .025								
Case 1	1.094	.13	.96	.04	97.27	3.05	97.11	2.45
Case 2	1.070	.14	.98	.06	97.03	2.20	97.42	1.85
Case 3	1.067	.08	.97	.08	96.95	3.05	96.72	2.88
σ = .05								
Case 1	1.061	.05	1.00	.02	97.42	2.84	96.64	3.41
Case 2	1.025	.04	.96	.05	96.95	1.99	96.56	2.48
Case 3	1.070	.09	.95	.06	95.86	1.88	95.16	2.88
σ = .1								
Case 1	1.164	.25	.98	.02	95.55	3.88	94.92	4.37
Case 2	1.056	.08	.97	.06	95.70	4.74	95.55	5.42
Case 3	1.044	.07	.98	.04	97.27	3.37	97.27	2.80
σ = .2								
Case 1	1.500	1.20	.98	.05	94.84	5.17	94.14	5.50
Case 2	1.177	.23	.94	.05	95.47	3.29	95.86	3.13
Case 3	1.186	.21	.97	.05	97.11	2.57	96.56	2.45

Table 3

n = 64

	ISUBV MEAN	ISUBV S.D.	VRATIO MEAN	VRATIO S.D.	PCI 95 MEAN	PCI 95 S.D.	CI 95 MEAN	CI 95 S.D.
σ = .0125								
Case 1	1.073	.09	.99	.10	96.09	3.37	95.94	2.72
Case 2	1.308	.59	.75	.39	96.56	2.19	80.31	33.16
Case 3	1.101	.12	.76	.25	96.09	3.14	91.25	6.26
σ = .025								
Case 1	1.094	.10	.98	.09	97.34	2.89	96.41	3.77
Case 2	1.332	.55	.65	.36	95.47	2.75	72.66	36.52
Case 3	1.059	.10	.80	.08	95.78	3.28	92.97	3.71
σ = .05								
Case 1	1.045	.04	.97	.08	95.78	4.94	95.16	4.81
Case 2	1.164	.36	.86	.23	97.66	2.13	94.06	10.89
Case 3	1.137	.19	.82	.24	96.72	2.36	92.66	6.05
σ = .1								
Case 1	1.832	.84	.86	.14	93.59	5.74	92.97	4.26
Case 2	1.049	.04	.97	.09	96.72	2.75	97.03	3.24
Case 3	1.215	.48	.83	.22	94.53	4.21	90.31	11.73
σ = .2								
Case 1	1.324	.52	.99	.09	97.97	2.52	97.03	3.46
Case 2	1.062	.07	1.01	.06	99.06	1.59	98.91	1.72
Case 3	1.446	.91	.87	.27	97.34	1.98	93.12	11.14

Table 4

n = 32

	ISUBV MEAN	ISUBV S.D.	VRATIO MEAN	VRATIO S.D.	PCI 95 MEAN	PCI 95 S.D.	CI 95 MEAN	CI 95 S.D.	DFSIGNAL MEAN	CIx
σ = .0125										
Case 1	1.141	.28	.79	.34	95.31	2.88	86.87	9.76	16.2	93
Case 2	1.242	.21	.16	.22	92.50	4.68	31.56	39.04	30.9	<80
Case 3	1.449	.25	.06	.13	94.37	4.38	12.19	26.01	31.6	<80
σ = .025										
Case 1	1.331	.62	.92	.37	95.94	5.05	85.94	28.95	14.8	93
Case 2	1.272	.26	.45	.33	95.31	4.25	67.19	35.96	26.8	89
Case 3	1.134	.08	.45	.31	95.00	2.07	74.06	21.92	26.8	90
σ = .05										
Case 1	1.950	1.08	.69	.44	92.19	6.13	74.06	29.91	16.0	93
Case 2	1.095	.11	.79	.42	93.44	5.13	82.81	28.10	21.0	92
Case 3	1.473	.62	.53	.45	92.19	4.25	65.63	35.88	24.4	91
σ = .1										
Case 1	1.200	.20	.94	.22	98.44	3.76	96.56	6.91	8.8	94
Case 2	1.070	.07	.90	.24	95.44	3.44	92.19	6.88	17.4	93
Case 3	1.114	.10	.74	.30	96.88	2.80	87.81	12.22	18.6	93
σ = .2										
Case 1	1.402	.80	.97	.17	91.56	7.79	91.25	11.08	6.5	94
Case 2	1.066	.09	.77	.18	93.75	4.42	90.00	5.38	14.9	93
Case 3	2.017	2.57	.76	.36	97.19	4.06	82.19	29.08	16.5	93

The replicate data for the 30 examples summarized in Tables 3 and 4 is also given in the Appendix (except for the DFSIGNAL numbers). By use of the Appendix, it can be seen that the failure of the confidence intervals to have the desired property is primarily traceable to a failure to estimate σ^2 accurately. In these very small σ cases where this happens the cross validation is coming close to choosing λ to interpolate the data. In these examples, it appears that σ^2 is either estimated reasonably well, and the resulting confidence interval is quite satisfactory, or $\hat{\sigma}^2(\hat{\lambda})$ is way off. If the experimenter has some idea of the order of magnitude of σ then, these failure cases can be spotted in practice by examination of $\hat{\sigma}^2(\hat{\lambda})$. For example, inspection of the 10 replications for the Case 2, $n = 64$, $\sigma = .025$, reveals that replications number 1,2,3,5,6,8 and 9 gave satisfactory confidence intervals whereas in replications 4 and 7 the smoothing spline very nearly interpolated the data (erroneously). The confidence intervals were meaningless for replicates 4 and 7 but this could easily have been detected with only a crude knowledge of the true σ^2. This ability to detect gross failure cases seems to be generally true in all the $n = 64$ and $n = 32$ cases where the confidence intervals were unsatisfactory. A gross underestimate of σ^2 will result in DFSIGNAL close to n, equivalently $EDF(\hat{\lambda})$ close to 0.

For the typical $n = 32$ case, 1.96 is not a good approximation to $t_{.025}(EDF(\hat{\lambda}))$. The column labeled CIx in Table 4 gives the true size of a confidence interval using 1.96 if the true distribution were t with $EDF(\hat{\lambda}) = n - DFSIGNAL$ degrees of freedom. From this, it appears that the replacement of 1.96 by $t_{.025}(EDF(\hat{\lambda}))$ would most probably have improved the overall results in the CI 95 column modestly. These numerical results as well as the asymptotic theory presented later suggests that the confidence interval results will only improve in similar examples, as n becomes larger. Several n = 16 cases were tried and the results were variable. The GCV method is recommended only with caution for n as small as 16.

Figures 11 and 12 give the figures analogous to figures 1 and 2, for an example of case 4. Case 4 is a triangular shaped hill function, and is not in W_2^2. None of the theory developed here necessarily applies to this case. Only the four examples for all combinations of n = 64, n = 128, $\sigma = 0.05$ and $\sigma = .1$ were tried. A summary of the results of 10 replications (analogous to Tables 2, 3 and 4) appears in Table 5. The average VRATIOS are about the same as for the smooth examples, 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.

In an attempt to defeat the method soundly we considered a function with a large discontinuity. Ordinarily one would not attempt to estimate such a function with a cubic spline, which has a continuous second derivative. Figures 13 and 14 are analogous to figures 1 and 2 for Case 5 with n = 64, $\sigma = .1$. Case 5 has a jump of 12 at t = .5. Table 5 gives summary data for Case 5 with n = 64 and 128, $\sigma = .05$ and .1. Here σ is overestimated by a factor between around 30 and 300. Overshoot (or "Gibbs effect") in $g_{n,\hat{\lambda}}$ near t = .55 is clearly visible. However in the n = 64 examples exactly 62 (=96.88%) and in the n = 128 case examples exactly 126 (98.44%) true values of g were covered by the confidence intervals in each replication.

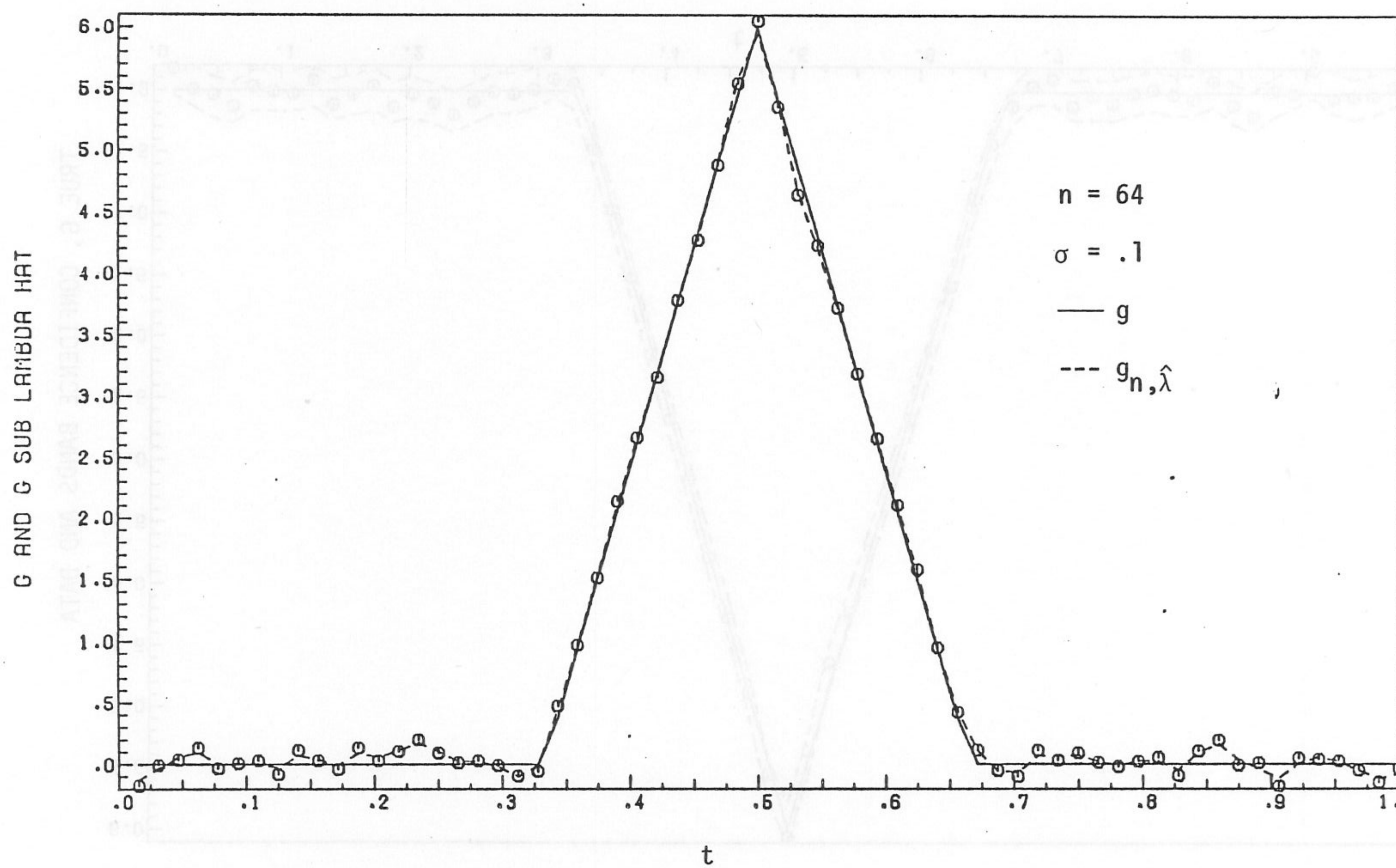

Figure 11. $g(t)$, simulated data, and $g_{n,\hat{\lambda}}(t)$ for an example of Case 4.

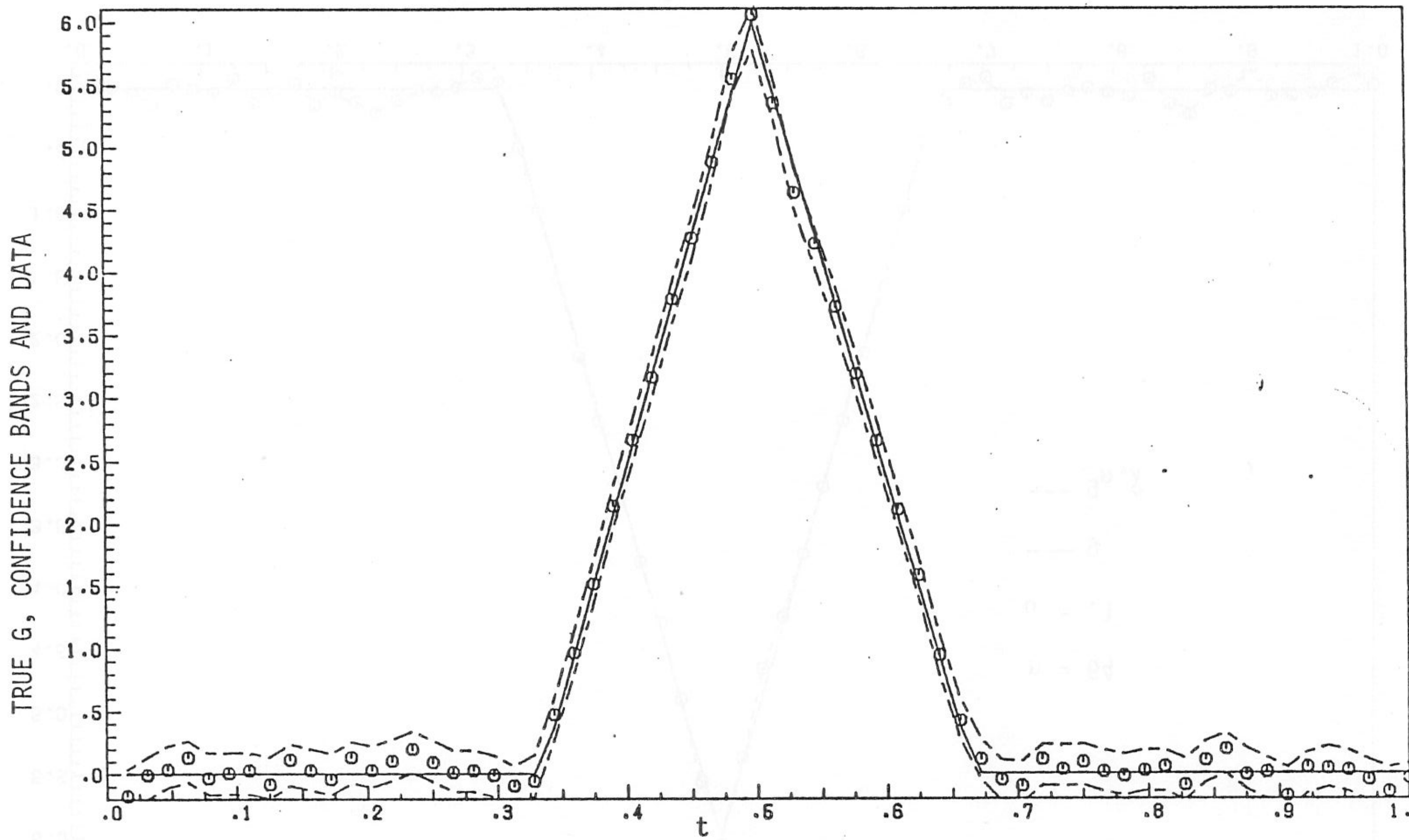

Figure 12. g(t) and data of Figure 11 with 95% confidence bands.

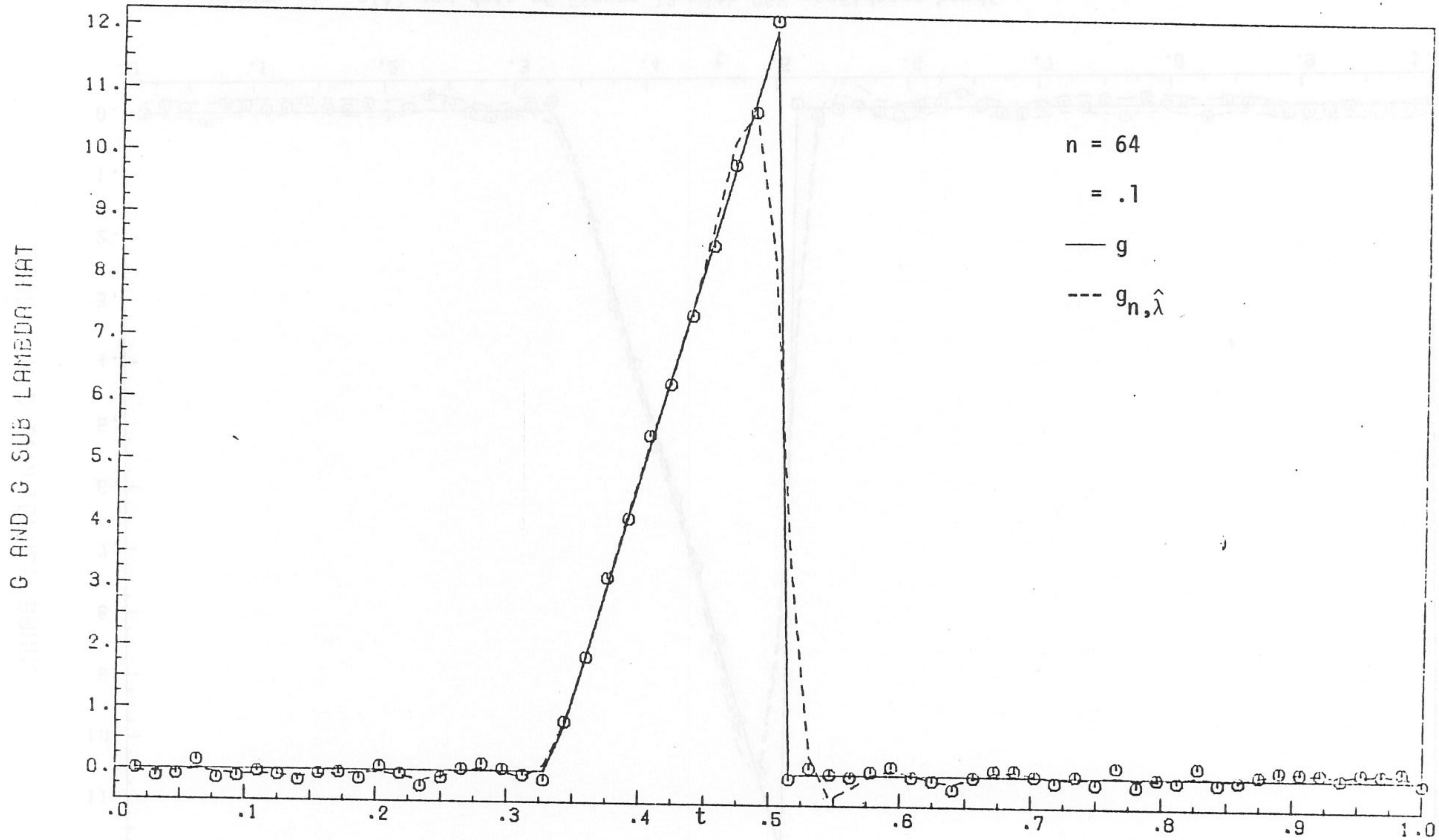

Figure 13. $g(t)$, simulated data, and $g_{n,\hat{\lambda}}(t)$ for an example of Case 5.

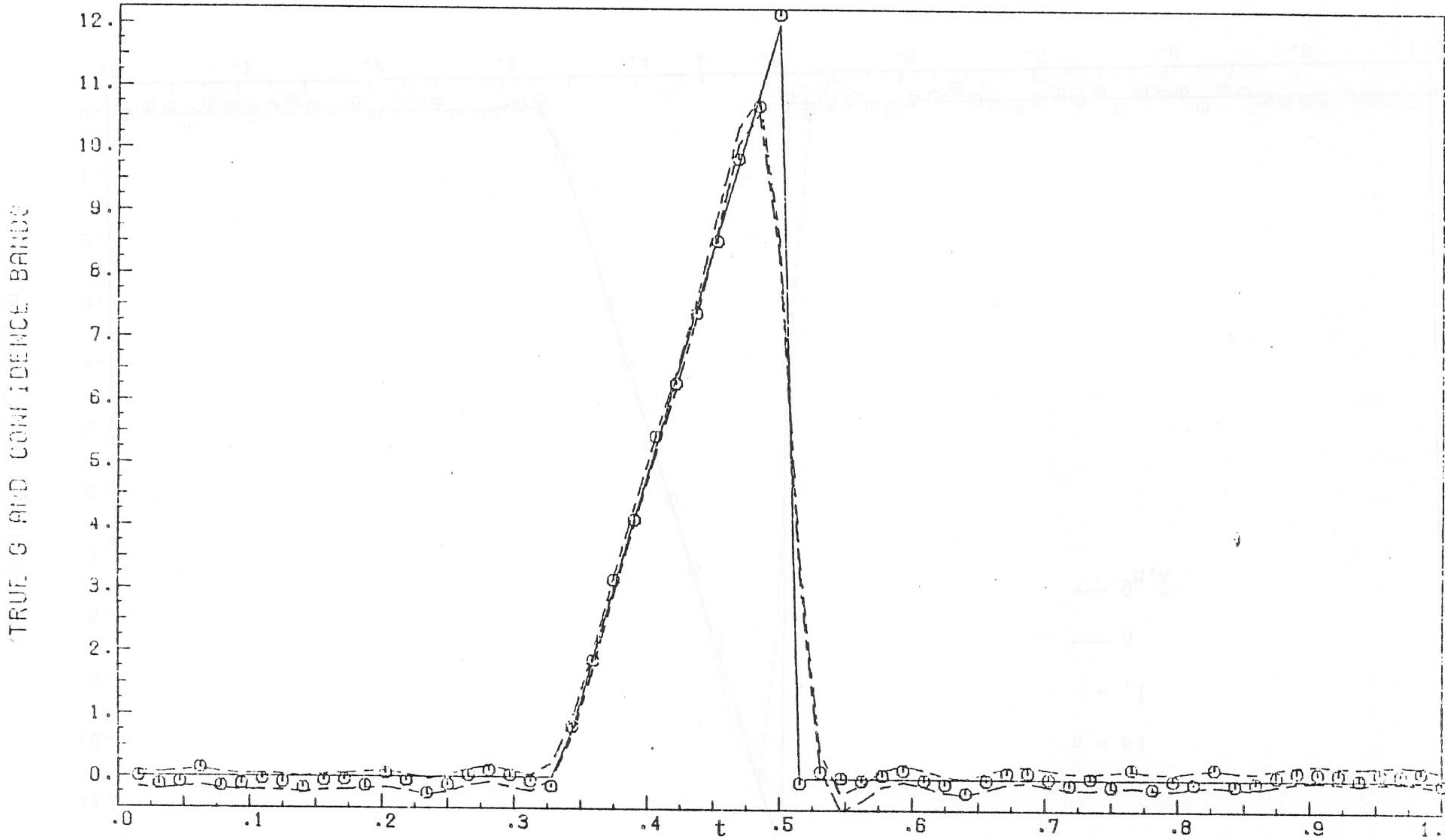

Figure 14. g(t) and data of Figure 13 with 95% confidence bands.

Table 5

	ISUBV MEAN	ISUBV S.D.	VRATIO MEAN	VRATIO S.D.	PCI 95 MEAN	PCI 95 S.D.	CI 95 MEAN	CI 95 S.D.
n = 64								
$\sigma = .05$								
Case 4	1.094	.15	1.08	.51	93.91	3.98	84.69	28.55
Case 5	43.33	5.4	298.8	51.3	84.2	2.15	96.88	0
$\sigma = .1$								
Case 4	1.225	.48	.97	.46	93.91	3.86	83.59	28.41
Case 5	23.18	4.96	70.81	14.83	87.81	2.40	96.88	0
n = 128								
$\sigma = .05$								
Case 4	1.138	.22	.92	.30	95.47	1.91	91.64	7.71
Case 5	41.27	6.39	131.02	11.38	92.66	1.41	98.44	0
$\sigma = .1$								
Case 4	1.091	.13	.92	.12	95.70	2.16	94.45	2.33
Case 5	14.78	2.58	32.25	6.00	92.81	1.84	98.44	0

It would be nice to have a better estimate of σ^2 in the small n case. It can be shown that the maximum likelihood estimate of σ^2, given λ, is

$$\sigma^2_{ML}(\lambda) = \frac{1}{n} y'(I-A(\lambda))y$$

which becomes, in the equally spaced periodic case

$$\hat{\sigma}^2_{ML}(\lambda) = \frac{2}{n} \sum_{\nu=1}^{n/2-1} \frac{\lambda}{\lambda_\nu+\lambda} (a_\nu^2+b_\nu^2) + \frac{1}{n} \frac{\lambda}{\lambda_{n/2}+\lambda} a^2_{n/2} \quad .$$

We computed $\hat{\sigma}^2(\hat{\lambda})$ and $\text{VRATIO}_{ML} = \hat{\sigma}^2_{ML}(\hat{\lambda})/\frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2$, for 10 replicates of several of the examples with worst VRATIO. In most of these cases $\hat{\sigma}^2(\hat{\lambda})$ tended to be too big by roughly the same factor as $\hat{\sigma}^2_{ML}(\hat{\lambda})$ was too small, and has a higher variance. This very brief study was by no means definitive, however and further study may be warranted.

Finally, we give a bivariate example, kindly provided to us by J. Wendelberger, using the computer program developed in his forthcoming thesis (Wendelberger (1981). Figure 15 depicts Franke's "Principal test function"

$$f(x,y) = .75\exp\left[-\frac{(9x-2)^2+(9y-2)^2}{4}\right]$$

$$+ .75\exp\left[-\frac{(9x+1)^2}{49} - \frac{9y+1}{10}\right]$$

$$+ .5\exp\left[-\frac{(9x-7)^2+(9y-3)^2}{4}\right]$$

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

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

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

with

$$\text{with } x_i = \frac{2i+1}{2N},\ y_j = \frac{2j+1}{2N},\ i,j = 1,2,\dots,N,$$

with $N = 13$, giving $n = N^2 = 169$ data points. The peak height of f was approximately 1.2 and σ was taken as .03. f was estimated as the so called "thin plate smoothing spline" which is the minimizer (in an appropriate space) of

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

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

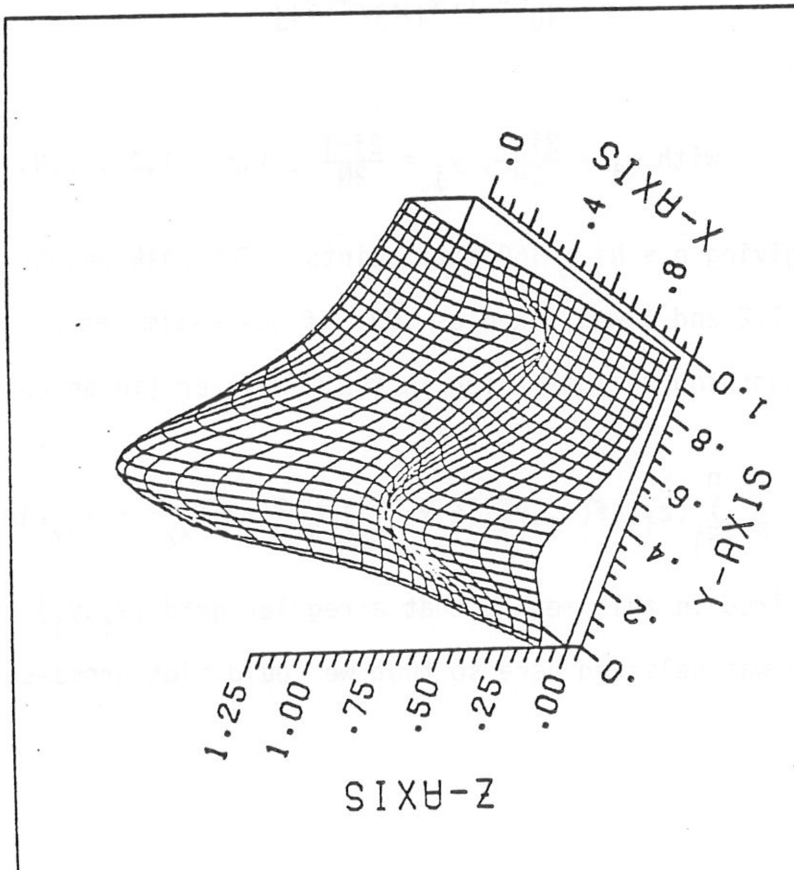

Figure 15.
FRANKE'S PRINCIPAL TEST FUNCTION

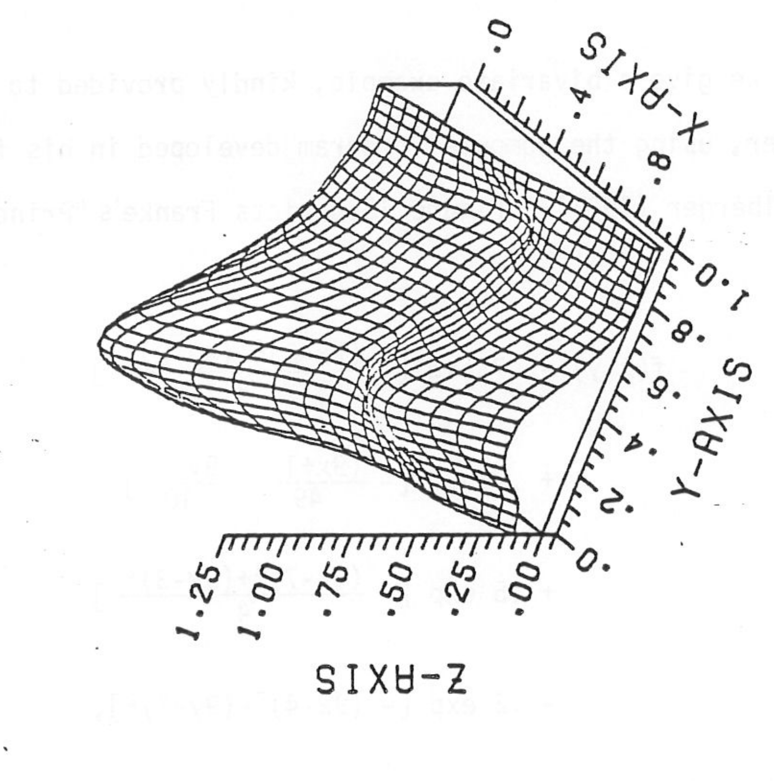

Figure 16.
SPLINE FIT
169 POINTS WITH SIGMA = .03

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 is given in Wendelberger (1981). Figure 16 gives the resulting thin plate smoothing spline estimate of f. Figure 17 gives 7 selected cross sections for 7 fixed values of x, $x=(2i+1)/N$ for $i = 1,3,5,7,9,11,13$. In each cross-section is plotted $f((2i+1)/N,y)$, $0 \leq y \leq 1$, (solid line), $f_{n,\hat{\lambda}}((2i+1)/N,y)$ $0 \leq y \leq 1$, 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

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

where $a_{ij,ij}(\lambda) = \dfrac{\partial g_{n,\lambda}(x_i,y_j)}{\partial z_{ij}}$. Of the 196 confidence intervals, 162 or 95.85% 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.

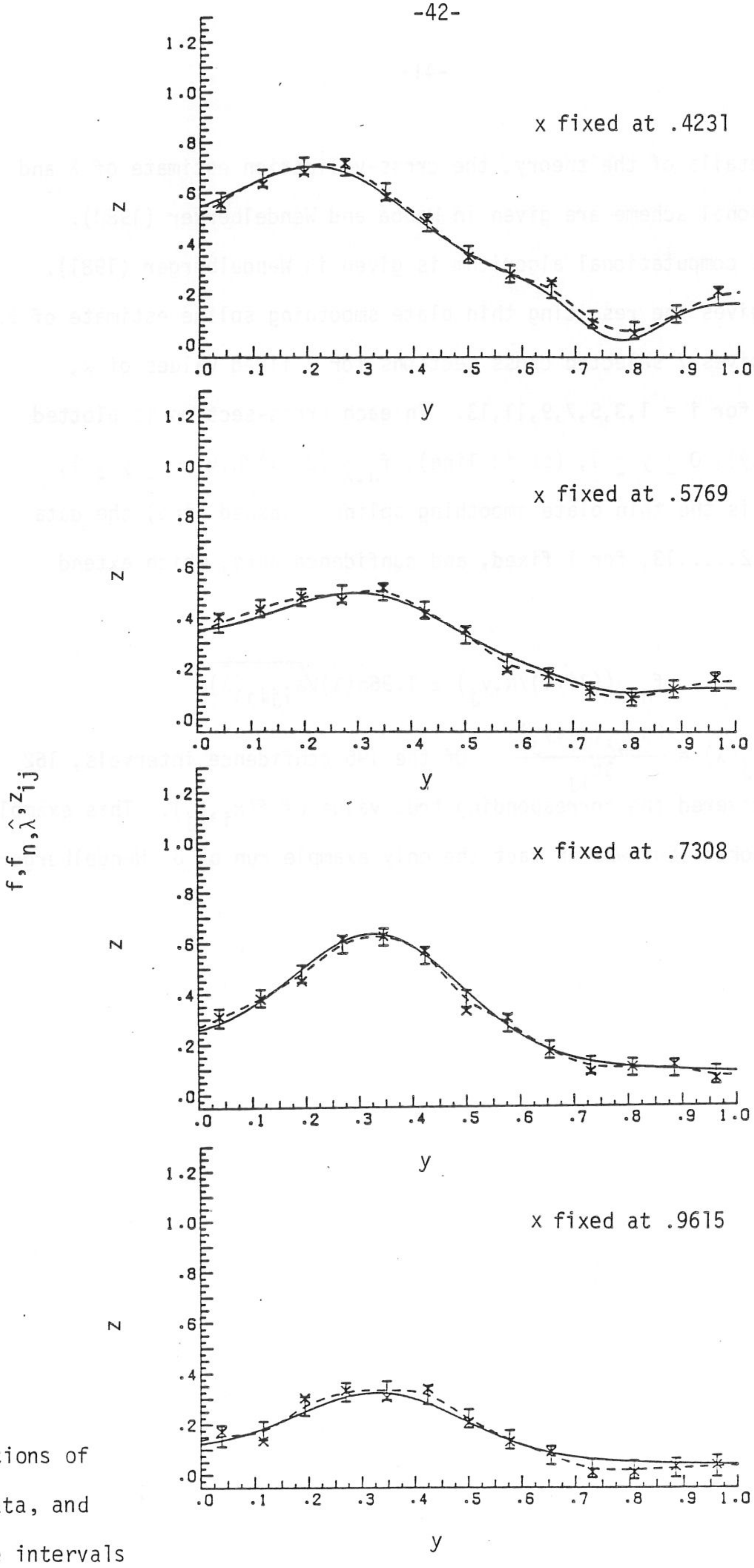

Figure 17.

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

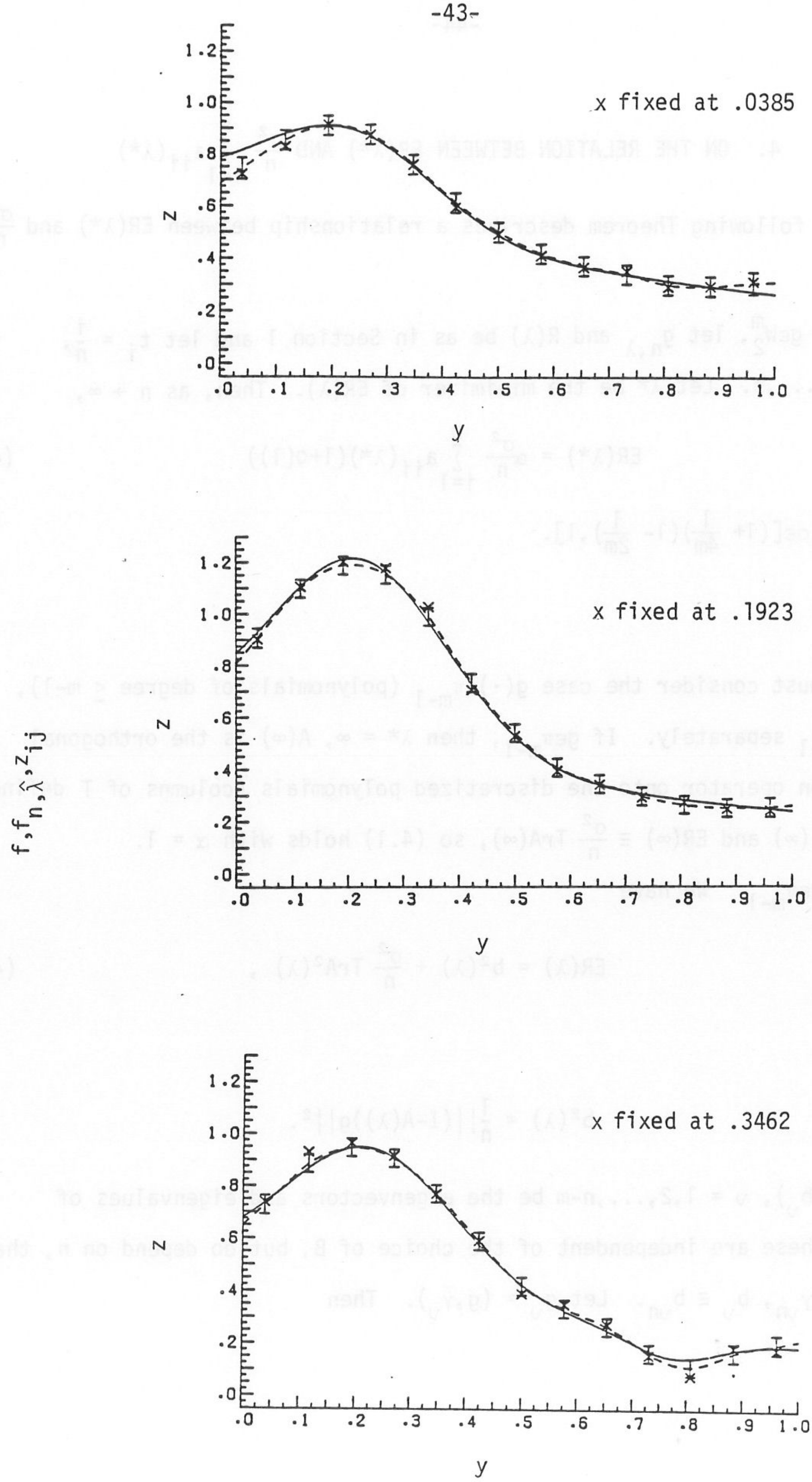

Figure 17 continued

4. 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 $\frac{\sigma^2}{n}\sum_{i=1}^{n} a_{ii}(\lambda^*)$.

Theorem:

Let $g\varepsilon W_2^m$, let $g_{n,\lambda}$ and $R(\lambda)$ be as in Section 1 and let $t_i = \frac{i}{n}$, $i = 1,2,\ldots,n$. Let λ^* be the minimizer of $ER(\lambda)$. Then, as $n \to \infty$,

$$ER(\lambda^*) = \alpha\frac{\sigma^2}{n}\sum_{i=1}^{n} a_{ii}(\lambda^*)(1+o(1)) \tag{4.1}$$

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

Argument:

We must consider the case $g(\cdot)\varepsilon\pi_{m-1}$ (polynomials of degree $\leq m-1$), and $g\notin\pi_{m-1}$ separately. If $g\varepsilon\pi_{m-1}$, then $\lambda^* = \infty$, $A(\infty)$ is the orthogonal projection operator onto the discretized polynomials (columns of T defined in Sect. 2) $A(\infty) = A^2(\infty)$ and $ER(\infty) \equiv \frac{\sigma^2}{n} TrA(\infty)$, so (4.1) holds with $\alpha = 1$.

Suppose $g\varepsilon\pi_{m-1}$. We have

$$ER(\lambda) = b^2(\lambda) + \frac{\sigma^2}{n} TrA^2(\lambda) \ , \tag{4.2}$$

where

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

Let (γ_ν, nb_ν), $\nu = 1,2,\ldots,n-m$ 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

$$b^2(\lambda) = \sum_{\nu=1}^{n-m} \frac{\lambda^2 g_\nu^2}{(b_\nu+\lambda)^2} \tag{4.3}$$

$$\frac{1}{n}TrA^2(\lambda) = \frac{1}{n}\sum_{\nu=1}^{n-m} \left(\frac{b_\nu}{\lambda+b_\nu}\right)^2 = \frac{1}{n}\sum_{\nu=1}^{n-m} \frac{1}{(1+\lambda/b_\nu)^2} . \tag{4.4}$$

$$\frac{1}{n}\sum_{i=1}^{n} \sigma_{ii}(\lambda) = \frac{1}{n}TrA(\lambda) = \frac{1}{n}\sum_{\nu=1}^{n-m} \frac{1}{(1+\lambda/b_\nu)} . \tag{4.5}$$

Utreras (1979b,1980) has shown (for $t_i = i/n$, $m = 2,3,\ldots$, and also more generally for $m = 2$), that there exist C_1, C_2, $0 < C_1 < C_2 < \infty$, independent of n such that $C_1\nu^{-2m} \le b_\nu \le C_2\nu^{-2m}$, $\nu = 1,2,\ldots,n-m$, $n = 1,2,\ldots$. (For an earlier, heuristic argument, see Craven and Wahba (1979.)

Thus, there exists some C satisfying $C_1 \le C^{2m} \le C_2$ such that

$$\frac{1}{n}TrA^2(\lambda) \approx \frac{C\ell_m}{n\lambda^{1/2m}}, \quad \frac{1}{n}TrA(\lambda) \approx \frac{C\tilde{\ell}_m}{n\lambda^{1/2m}} ,$$

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

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

It is shown in Craven and Wahba (1979) that

$$b^2(\lambda) \le \lambda J_m(g) \tag{4.6}$$

where

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

((4.6) actually holds for all the seminorms considered in section 3 of W, whether or not the $\{t_i\}$ are equally spaced). More generally, if

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

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

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

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

$$\sum_{\nu=1}^{n-m} \frac{g_\nu^2}{b_\nu^2} \leq J_{2m}(g) \tag{4.7}$$

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 (4.7) 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

$$R(\lambda) \leq (\lambda^p g_p + \frac{\sigma^2 C\ell_m}{n\lambda^{1/2m}})(1+o(1)) \tag{4.8}$$

as $\lambda \to 0$, $n\lambda^{1/2m} \to \infty$, for each $p\varepsilon[1,2]$ for which g_p is finite, with equality for p = 2, if $g < \infty$. The minimizer of the right hand side of (4.8) is

$$\lambda^*_p = \left(\frac{\sigma^2}{g_p}\frac{C\ell_m}{2mp}\frac{1}{n}\right)^{\frac{2m}{2mp+1}}$$

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

$$R(\lambda^*_p) \lesssim \left(\frac{\sigma^2 C\ell_m}{2mpn}\right)^{\theta} g_p^{1-\theta}(2mp+1) \tag{4.9}$$

$$\frac{\sigma^2}{n}\mathrm{Tr}A(\lambda^*_p) = \left(\frac{\sigma^2 C\ell_m}{2mpn}\right)^{\theta} g_p^{(1-\theta)}(2mp)\left(\frac{2m}{2m-1}\right) \tag{4.10}$$

where we have used $\tilde{\ell}_m/\ell_m = (2m)/(2m-1)$. Arguing heuristically that equality in (4.9) must hold for some p between 1 and 2 gives

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

for some $p\varepsilon[1,2]$. Since this quantity is between $(1+\frac{1}{4m})(1-\frac{1}{2m})$ and $(1+\frac{1}{2m})(1-\frac{1}{2m})$, the result follows.

We remark that this argument can be repeated in the general context of W 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).

ACKNOWLEDGMENTS

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 the fall of 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 USARO under Grant No. DAAG29-80-K0042.

REFERENCES

Adams, R.A. (1975). Sobolev Spaces. Academic Press, New York.

Agarwal, G.G. and Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines, Ann. Statist. 8, 6, 1307-1325.

Anderson, T.W. (1958). An introduction to multivariate statistical analysis, John Wiley & Sons, N.Y., p. 28.

Anderssen, R. and Bloomfield, P. (1974). A time series approach to numerical differentiation, Technometrics 16, 69-75.

Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, 4, 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, 14, 1437-1446.

Gamber, H.A. (1979b). Choice of an optimal shape parameter when smoothing noisy data. Commun. Statist., A8, 14, 1425-1436.

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

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

Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. J. Math. Anal. Applic. 33, 82-95.

Knafl, G., Sacks, J. and Ylvisaker, D. (1981). Model robust confidence intervals. Manuscript.

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 #67.

Stone, C. (1980). Optimal rates of convergence for nonparametric estimators, Ann. Statist. 8, 6, 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.

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

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

Wahba, G. (1977a). Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numerical Analysis, 14, 4.

Wahba, G. (1977b). Invited discussion to "Consistent nonparametric regression by C.J. Stone, Ann. Stat., 5, 4, 637-645.

Wahba, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression. J. Roy. Stat. Soc. B, 49, 3, 364-372.

Wahba, G. (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.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere. SIAM J. Scientific and Statistical Computing 2, 1, 5-16.

Wahba, G. and Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation, Monthly Weather Review 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, manuscript.

Wendelberger, J. (1981). Ph.D. thesis, Department of Statistics, University of Wisconsin, Madison, in preparation.

APPENDIX

Tables of ISUBV, VRATIO, PCI 95 and CI 95 for 10 replications each of Cases 1, 2, and 3, σ = .0125, .025, .05, .1 and .2, n = 128, 64, 32.

n = 128
σ = .0125

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.127	.57	100.00	100.00	1.024	1.00	97.66	98.44	1.017	1.03	97.66	96.09
2	1.000	.99	100.00	98.44	1.019	.96	92.97	91.41	1.005	.88	94.53	92.97
3	1.062	1.02	98.44	100.00	1.185	.95	99.22	99.22	1.001	.95	96.09	96.88
4	1.091	1.03	100.00	100.00	1.005	1.01	96.88	96.88	1.055	.88	98.44	97.66
5	1.115	.97	100.00	100.00	1.107	.90	99.22	98.44	1.029	.94	96.09	94.53
6	1.002	.91	92.19	91.41	1.014	1.07	100.00	100.00	1.005	.93	99.22	98.44
7	1.035	1.01	100.00	100.00	1.000	1.01	96.88	97.66	1.064	.96	93.75	93.75
8	1.085	.96	99.22	97.66	1.000	1.00	93.75	93.75	1.646	.76	96.09	94.53
9	1.116	.95	88.28	86.72	1.000	1.00	98.44	98.44	1.020	.91	97.66	98.44
10	1.000	1.06	100.00	100.00	1.002	.98	96.88	94.53	1.019	.95	97.66	97.66
SAMPLE MEAN	1.063	.99	97.81	97.42	1.036	.99	97.19	96.88	1.086	.92	96.72	96.09
SAMPLE S.D.	.048	.04	3.92	4.38	.058	.04	2.19	2.61	.188	.07	1.63	1.91

σ = .0250

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.428	.94	99.22	95.31	1.006	.95	93.75	94.53	1.274	.88	94.53	94.53
2	1.179	.88	93.75	93.75	1.011	1.01	95.31	95.31	1.021	1.05	100.00	99.44
3	1.009	.99	100.00	98.44	1.022	.94	97.66	97.66	1.059	1.04	100.00	98.44
4	1.020	.96	96.88	97.66	1.038	.99	100.00	100.00	1.030	.95	97.66	98.44
5	1.026	.97	97.66	97.66	1.022	.97	96.88	96.88	1.011	1.01	98.44	100.00
6	1.112	.99	91.41	94.53	1.486	.87	97.66	96.09	1.009	.90	100.00	97.66
7	1.000	.99	100.00	100.00	1.012	1.02	93.75	98.44	1.044	.88	94.53	94.53
8	1.000	.94	93.75	93.75	1.006	.99	100.00	100.00	1.150	1.14	99.22	100.00
9	1.141	.95	100.00	100.00	1.080	1.11	99.22	99.22	1.048	.93	90.63	90.63
10	1.023	1.03	100.00	100.00	1.019	.90	96.09	96.09	1.024	.94	94.53	94.53
SAMPLE MEAN	1.094	.96	97.27	97.11	1.070	.98	97.03	97.42	1.067	.97	96.95	96.72
SAMPLE S.D.	.127	.04	3.05	2.45	.140	.06	2.20	1.85	.079	.08	3.05	2.88

Table A-1. ISUBV, VRATIO, PCI 95 and CI 95 for n = 128, σ = .0125, .025, .05, .1 and .2 for 10 replications of cases 1, 2 and 3.

n = 128

σ = .0500

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.129	1.05	100.00	100.00	1.000	.97	96.88	96.88	1.000	.97	96.88	96.88
2	1.027	.98	100.00	100.00	1.107	.94	96.88	97.66	1.002	.92	96.88	96.09
3	1.000	1.00	100.00	98.44	1.000	.91	92.97	91.41	1.126	.89	95.31	95.31
4	1.039	1.00	100.00	100.00	1.004	.98	96.09	96.88	1.234	1.07	96.09	97.66
5	1.124	1.00	100.00	100.00	1.052	.98	100.00	98.44	1.007	.92	96.09	96.09
6	1.155	.95	96.09	96.09	1.001	.88	94.53	92.97	1.012	.95	93.75	93.75
7	1.015	.98	96.88	96.09	1.001	.93	97.66	95.31	1.000	1.03	98.44	98.44
8	1.041	1.01	93.75	92.19	1.000	.99	98.44	99.22	1.224	.88	98.44	94.53
9	1.003	.97	92.19	92.19	1.026	.99	96.88	97.66	1.024	.87	94.53	87.50
10	1.078	1.01	95.31	91.41	1.062	1.06	99.22	99.22	1.072	1.00	92.19	95.31
SAMPLE MEAN	1.061	1.00	97.42	96.64	1.025	.96	96.95	96.56	1.070	.95	95.86	95.16
SAMPLE S.D.	.054	.02	2.84	3.41	.035	.05	1.99	2.48	.089	.06	1.88	2.88

σ = .100

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.217	1.01	90.63	90.63	1.025	.94	90.63	91.41	1.236	.88	100.00	99.22
2	1.027	.97	92.19	92.19	1.031	1.03	100.00	100.00	1.005	.99	100.00	98.44
3	1.440	.98	100.00	100.00	1.023	1.01	100.00	100.00	1.049	.98	100.00	100.00
4	1.009	.97	92.19	92.97	1.016	1.01	100.00	100.00	1.018	.97	91.63	92.19
5	1.091	1.00	92.97	93.75	1.144	1.00	89.06	89.06	1.019	1.00	100.00	100.00
6	1.010	.95	96.09	88.28	1.000	.84	86.72	83.59	1.073	.97	95.31	95.31
7	1.023	.94	91.41	91.41	1.051	.92	96.88	97.66	1.096	1.04	100.00	100.00
8	1.810	.96	100.00	100.00	1.250	.95	97.66	98.44	1.016	.97	93.44	98.44
9	1.003	1.01	100.00	100.00	1.012	.96	96.88	95.31	1.008	.95	96.09	96.09
10	1.007	.98	100.00	100.00	1.004	1.04	99.22	100.00	1.015	1.00	92.19	92.97
SAMPLE MEAN	1.164	.98	95.55	94.92	1.056	.97	95.70	95.55	1.044	.98	97.27	97.27
SAMPLE S.D.	.253	.02	3.88	4.37	.076	.06	4.74	5.42	.067	.04	1.37	2.80

Table A-1, continued

n = 128

σ = .200

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.010	.99	100.00	100.00	1.285	.87	96.88	96.09	1.121	1.04	96.88	96.88
2	1.020	.98	100.00	100.00	1.002	.99	100.00	100.00	1.157	.95	100.00	100.00
3	1.031	1.03	91.41	91.41	1.719	.83	92.19	92.19	1.022	1.01	97.66	96.09
4	1.678	1.03	92.19	92.19	1.133	.93	95.31	94.53	1.728	.92	100.00	100.00
5	1.009	.96	85.94	85.16	1.044	.95	90.63	94.53	1.047	1.01	96.09	96.09
6	1.023	.99	100.00	100.00	1.036	1.03	94.53	94.53	1.188	1.05	94.53	94.53
7	1.037	.95	92.19	89.06	1.029	.95	100.00	100.00	1.284	.92	97.66	95.31
8	1.192	.93	88.28	87.50	1.446	.96	97.66	100.00	1.017	.94	91.41	91.41
9	5.036	.84	98.44	96.09	1.002	.96	90.63	90.63	1.000	1.01	96.88	96.88
10	1.023	1.02	100.00	100.00	1.076	.93	96.88	96.09	1.300	.90	100.00	98.44
SAMPLE MEAN	1.500	.98	94.84	94.14	1.177	.94	95.47	95.86	1.186	.97	97.11	96.56
SAMPLE S.D.	1.195	.05	5.17	5.50	.226	.05	3.29	3.13	.208	.05	2.57	2.45

Table A-1, continued

n = 64
σ = .0125

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.001	.92	95.31	93.75	1.008	.07	98.44	96.88	1.052	.62	100.00	85.94
2	1.010	.97	100.00	93.75	1.004	1.09	96.88	98.44	1.000	1.22	98.44	100.00
3	1.000	1.15	100.00	100.00	1.021	.85	95.31	95.31	1.209	.50	95.31	85.94
4	1.004	.88	93.75	93.75	2.220	.00	93.75	.00	1.258	.46	93.75	81.25
5	1.074	.80	93.75	93.75	1.005	.95	100.00	100.00	1.028	1.08	95.31	95.31
6	1.087	.95	100.00	100.00	2.730	.02	98.44	31.25	1.000	.97	98.44	98.44
7	1.236	1.02	93.75	95.31	1.004	1.12	96.88	96.88	1.000	.90	95.31	95.31
8	1.073	1.13	93.75	95.31	1.010	1.01	93.75	93.75	1.020	.80	100.00	96.88
9	1.010	1.05	100.00	100.00	1.034	.67	93.75	92.19	1.090	.58	89.06	85.94
10	1.235	1.06	90.63	93.75	1.047	.90	98.44	98.44	1.342	.52	95.31	87.50
SAMPLE MEAN	1.073	.99	96.09	95.94	1.308	.75	96.56	80.31	1.101	.76	96.09	91.25
SAMPLE S.D.	.087	.10	3.37	2.72	.595	.39	2.19	33.16	.118	.25	3.14	6.26

σ = 0.250

REPL	Case 1 ISUBV	Case 1 VRATIO	Case 1 PCI 95	Case 1 CI 95	Case 2 ISUBV	Case 2 VRATIO	Case 2 PCI 95	Case 2 CI 95	Case 3 ISUBV	Case 3 VRATIO	Case 3 PCI 95	Case 3 CI 95
1	1.065	.65	100.00	89.06	1.000	1.06	98.44	98.44	1.348	.71	100.00	95.31
2	1.058	1.10	96.88	100.00	1.073	.96	92.19	92.19	1.007	.78	100.00	93.75
3	1.347	.91	100.00	96.88	1.088	.68	98.44	89.06	1.023	.95	96.88	96.88
4	1.069	1.09	95.31	98.44	2.630	.00	98.44	.00	1.084	.74	96.88	96.88
5	1.064	1.07	92.19	93.75	1.143	.58	98.44	87.50	1.046	.74	90.63	90.63
6	1.178	.88	95.31	93.75	1.025	.82	93.75	90.63	1.019	.83	96.88	95.31
7	1.002	1.05	100.00	100.00	2.209	.00	93.75	.00	1.014	.74	90.63	89.06
8	1.007	.96	93.75	92.19	1.050	.92	95.31	89.06	1.007	.76	92.19	84.38
9	1.139	.91	100.00	100.00	1.051	.93	95.31	95.31	1.018	.87	96.88	93.75
10	1.005	1.00	100.00	100.00	1.050	.53	90.63	84.38	1.024	.91	96.88	93.75
SAMPLE MEAN	1.094	.98	97.34	96.41	1.332	.65	95.47	72.66	1.059	.80	95.78	92.97
SAMPLE S.D.	.100	.09	2.89	3.77	.553	.36	2.75	36.52	.099	.08	3.28	3.71

Table A-2. ISUBV, VRATIO, PCI 95 and CI 95 for n = 64, σ = .0125, .025, .05, .1 and .2 for 10 replications of cases 1, 2 and 3.

n = 64

σ = .0500

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.101	.87	100.00	96.88	1.041	.84	98.44	98.44	1.008	1.18	100.00	100.00
2	1.027	.97	87.50	89.06	1.053	.96	98.44	100.00	1.000	.95	95.31	93.75
3	1.062	.81	85.94	85.94	1.235	.90	100.00	100.00	1.015	.99	93.75	93.75
4	1.000	.90	96.88	90.63	2.219	.20	95.31	62.50	1.371	.41	98.44	79.69
5	1.049	1.08	100.00	100.00	1.047	.95	95.31	92.19	1.337	.59	96.88	89.06
6	1.008	1.05	96.88	98.44	1.000	1.03	96.88	98.44	1.076	.96	96.88	96.88
7	1.076	.93	96.88	96.88	1.000	.86	93.75	92.19	1.004	.64	92.19	89.06
8	1.096	.95	93.75	93.75	1.005	.91	100.00	98.44	1.527	.55	96.88	87.50
9	1.014	1.00	100.00	100.00	1.013	1.01	100.00	100.00	1.013	1.05	100.00	100.00
10	1.019	1.08	100.00	100.00	1.026	.95	98.44	98.44	1.018	.92	96.88	96.88
SAMPLE MEAN	1.045	.97	95.78	95.16	1.164	.86	97.66	94.06	1.137	.82	96.72	92.66
SAMPLE S.D.	.035	.08	4.94	4.81	.358	.23	2.13	10.89	.167	.24	2.36	6.05

σ = .1000

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.968	.51	95.31	93.63	1.023	.90	96.88	96.88	1.002	.98	93.75	100.00
2	1.611	.99	100.00	100.00	1.134	.85	95.31	95.31	1.140	.95	96.88	100.00
3	1.652	.82	96.88	96.88	1.000	1.02	95.31	95.31	1.004	.86	90.63	90.63
4	2.100	.76	98.44	93.75	1.070	.96	100.00	100.00	2.660	.23	98.44	60.94
5	3.363	1.05	100.00	100.00	1.001	.81	90.63	89.06	1.072	.89	96.88	96.88
6	1.024	.89	90.63	89.06	1.012	1.03	96.88	100.00	1.084	.67	98.44	79.69
7	1.003	.91	84.38	90.63	1.024	1.14	100.00	100.00	1.141	.93	90.63	96.88
8	3.262	.65	93.75	90.63	1.089	.95	100.00	100.00	1.017	.85	85.94	84.38
9	1.034	.92	82.81	87.50	1.057	1.00	96.88	96.88	1.010	1.00	93.75	95.31
10	1.109	.97	93.75	90.63	1.081	1.05	95.31	96.88	1.023	.95	100.00	98.44
SAMPLE MEAN	1.832	.86	93.59	92.97	1.049	.97	96.72	97.03	1.215	.83	94.53	90.31
SAMPLE S.D.	.841	.14	5.74	4.26	.042	.09	2.75	3.24	.484	.22	4.21	11.73

Table A-2, continued

n = 64

σ = .2000

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.088	1.00	100.00	100.00	1.005	1.05	100.00	100.00	1.006	.98	100.00	100.00
2	1.239	1.01	95.31	98.44	1.010	.99	96.88	96.88	1.183	.82	95.31	92.19
3	1.052	.94	100.00	98.44	1.082	1.05	100.00	100.00	4.089	.16	98.44	60.94
4	1.095	1.09	100.00	100.00	1.137	.98	100.00	100.00	1.029	.85	96.88	92.19
5	1.277	1.02	95.31	95.31	1.004	.95	100.00	95.31	1.746	.68	95.31	93.75
6	1.136	1.09	100.00	100.00	1.227	.90	100.00	100.00	1.000	1.02	100.00	100.00
7	1.021	.99	100.00	95.31	1.002	1.09	98.44	100.00	1.153	1.16	100.00	100.00
8	1.151	1.08	100.00	100.00	1.000	.96	95.31	96.88	1.007	1.00	95.31	95.31
9	1.238	.87	93.75	89.06	1.047	1.09	100.00	100.00	1.050	1.06	95.31	98.44
10	2.865	.81	95.31	93.75	1.101	1.02	100.00	100.00	1.201	.97	96.88	98.44
SAMPLE MEAN	1.324	.99	97.97	97.03	1.062	1.01	99.06	98.91	1.446	.87	97.34	93.12
SAMPLE S.D.	.521	.09	2.52	3.46	.072	.06	1.59	1.72	.906	.27	1.98	11.14

Table A-2, continued

n = 32

σ = .0125

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.015	1.55	100.00	100.00	1.389	.24	90.63	75.00	1.061	.00	96.88	.00
2	1.013	.99	90.63	96.88	1.234	.00	84.38	.00	1.491	.00	96.88	.00
3	1.109	.76	96.88	78.13	1.504	.35	96.88	71.88	1.329	.00	100.00	.00
4	1.039	.83	93.75	90.63	1.601	.00	93.75	.00	1.931	.15	96.88	40.63
5	1.022	.62	93.75	84.38	1.302	.00	90.63	.00	1.096	.41	84.38	81.25
6	1.034	.76	100.00	93.75	1.176	.00	93.75	.00	1.293	.00	96.88	.00
7	1.959	.23	93.75	65.63	1.000	.00	84.38	.00	1.543	.00	90.63	.00
8	1.200	.57	96.88	84.38	1.093	.29	96.88	75.00	1.431	.00	96.88	.00
9	1.019	1.06	93.75	93.75	1.039	.68	96.88	93.75	1.654	.00	93.75	.00
10	1.000	.56	93.75	91.25	1.000	.00	96.88	.00	1.661	.00	90.63	.00
SAMPLE MEAN	1.141	.79	95.31	86.87	1.242	.16	92.50	31.56	1.449	.06	94.37	12.19
SAMPLE S.D.	.279	.34	2.38	9.76	.206	.22	4.68	39.04	.254	.13	4.38	26.01

σ = .0250

REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.410	.87	100.00	100.00	1.600	.00	96.88	.00	1.249	.49	93.75	87.50
2	1.013	.89	87.50	90.63	1.860	.00	96.88	.00	1.169	.17	93.75	68.75
3	1.002	1.11	100.00	96.88	1.018	.99	100.00	96.88	1.059	.44	96.88	84.38
4	1.442	.70	100.00	87.50	1.013	.78	93.75	93.75	1.071	1.02	93.75	96.88
5	1.031	.95	87.50	90.63	1.232	.21	87.50	62.50	1.154	.05	90.63	46.88
6	1.109	1.06	90.63	96.88	1.133	.53	90.63	87.50	1.090	.16	96.88	43.75
7	1.055	1.17	100.00	100.00	1.042	.42	90.63	81.25	1.002	.87	96.88	96.88
8	1.124	1.48	100.00	100.00	1.149	.09	100.00	100.00	1.188	.37	96.88	87.50
9	1.120	.03	96.88	.00	1.209	.24	96.88	59.38	1.287	.09	96.88	37.50
10	1.073	.98	96.88	96.88	1.387	.49	100.00	90.63	1.075	.63	93.75	90.63
SAMPLE MEAN	1.331	.92	95.94	85.94	1.272	.45	95.31	67.19	1.134	.45	95.00	74.06
SAMPLE S.D.	.619	.37	5.05	28.95	.262	.33	4.25	35.96	.085	.31	2.07	21.92

Table A-3. ISUBV, VRATIO, PCI 95 and CI 95 for n = 32, σ = .0125, .025, .05, .1 and .2 for 10 replications of cases 1, 2 and 3.

n = 32

σ = .0500

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.218	.54	96.88	84.38	1.318	.00	93.75	.00	1.023	.79	87.50	93.75
2	1.195	1.12	100.00	100.00	1.000	1.48	100.00	100.00	2.448	.15	96.88	56.25
3	1.004	1.17	90.63	100.00	1.029	.76	93.75	87.50	1.005	1.17	93.75	100.00
4	4.219	.00	87.50	.00	1.001	.77	90.63	90.63	1.244	.39	90.63	78.13
5	1.856	.61	93.75	90.63	1.035	.59	93.75	90.63	1.069	.47	100.00	84.38
6	1.069	1.11	90.63	90.63	1.053	.56	90.63	81.25	2.105	.00	90.63	.00
7	1.314	1.14	90.63	93.75	1.030	1.26	87.50	100.00	1.083	1.28	87.50	96.88
8	2.953	.15	93.75	56.25	1.223	1.29	84.38	93.75	1.045	.89	96.88	90.63
9	1.351	.89	78.13	78.13	1.240	.58	100.00	93.75	2.643	.00	87.50	.00
10	3.317	.14	100.00	46.88	1.021	.57	100.00	90.63	1.062	.20	90.63	56.25
SAMPLE MEAN	1.950	.69	92.19	74.08	1.095	.79	93.44	82.81	1.473	.53	92.19	65.63
SAMPLE S.D.	1.076	.44	6.13	29.91	.112	.42	5.13	28.10	.621	.45	4.25	35.88

σ = .1000

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.451	.63	100.00	84.38	1.173	.51	100.00	78.13	1.046	.67	96.88	96.88
2	1.625	1.26	96.88	100.00	1.004	.95	93.75	93.75	1.272	.29	96.88	68.75
3	1.049	.52	87.50	81.25	1.001	.80	93.75	93.75	1.014	1.38	96.88	100.00
4	1.340	1.06	100.00	100.00	1.107	1.16	93.75	93.75	1.161	.89	100.00	100.00
5	1.065	1.07	100.00	100.00	1.205	.93	100.00	100.00	1.244	.34	100.00	62.50
6	1.129	1.08	100.00	100.00	1.050	.83	90.63	90.63	1.178	.94	96.88	87.50
7	1.044	1.04	100.00	100.00	1.063	.50	93.75	81.25	1.018	.60	93.75	84.38
8	1.048	.97	100.00	100.00	1.032	1.28	100.00	100.00	1.019	.73	90.63	90.63
9	1.205	.79	100.00	100.00	1.017	1.05	100.00	96.88	1.001	.98	96.88	96.88
10	1.014	1.00	100.00	100.00	1.044	1.00	93.75	93.75	1.192	.62	100.00	90.63
SAMPLE MEAN	1.200	.94	98.44	96.56	1.070	.90	95.94	92.19	1.114	.74	96.88	87.81
SAMPLE S.D.	.202	.32	3.76	6.91	.067	.24	3.44	6.88	.100	.30	2.80	12.22

Table A-3, continued

n = 32

σ = .2000

	Case 1				Case 2				Case 3			
REPL	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95	ISUBV	VRATIO	PCI 95	CI 95
1	1.331	1.01	100.00	100.00	1.293	.57	93.75	93.75	1.055	1.18	87.50	96.88
2	1.307	1.04	81.25	87.50	1.007	1.20	100.00	100.00	1.117	.75	96.88	96.88
3	1.016	.96	100.00	100.00	1.171	.48	93.75	87.50	1.477	1.14	100.00	100.00
4	1.021	1.17	90.63	100.00	1.027	.77	90.63	90.63	1.469	.66	100.00	93.75
5	1.019	1.13	100.00	100.00	1.045	.77	90.63	87.50	1.223	1.25	93.75	100.00
6	1.243	1.05	87.50	90.63	1.047	.83	87.50	87.50	9.703	.00	100.00	.00
7	3.762	.52	87.50	78.13	1.052	.81	100.00	96.88	1.031	.54	100.00	68.75
8	1.241	.85	78.13	65.63	1.012	.79	90.63	81.25	1.002	.91	100.00	100.00
9	1.011	.95	90.63	90.63	1.009	.64	90.63	84.38	1.039	.61	93.75	84.38
10	1.065	.98	100.00	100.00	1.000	.81	100.00	90.63	1.079	.52	100.00	81.25
SAMPLE MEAN	1.402	.97	91.56	91.25	1.066	.77	93.75	90.00	2.017	.76	97.19	82.19
SAMPLE S.D.	.796	.17	7.79	11.08	.089	.18	4.42	5.38	2.568	.36	4.06	29.08

Table A-3, continued

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 645	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BAYESIAN CONFIDENCE INTERVALS FOR THE CROSS VALIDATED SMOOTHING SPLINE		5. TYPE OF REPORT & PERIOD COVERED Scientific Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-K0042
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics 1210 W. Dayton St. Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, N.C. 27709		12. REPORT DATE July 1981
		13. NUMBER OF PAGES 60
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Spline smoothing, cross-validation, confidence intervals		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (see attached sheets)		

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

We consider the model $Y(t_i) = g(t_i) + \varepsilon_i$, $i = 1,2,\ldots,n$, where $g(t)$, $t\varepsilon[0,1]$ is a smooth function and the $\{\varepsilon_i\}$ are independent $N(0,\sigma^2)$ errors with σ^2 unknown. The cross validated smoothing spline will be used to estimate g nonparametrically 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. First, properties of smoothing splines as Bayes estimates are used to derive confidence intervals based on the posterior covariance function of the estimate. To compute the confidence intervals it is necessary to know or to estimate σ^2. We estimate σ^2 here by the residual sum of squares divided by the equivalent degrees of freedom, both of which are determined using the generalized cross validation estimate of the smoothing parameter. A Monte Carlo study is carried out to suggest by example to what extent the resulting 95% confidence intervals can be expected to cover about 95% of the true (but in practice unknown) values of $g(t_i)$, $i = 1,2,\ldots,n$. Three smooth example functions, 5 values of σ^2, and n = 32, 64 and 128 were tried. Confidence intervals based on known σ^2 were extremely reliable for all 3 n's, generally covering close to 95% of the true $\{g(t_i)\}$. Confidence intervals based on estimated σ^2's were also highly reliable for all n = 128 and most n = 64 examples tried. Degraded results were sometimes seen for n = 32. Failure of the method for small n appears to be accompanied by estimates of σ^2 off by orders of magnitude. which would frequently be evident to an experimenter. The method was also applied to one example of a two dimensional thin plate smoothing spline with n = 169, and 162 or 95.8% of the true values were covered by the 95% confidence intervals. 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.