A FREQUENCY INTERPRETATION OF BAYESIAN "CONFIDENCE" INTERVALS FOR SMOOTHING SPLINES.

by

DOUGLAS NYCHKA

Institute of Statistics Mimeograph Series no. 1699.

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

A frequency interpretation of Bayesian "confidence" intervals for smoothing splines.

Douglas Nychka

Department of Statistics

North Carolina State University

Abstract

The frequency properties of Wahba's Bayesian "confidence" intervals for smoothing splines are investigated by a large sample approximation and by a simulation study. When the coverage probabilities for these pointwise confidence intervals is averaged across the observation points, we explain why the average coverage probability (ACP) should be close to the nominal level. From a frequency point of view, the ACP is close to the nominal level because the average posterior variance for the spline is similar to a consistent estimate of the mean squared error and also because the mean squared bias is a modest fraction of the total mean squared error. These properties are independent of the Bayesian assumptions used to derive this confidence procedure and explain why the ACP is accurate for functions that are much smoother than the sample paths prescribed by the prior. The main disadvantage with this approach is that these confidence intervals are only valid in an average sense and may not be reliable if only evaluated at peaks or troughs in the estimate.

This analysis accounts for the choice of the smoothing parameter (bandwidth) using cross validation and in the case of natural splines, we consider an adaptive method for avoiding boundary effects.

Subject Classification: (Amos-mos): Primary 62G05 Secondary 62G15

Key Words and Phrases: Nonparametric regression, cross validation, smoothing spline.

SECTION 1

Consider the additive model:

$$Y_k = f(t_k) + e_k \quad k=1,n$$

where the observation vector $\underset{\sim}{Y}' = (Y_1, Y_2, \ldots, Y_n)$ depends on a smooth, unknown function f evaluated at the points: $0 \leq t_1 \leq t_2 \ldots \leq t_n \leq 1$ and a vector of independent and identically distributed errors: $\underset{\sim}{e}' = (e_1, e_2, \ldots e_n)$ with $E(\underset{\sim}{e}) = \underset{\sim}{0}$ and $E(\underset{\sim}{e}\,\underset{\sim}{e}') = \sigma^2 I$. The statistical problem posed by this model is to estimate f from $\underset{\sim}{Y}$ without having to to assume that f is contained in a specific parametric family. One solution is a smoothing spline estimate for f where the appropriate amount of smoothing is determined by generalized cross validation. Splines have been used successfully in a diverse range of applications and eventually may provide an alternative to standard parametric regression models. One limitation in applying spline methods in practice, however, is the difficulty in constructing confidence intervals or specifying other measures of the estimates accuracy. Wahba (1983), using the elegant interpretation of a smoothing spline as a posterior mean when f is viewed as the realization of a Gaussian stochastic process, suggested a pointwise confidence interval for $f(t_k)$ based on the posterior distribution for f at t_k. An intriguing feature of these confidence intervals is that although they are derived from a Bayesian viewpoint, they work well when evaluated by a frequency criterion. If $C(\alpha,t)$ is the $(1-\alpha)100\%$ Bayesian "confidence" interval for f(t) then Wahba's simulations show that the average coverage probability (ACP):

$$E_{\underset{\sim}{e}} \frac{1}{n} \sum_{k=1}^{n} 1\left[f(t_k)\right]$$

$$\left\{f(t_k) \in C(\alpha, t_k)\right\}$$

is surprisingly close to the nominal level, $1-\alpha$. An interesting twist in Wahba's analysis is that rather than simulating f as a realization of the continuous stochastic process prescribed by the prior, she used only several different fixed functions.

Figures 1a and 1b contrast the difference between these two choices for f. The first is a plot of three realizations of the stochastic process corresponding to Wahba's prior distribution for f. In the second plot are three of the curves used in her simulations. These two groups of curves clearly represent different degrees of smoothness and it may be puzzling why Wahba's method is so successful when the prior distribution for f is a poor reflection of the "true" function.

The goal in this paper is to remove some of the mystery concerning the remarkable simulation results reported by Wahba by giving a frequency interpretation to these confidence intervals. This is accomplished by switching the fixed and random components of this model. Rather than considering a confidence interval for $f(\tau)$ where $f(\cdot)$ is the realization of a stochastic process and τ is fixed we will consider confidence intervals for $f(\tau_n)$ where f is now a fixed function and τ_n is a point randomly selected from $\{t_k\}_{k=1,n}$. From this second point of view, the average coverage probability computed by Wahba in her simulations is really just $P\{f(\tau_n) \in C(\alpha,\tau_n)\}$. Moreover, we show that

$$(1.1) \qquad P\{f(\tau_n) \in C(\alpha,\tau_n)\} = P\left(|\mathcal{U}| \le Z_{\frac{\alpha}{2}}\right) + o(1) \text{ as } n\to\infty$$

where $\mathcal{U}$ is close to a standarized random variable that is the sum of a normal random variable and a discrete random variable related to the bias of spline estimate at the observation points. Z_α is the $1-\alpha$ normal percentile.

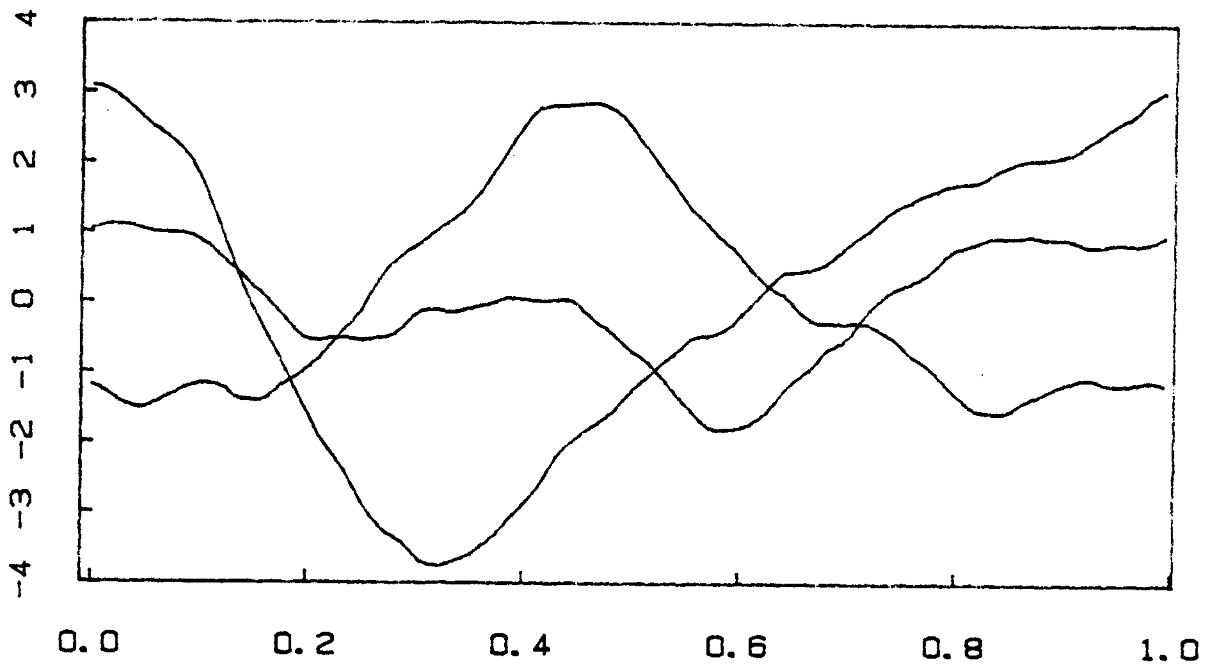

Figure 1a Three realizations from the Bayesian prior for the unknown function. For a periodic second order smoothing spline the prior considered by Wahba is an integrated Brownian bridge adjusted to be a periodic function. To facilitate graphing the sample paths, they have been scaled to come from a Brownian bridge with unit variance and translated to have zero mean. This adjustment does not change the qualitative impression of the local smoothness of these functions.

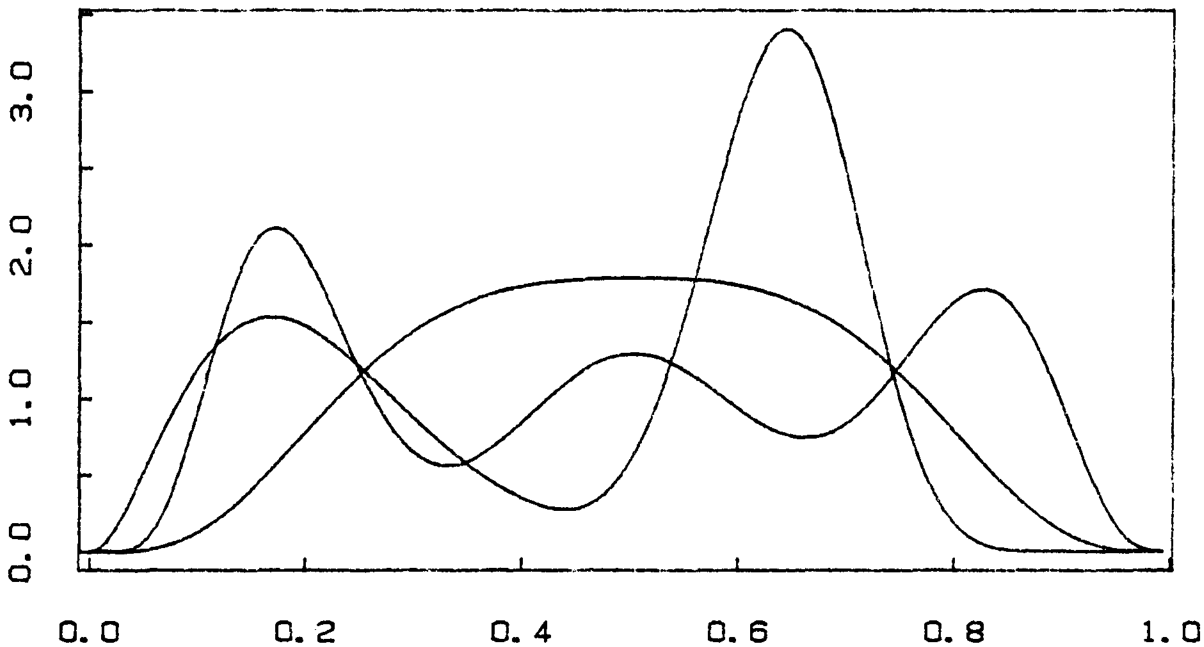

Figure 1b The three smooth test functions used for the simulations in Wahba (1983). These three cases all have at least 1 continuous derivative satisfying periodic boundary conditions and are mixtures of beta densities:

Case I: $\frac{1}{3}\beta_{10,5} + \frac{1}{3}\beta_{7,7} + \frac{1}{3}\beta_{5,10}$

Case II: $\frac{6}{10}\beta_{30,17} + \frac{4}{10}\beta_{3,11}$

Case III: $\frac{1}{3}\beta_{20,5} + \frac{1}{3}\beta_{12,12} + \frac{1}{3}\beta_{7,30}$

where $\beta_{m,n}$ is the standard Beta density function on [0,1].

Although $\mathcal{U}$ is a sum of two random variables, for the cases considered by Wahba, the variance of the normal component is substantially larger than the variance of the component related to the bias. Thus, the agreement of the ACP with the nominal level is a consequence of the fact that the distribution of $\mathcal{U}$ is close to a standard normal distribution.

One advantage of this interpretation is that these confidence intervals may be justified independently of their Bayesian derivation. From a frequency point of view, there are two main factors that contribute to the accuracy of the average coverage probability. The average posterior variance of the cross validated spline is proportional to a consistent estimate of the mean squared error and the mean squared bias is a small fraction of the total mean squared error. Suppose $\hat{f}$ denotes a spline estimate where the smoothing parameter has been determined by cross validation and let $\hat{T}$ be an consistent estimate of the mean squared error. Our results suggest that if the observations are uniformly distributed, if the spline estimate is not influenced by boundary effects and if f is sufficiently smooth, then the intervals:

$$\hat{f}(t_k) \pm Z_{\frac{\alpha}{2}} \sqrt{\hat{T}}$$

will have an ACP close to $1-\alpha$. Moreover, because our analysis does not specifically depend on $\hat{f}$ being a smoothing spline, we suspect that these results may hold in general for non parametric regression type estimators.

This frequency interpretation emphasizes a limitation of these confidence intervals for a fixed function. They may not be reliable at specific points and are only valid when averaged over the knots (see Wahba (1985)). In particular, points where the bias is large will result in a coverage below $1-\alpha$. Figures 2a and 2b illustrate this relationship for one of the cases

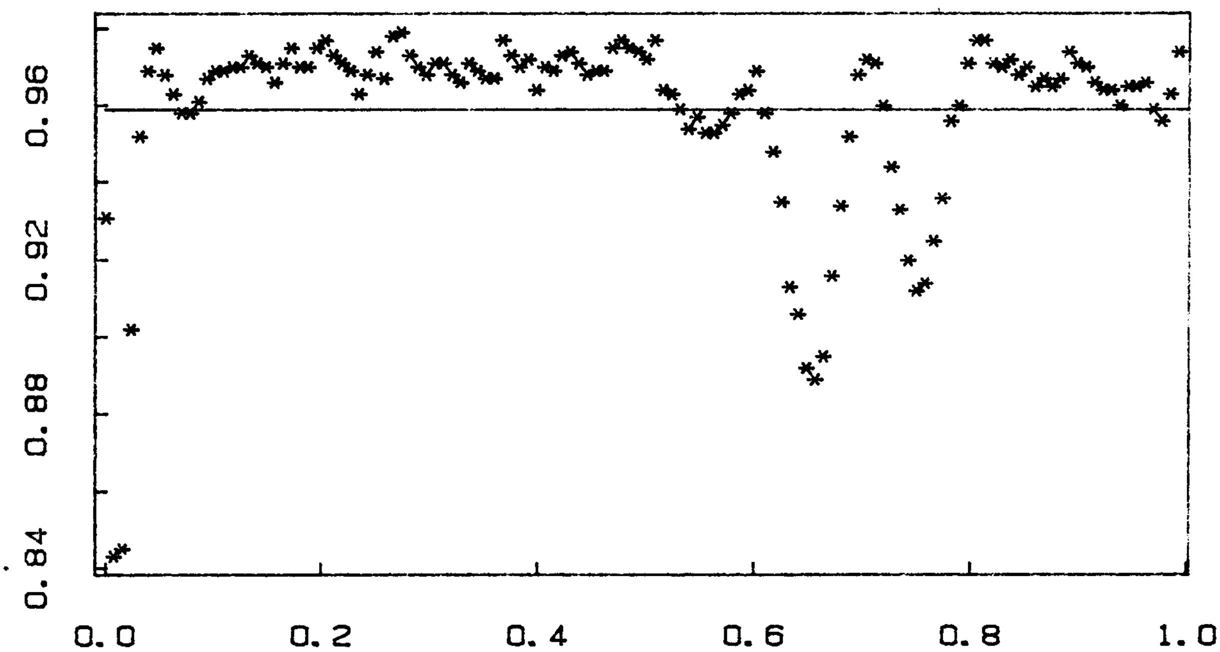

Figure 2a Individual coverage probabilities of Wahba's "Bayesian" confidence intervals. Plotted are the estimated coverage probabilities from a simulation of 1000 trials. The true function is case II with n=128, σ =.05 and a normal distribution for the errors. The smoothing for each estimate was determined by generalized cross validation. The estimated ACP for this case is .959 with a standard error of .0013.

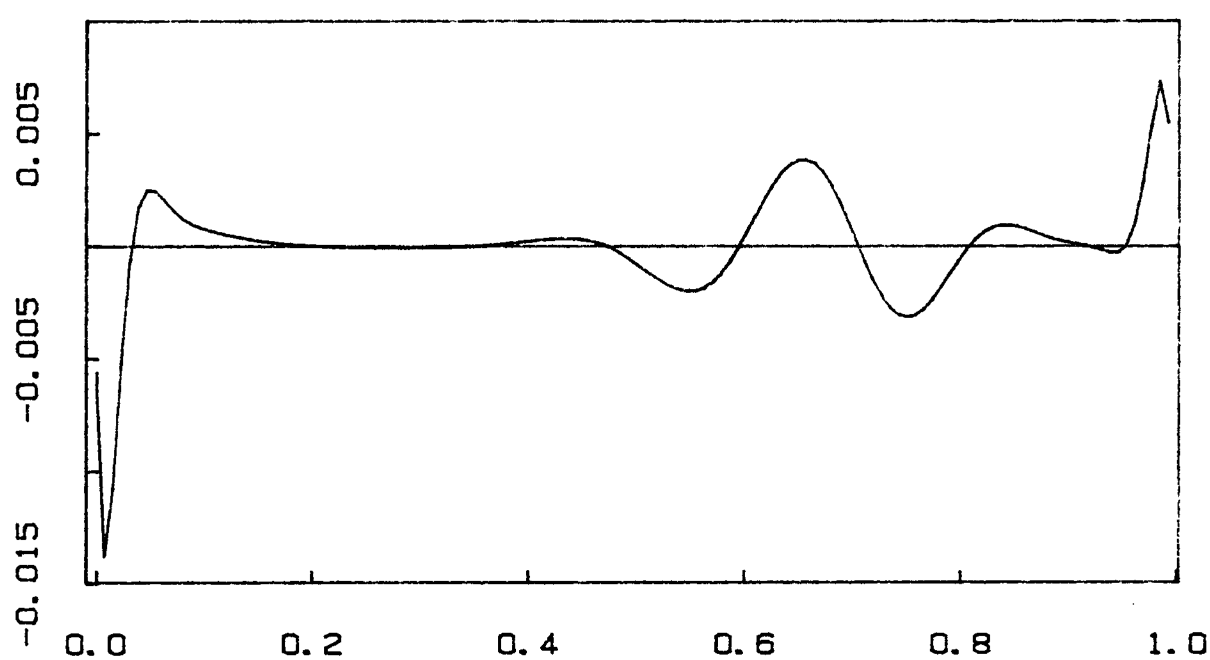

Figure 2b Bias of the spline estimate for the same parameters as in Figure 2a except λ is taken to be the minimizer of the expected mean squared error. (λ^o=4.09E-4). Note that points where the estimate is significantly biased correspond to lower coverage probabilities.

from Wahba's Monte Carlo study. This problem is compounded by the fact that one would expect the bias of f_λ to be large precisely at the points where f has more complicated and interesting structure. Thus, we would not expect this confidence procedure to be reliable if intervals were only computed where f_λ has a peak. Despite this difficulty, this type of interval appears to provide a reasonable measure of the spline estimates accuracy provided that the point for evaluation is chosen independently of the shape of f. For another approach to confidence intervals see Cox (1986).

The next section gives the details of the stochastic interpretation of a smoothing spline and discusses the connection between the average posterior variance and the mean squared error. Section 3 defines the random variable, $\mathfrak{U}$, that is related to the average coverage probability. The distribution of $\mathfrak{U}$ has been computed for the different cases considered by Wahba and found to be very close to a standard normal. These results are also reported in this section. Section 4 explains how the ideas in the previous section generalize to natural smoothing splines and gives some simulation results . As part of this discussion we also suggest a simple adaptive method for avoiding large biases in the estimate at the end points. Section 5 gives a proof of the relationship in (1.1) using some technical results given in an appendix. Our analysis accounts for the fact that the smoothing parameter is determined adaptively by cross validation. At least in this limited context we show that inference based on this estimated smoothing parameter is asymptotically equivalent to the case when the "optimal" value is known.

SECTION 2

One property of a spline that distinguishes it from other non parametric regression estimators is that it is the solution to a variational problem. Let

$$\mathcal{H} = W_2^2[0,1] = \left\{h : h, h' \text{ are absolutely continious and } h'' \in L^2[0,1]\right\} ,$$

then for $\lambda > 0$ the spline estimate, f_λ, is the minimizer of

$$\frac{1}{n} \sum_{k=1}^{n} \left(Y_k - h(t_k)\right)^2 + \lambda \int_{[0,1]} \left(h''(t)\right)^2 dt \tag{2.1}$$

for all $h \in \mathcal{H}$. When λ is fixed, f_λ is a linear function of $\underset{\sim}{Y}$ and it is convenient to define an n×n "hat" matrix, $A(\lambda)$, depending only on $\{t_k\}$ such that

$$\underset{\sim}{f}_\lambda = A(\lambda)\, \underset{\sim}{Y} \quad \text{where } \underset{\sim}{f}_\lambda' = \left\{f_\lambda(t_1),\ f_\lambda(t_2), \ldots, f_\lambda(t_n)\right\} .$$

For details concerning the form of $A(\lambda)$ and its computation see Bates (1985) and Hutchinson and de Hoog (1985).

f_λ is referred to as a natural spline because it satisfies the "natural" boundary conditions:

$$f_\lambda^{(2)}(0) = f_\lambda^{(2)}(1) = 0, \quad f_\lambda^{(3)}(0) = f_\lambda^{(3)}(1) = 0.$$

If $\mathcal{H}$ is taken to be the periodic set of functions:

$$\left\{h \in W_2^2\,[0,1]:\ h^{(j)}(0) = h^{(j)}(1),\ j = 0,1\right\}$$

then f_λ will be a periodic spline satisfying the periodic boundary conditions: $f_\lambda^{(j)}(0) = f_\lambda^{(j)}(1) \quad j = 0,3.$

It is important to distinguish between these two estimates because Wahba's simulation results pertain to periodic splines while natural splines are more likely to be used in applications.

The smoothing parameter, λ, is analagous to the bandwidth in a non-parametric regression kernel estimate and this paper will concentrate on the statistical properties of f_λ when λ is selected by a data-based procedure. Specifically, let

$$V(\lambda) = \frac{\frac{1}{n} \left\| \left[I-A(\lambda)\right] \underset{\sim}{Y} \right\|^2}{\left[\frac{1}{n} \operatorname{tr}(I-A(\lambda)\right]^2}$$

be the generalized cross validation function and let $\hat{\lambda}$ denote the global minimum of V. We will consider the spline estimate: $\hat{f} = f_{\hat{\lambda}}$.

The rationale for Wahba's confidence intervals comes from the correspondence between f_λ and a conditional expectation.
Suppose that f is a sample path from the Gaussian process:

$$(2.2) \qquad f(t) = \alpha_o + \alpha_1 t + \frac{\sigma^2}{n\lambda} \int_o^t (t-s)\, dW(s)$$

where $W(\cdot)$ is the standard Weiner process and $\underset{\sim}{\alpha} \sim N(0, \zeta I)$. With this model for f, a natural spline satisfies the relationships:

$$f_\lambda(t) = \lim_{\zeta \to \infty} E\left(f(t) \mid \underset{\sim}{Y}\right)$$

and

$$\sigma^2 A(\lambda) = \lim_{\zeta \to \infty} \operatorname{cov}\left[\underset{\sim}{f} \mid \underset{\sim}{Y}\right]$$

with $\underset{\sim}{f}' = \left[f(t_1), f(t_2), \ldots, f(t_n)\right]$.

The same conditional moments also hold for the periodic spline if f is assumed to be a sample path from a periodic version of (2.2). This connection lead Wahba to propose

$$\hat{f}(t_k) \pm Z_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2 [A(\hat{\lambda})]_{kk}} \qquad \text{where} \quad \hat{\sigma}^2 = \frac{\left\| \left(I-A(\hat{\lambda})\right) Y \right\|^2}{\mathrm{tr}\left[I-A(\hat{\lambda})\right]}$$

as a $(1-\alpha)$ 100% confidence interval for $f(t_k)$. What is surprising about this confidence procedure is that the ACP is close to $1-\alpha$ even when f is much smoother (has more derivatives) than the process in (2.2).

Part of the reason for the success of these intervals lies in the fact that the standard error is close to the mean squared error. This observation was first made by Wahba (1983) and will be used in the next section to explain the accuracy of the ACP for these intervals. We end this section by briefly discussing this connection. Suppose that $\hat{f}$ is a periodic spline and the observation points are equally spaced. In this case $A(\lambda)$ will be a circulant matrix and thus, $\left[A(\lambda)\right]_{kk} = \frac{1}{n} \mathrm{tr}(A(\lambda))$. If f satisfies the conditions F1-F3 stated in Section 4, then Nychka (1986) shows

$$\frac{\frac{\hat{\sigma}}{n} \mathrm{tr}\, A(\hat{\lambda})}{E\, T_n(\lambda^o)} \xrightarrow{P} \left(\frac{32}{27}\right) \tag{2.3}$$

where $T_n(\lambda) = \frac{1}{n} \Sigma \left(\hat{f}(t_k) - f(t_k)\right)^2$ and λ^o is the global minimum of $E\, T_n(\lambda)$. This result is established by observing that the average posterior variance is proportional to a consistent estimate of $E\, T_n(\lambda^o)$ based on the cross validation function:

$$\hat{T}_n = V(\hat{\lambda}) - \hat{S}_n^2 \tag{2.4}$$

with $$\hat{S}_n^2 = \frac{\left\| \left(I - A(\hat{\lambda})\right) Y \right\|^2}{\mathrm{tr}\left[I - \mathcal{C}\, A(\hat{\lambda})\right]} \quad \text{and} \quad \mathcal{C} = \left(\frac{37}{32}\right) .$$

In general $\hat{T}_n$ will be a consistent estimate of E $T(\lambda^o)$ provided the mean squared bias of the spline estimate is not influenced by effects at the boundaries (see Nychka (1986) for details). In Section 4 we propose to reduce boundary effects by restricting the evaluation of the spline estimate to an interval smaller than range of $\{t_k\}_{k=1,n}$. The estimate of the mean squared error over this interior interval worked well in simulations, and in Section 5 we outline an argument for its consistency.

SECTION 3

In this section we will focus on the case when f_λ is a periodic spline and the knots, $\{t_k\}_{k=1,n}$ are equally spaced. This case is considered for two reasons. First, it is the estimate computed by Wahba in her Monte Carlo study. Secondly, in this simple case, the frequency interpretation of these confidence intervals is the least complicated. A general discussion for the natural spline estimate will be given in section 4.

The random variable $\mathfrak{U}$ in (1.1) will be defined for f_λ evaluated at an "optimal" value of the smoothing parameter. Set $f^o = f_{\lambda^o}$. Then f^o is the spline that minimizes the expected mean squared error and has the advantage that λ^o is not a random quantity. This property makes it easy to decompose f^o into fixed and random components related to the bias and variance of the estimate. Given this decomposition, it is straightforward to derive a frequency interpretation of the ACP for f^o .

Of course in practice, both $E\ T_n(\lambda)$ and λ^o are unknown and have to be estimated. The main mathematical result of this paper is to show that the difference between the ACP for $\hat{f}$ when λ is determined by generalized cross validation and the ACP for when f^o and λ^o are known converges to zero as the sample size increases.

When f_λ is a periodic spline with equally spaced knots, the diagonal elements will be equal and in this case all the pointwise confidence intervals will have the same width. For the moment, they will be represented as

(3.1) $\quad f^o(t_k) \pm Z_{\frac{\alpha}{2}}\ B.$

Let $b(t) = f(t) - E\ f^o(t)$, $v(t) = \left[f^o(t) - E\ f^o(t)\right]$

and recall that τ_n has a distribution independent of $\underset{\sim}{e}$ putting equal mass on the points $\{t_k\}_{k=1,n}$. These quantities will be used to define $\mathfrak{U}$.

Taking $b = b(\tau_n)$ and $v = v(\tau_n)$

then

$$E\left[\frac{1}{n}\sum_{k=1}^{n} I\left(f^o(t_k)\right)\left\{\left|f^o(t_k) - f(t_k)\right| \le Z_{\frac{\alpha}{2}} B\right\}\right] = P\left[\left|b(\tau_n) + v\,(\tau_n)\right| \le Z_{\frac{\alpha}{2}} B\right]$$

$$= P\left[\left|\frac{b + v}{B}\right| \le Z_{\frac{\alpha}{2}}\right]$$

.

Thus, the ACP for the intervals in 3.1 will be close to $1 - \alpha$ provided that the distribution of $\mathfrak{U} = (b + v) \big/ B$ is close to a standard normal. It is easy to argue that this should be so from the properties of b and v summarized below.

Lemma 3.1

If f_λ is a periodic spline, $t_k = k/_n$ for k=0, n-1, and $\underset{\sim}{e} \sim N\left(0,\sigma^2 I\right)$ then

a) $E(b) = 0$,

b) $v \sim N\left(0, \frac{\sigma^2}{n} \text{tr}\left[A(\lambda^o)^2\right]\right)$

c) b and v are independent

d) $VAR(b + v) = E\, T_n(\lambda^o)$

Proof

a) Because the constant function is in the null space of the penalty function and $A(\lambda^o)$ is symmetric, $\left(I-A(\lambda^o)\right)\underset{\sim}{1} = 0$ and we have

$$E(b) = \frac{1}{n}\sum_{k=1}^{n} (Ef_\lambda(t_k) - f(t_k)) = \frac{1}{n}\underset{\sim}{1}'\left[I-A(\lambda^o)\right]\underset{\sim}{f} = 0 \quad .$$

b) Let $\underset{\sim}{v} = A(\lambda^o)\ \underset{\sim}{e}$ where $\underset{\sim}{v}' = \left[v(t_1),\ldots v(t_n)\right]$. Because $\underset{\sim}{e}$ is normally distributed $\underset{\sim}{v} \sim N\left[0,\sigma^2 A^2(\lambda_o)\right]$. $A^2(\lambda^o)$ will also be a circulant matrix and the diagonal elements can be written as $\frac{1}{n}\ \mathrm{tr}\ A^2(\lambda^o)$.

Thus, $v(t_k) = E\left[v \middle| \tau_n = t_k\right] \sim N\left[0,\ \frac{\sigma^2}{n}\ \mathrm{tr}\ A^2(\lambda^o)\right]$. Because this distribution does not depend on k the unconditional distribution is the same.

c) From the results above it follows that $E\left[\varphi(v) \middle| b\right]$ will not depend on b for any measurable function φ .

d) $$\mathrm{Var}\ (b + v) = E\ (b + v)^2 = \frac{1}{n}\sum_{k=1}^{n} E\left[(b + v)^2 \middle| b = b_k\right]$$

$$= \frac{1}{n}\sum_{k=1}^{n} b\ (t_k)^2 + \frac{1}{n}\ \sigma^2 \mathrm{tr} A^2(\lambda^o)$$

$$= E\ T_n(\lambda^o)$$

If we take $B = \sqrt{E\ T_n(\lambda^o)}$ then from the discussion above, $E(\mathcal{U}) = 0$ and $\mathrm{Var}(\mathcal{U}) = 1$. Also, for the cases considered by Wahba $\mathrm{Var}\ (b) / \mathrm{var}\ (v)$ is less than .25 . The distribution of $\mathcal{U}$ should be close to a normal since it is the convolution of two independent random variables one of which is normal while the other has a variance that is small relative to the normal component. Figure 3 is a histogram of the distribution of b for the Case II with n = 128 and σ = .05 (see Figure 2b) along with an appropriately scaled version of the normal density for v. When plotted, the distribution function of

$\mathcal{U} = \dfrac{b + v}{\sqrt{E\ T_n(\lambda^o)}}$, cannot be distinguished from a standard normal.

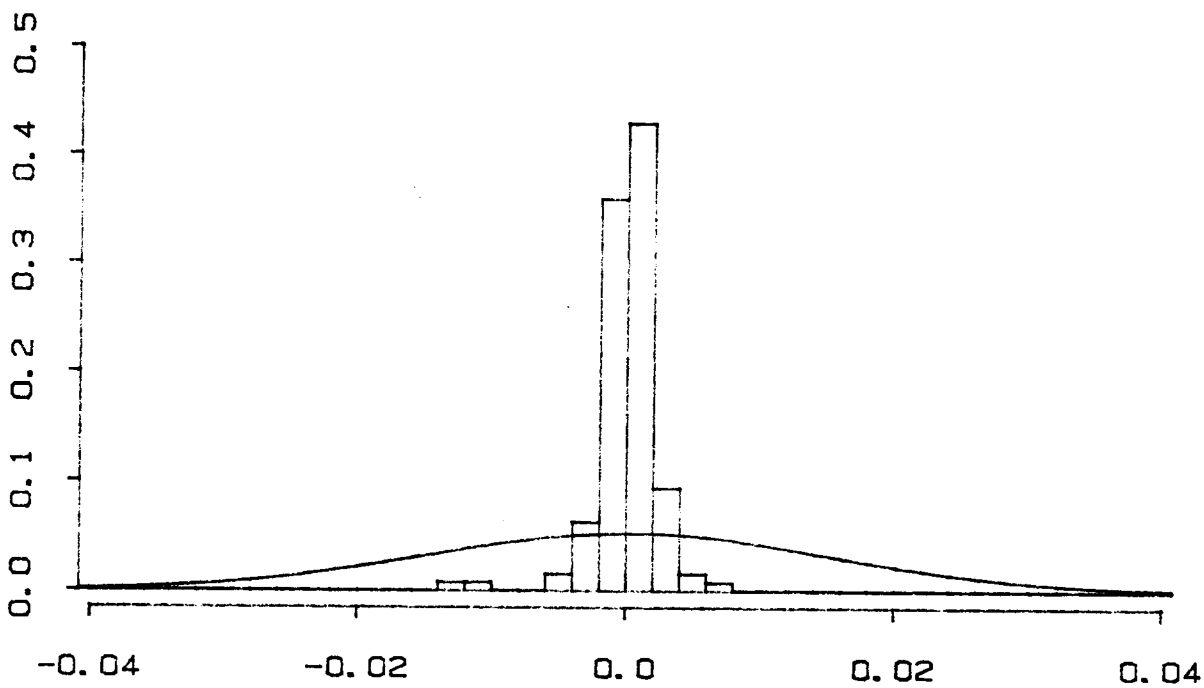

Figure 3 Comparison of the densities of b and v.

The histogram represents the discrete distribution of b for the same parameters as in Figure 2b. For this case var(b) = 5.54E-5. The random variable v is distributed $N\left(0, \frac{1}{n} \text{ tr } A^2(\lambda^o)\right)$ where $\frac{1}{n}$ tr $\left(A^2(\lambda^o)\right)$ = 3.58 E-4. This normal density is included in the figure, appropriately scaled by the histogram bin width. The convolution of these two distributions yields a distribution very close to a normal.

Table 1 reports the maximum absolute difference between the distribution function of $\mathcal{U}$ and a standard normal over the range of cases considered by Wahba. The ratio of $\frac{\text{var}(b)}{\text{var}(\nu)}$ is also given. We note that the ratios $\frac{\text{var}(b)}{\text{var}(\nu)}$ in Table I are in rough agreement with asymptotic theory. If f is a m^{th} order smoothing spline such that

$$\frac{1}{n} \sum_{k=1}^{n} \left[f(t_k) - E\, f_\lambda(t_k) \right]^2 = \gamma\, \lambda^{2\alpha} \left[1 + o(1) \right]$$

for some $\gamma > 0$ as $\lambda \to 0$ then a simple calculation yields (see Lemma 3.1 Nychka (1986)) $\frac{\text{var}(b)}{\text{var}(\nu)} = \frac{1}{4\alpha m} \left(1 + o(1) \right)$ where m refers to the number of derivatives in the penalty function. For the smooth test functions considered by Wahba $\alpha = 1$ for cases I and III because these functions have 3 periodically continuous derivatives and $f^{(4)} \in L^2\,[0,1]$. Thus, with $m = 2$ we would expect $\frac{\text{var}(b)}{\text{var}(\nu)} \approx .125$. For case II $\alpha = ¼$ because this function does not have a periodic second derivative and the ratio of variances should be compared to .25. The difference between these actual fractions and the asymptotic values may be due to the fact that the beginning Fourier coefficients for these functions decrease at a different rate than what would be expected in the limit.

For a periodic spline, Wahba's confidence procedure suggests that B should be equal to $\frac{\sigma^2}{n} \text{tr}\, A(\lambda^o)$, however, at least from a frequency point of view, $B = E\, T_n(\lambda^o)$ appears to be a more defensible choice because it standardizes the variance of $\mathcal{U}$. Using $\frac{\hat{\sigma}^2}{n} \text{tr}\, A(\hat{\lambda})$ should result in slightly conservative intervals with an ACP slightly higher than $1-\alpha$. This effect is apparent in the original simulations done by Wahba because the ACP tends to be above .95. In fact, the practical difference between these two choices for B is slight as the resulting average coverage probabilities differ by less than a few percent.

Table 1 Comparison of the distribution of $\mathfrak{U}$ with a standard normal.

True function	n	σ = .0125 %K-S	σ = .0125 var(b)/var(v)	σ = .05 %K-S	σ = .05 var(b)/var(v)	σ = .2 %K-S	σ = .2 var(b)/var(v)
	64	.06	.14	.10	.17	.16	.22
Case I	128	.09	.14	.06	.16	.17	.21
	256	.11	.14	.04	.15	.16	.19
	64	.44	.16	.16	.15	.06	.16
Case II	128	.70	.17	.27	.16	.04	.15
	256	.86	.18	.37	.16	.07	.15
	64	.13	.13	.12	.15	.09	.19
Case III	128	.14	.13	.13	.14	.08	.18
	256	.14	.13	.13	.14	.10	.17

Note: $\%K\text{-}S = 100 \times \left\{\sup_z \left| P(\mathfrak{U} < z) - \Phi(z)\right|\right\}$

where Φ is the standard normal c.d.f.

$\mathfrak{U} = \dfrac{b + v}{\sqrt{\operatorname{var}(b+v)}}$ where $v \sim N\left(0, \frac{\sigma^2}{n} \operatorname{tr} A^2(\lambda^o)\right)$

and b has a discrete distribution taking on the values

$\left\{E f^o(t_k) - f(t_k)\right\}$, $k = 1,n$ with equal probability.

$\operatorname{var}(b + v)$ is equal to the expected mean squared error of the spline estimate.

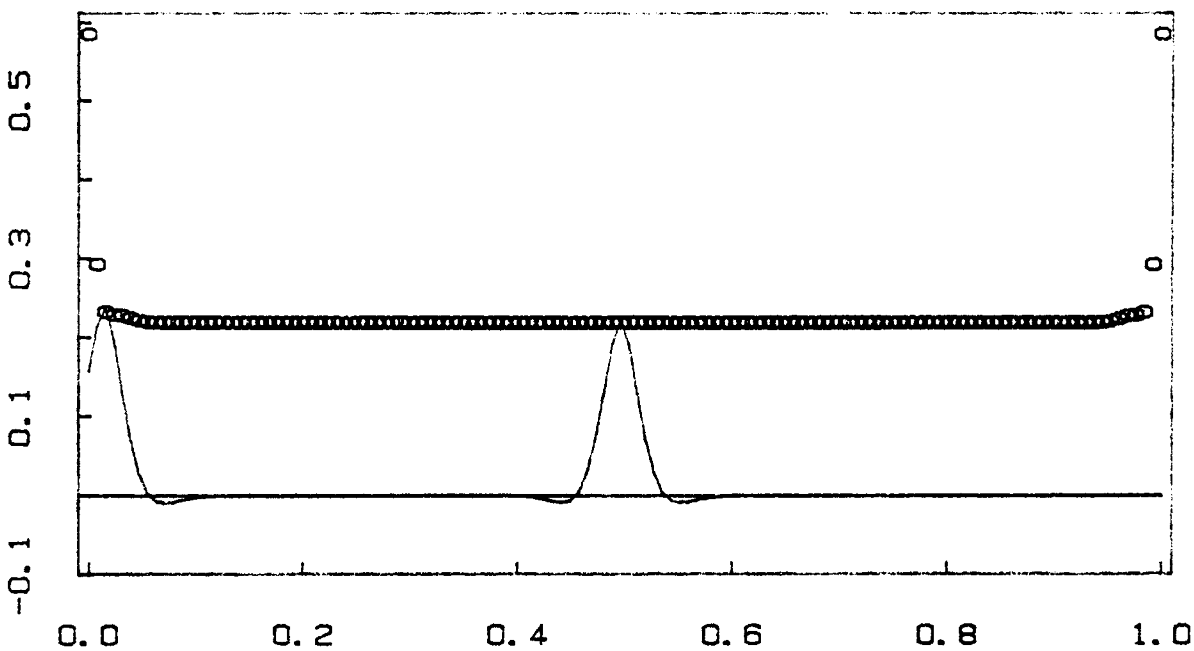

Figure 4 Influence of the boundaries on the matrix $A(\lambda)$.
Plotted points are $A_{kk}(\lambda)$ verses $t_k = \frac{k}{n}$ when n=128 and λ = 4.09 E-4 (the optimal choice for case II when σ =.05). To give an idea to the extent that $f_\lambda(t_k)$ is a locally weighted average, $A_{3k}(\lambda)$ and $A_{64k}(\lambda)$ are plotted versus t_k. (The points are connected by lines to make the weight functions easier to distinguish). Note that the weight function close to zero is skewed but for this case if $3 \leq k \leq 126$ then

$$\frac{A_{kk}(\lambda)}{\inf A_{jj}(\lambda)} \leq 1.1 .$$

SECTION 4

So far the discussion has been limited to periodic splines where f must satisfy a specific set of periodic boundary conditions. Although this narrow scope is adequate to analyze Wahba's Monte Carlo study these restrictions on the estimator and f limit its usefulness. In this section we generalize the frequency interpretation of these Bayesian confidence intervals to a natural spline estimate and give a straight forward solution to the problem of boundary effects.

To make the presentation less complicated it will be assumed that the observation points are uniformly distributed. It is possible to extend these ideas to other distributions but this is beyond the scope of this paper.

The main difficulty in generalizing the ideas in the previous section from periodic splines to natural splines is that the diagonal elements of $A(\lambda)$ will no longer be equal. This complicates the definition of v because the random variable $v\left(t_k\right)$, must now depend on k and hence will not be independent of b. For uniformly distributed $\left\{t_k\right\}$ the variation in $A_{kk}(\lambda)$ reflects the fact that f can be estimated more accurately in the middle of the interval compared to the endpoints (see Figure 4). (Also, if the observation points were not distributed uniformly these diagonal elements would adjust for the relative accuracy of $\hat{f}$ due to the variation in the density of $\left\{t_k\right\}$.)

Another problem with natural splines is that the mean squared error (and hence the cross validation function) may be dominated by the bias of the estimate in small neighborhoods of 0 and 1 (Rosenblatt and Rice (1983) and Messer (1986)). The diagonal elements of $A(\lambda)$ do not adjust for this effect and in fact the ACP will approach 1 as $n \to \infty$ because the mean squared error for almost all the points in the interval will be negligible with respect to the total mean squared error.

A simple way to deal with both of these problems is to restrict the evaluation of the spline to a sub interval of $[0,1]$:

$$\mathcal{D} = \left[\delta_1,\ 1-\delta_2\right] \text{ for } \delta_1,\ \delta_2 > 0 .$$

If $\mathcal{D}$ remains fixed as the sample size increases then $f_\lambda(t)$ for $t \in \mathcal{D}$ will not be influenced by boundary effects. Moreover, if the observation points are uniformly distributed then the diagonal elements, $A_{kk}(\lambda)$, will converge uniformly to a function independent of k for $t_k \in \mathcal{D}$. Thus, at least for large sample sizes a natural spline estimate restricted to an interior interval will have similar properties to a periodic spline.

The frequency interpretation for confidence intervals is also similar. Because $\mathcal{D}$ may be chosen so that the diagonal elements of $A(\lambda)$ for $t_k \in \mathcal{D}$ are essentially constant, we consider pointwise confidence intervals of equal width. Let $n_\delta = \#\left\{t_k \in \mathcal{D}\right\}$, $T_\delta(\lambda) = \frac{1}{n} \sum_{t_k \in \mathcal{D}} \left(\hat{f}_\lambda(t_k) - f(t_k)\right)^2$ and let λ_δ^o be the minimizer of $E\, T_\delta(\lambda)$. Following the development in Section 3, the ACP for the intervals:

$$f_{\lambda_\delta^o}(t_k) \pm Z_{\frac{\alpha}{2}} B \quad \text{for } t_k \in \mathcal{D}$$

is given by $P\left(\left|\mathcal{U}_\delta\right| \le Z_{\frac{\alpha}{2}}\right)$ where now

$$\mathcal{U}_\delta = \frac{\hat{f}_{\lambda_\delta^o}(\tau_n) - f(\tau_n)}{B}$$ and τ_n assumes the values $t_k \in \mathcal{D}$ with equal probability.

Similar to the periodic case $\mathcal{U}_\delta$ can be decomposed into the sum of a discrete and a normal random variable. Because $A(\lambda)$ is not circulant, however, these two components will not be independent. Also, the expectation of $\mathfrak{U}_\delta$ is not identically equal to zero because of the restriction of $\hat{f}$ to points in $\mathfrak{D}$. Note that $E\left(\mathfrak{U}_\delta^2\right) = E\ T\left(\lambda_\delta^o\right)$ and thus in order to standardize $\mathfrak{U}_\delta$ a natural choice for B is $\sqrt{E\ T_\delta\left(\lambda_\delta^o\right)}$. Despite these departures from the exact relationships in Section 3, the distribution of $\mathfrak{U}_\delta$ was found to be close to a standard normal with this choice for B. Before discussing these results, however, an adaptive method for determining $\mathfrak{D}$ will be presented.

For $\rho > 1$ let $\mathfrak{D}(\lambda)$ be the smallest interval containing

$$\left\{t_j : \alpha_j(\lambda) \leq \rho\right\} \quad \text{where } \alpha_j(\lambda) = \frac{A_{jj}(\lambda)}{\inf\limits_{1 \leq k \leq n} A_{kk}(\lambda)}$$

We will use $\mathfrak{D}(\lambda^o)$ to evaluate the distribution of $\mathfrak{U}_\delta$. Note that two values of λ are entered into this computation: the interval for evaluation depends on λ^o and the spline estimate on this interval uses λ_δ^o . Initially, we considered using a single value of λ defined as the minimizer of

$$E\left[\frac{1}{n_\delta} \sum_{t_k \varepsilon\, \mathfrak{D}(\lambda)} \left(f(t_k) - f_\lambda(t_k)\right)^2\right] \quad .$$

Unfortunately, this criterion tends to have multiple minima. Also, the estimate for this value of the smoothing parameter using cross validation had more variability than the two step procedure described below. In either case, the choice of $\mathfrak{D}$ is not completely objective because ρ must be specified. In fact, for boundary effects to be asymptotically negligible we must have $\rho \to 1$ at a rate that can be deduced from the functional form of $A_{kk}(\lambda)$ and the convergence rates of λ^o and λ_δ^o .

Table 2 compares the distribution of $\mathfrak{U}_\delta$ when $B = \sqrt{E\ T_\delta(\lambda_\delta^o)}$ to a standard normal in a format similar to Table 1 except that n_δ / n is also included. In these computations $\rho = 1.1$, that is, the diagonal elements of $A(\lambda)$ in $\mathfrak{I}(\lambda^o)$ differ by at most 10%. Surprisingly, this restriction does not exclude many data points. In order to see how much boundary effects influence the distribution of $\mathfrak{U}_\delta$, we also considered a version of case III where the function has been shifted periodically so that the maximum of the absolute second derivative is at zero $\left(f_{shifted}(t) = f\ (t + .164 \mod 1)\right)$. According to the asymptotic theory, a large second derivative at zero should induce large biases at this endpoint. Surprisingly, for the ranges of n and σ considered this change in the function only increases the Kolomogorov-Smirnov distance by at most 1%. The last case in this table uses the same shifted function but no boundary correction. That is, $\mathfrak{I} = [0,1]$. Once again, we see little difference between the distribution of $\mathfrak{U}$ and a standard normal.

We end this section by summarizing the results of a simulation study when λ^o, λ_δ^o and $E\ T\ (\lambda_\delta^o)$ are estimated. These estimates require the use of cross validation twice: first to determine $\mathfrak{I}$ and then to estimate λ_δ^o. Let $\hat{\lambda}$ be the minimizer of the generalized cross validation function given in Section 2 and take the interval for evaluation to be $\mathfrak{I}(\hat{\lambda})$. Let $\hat{\lambda}_\delta$ be the minimizer of the restricted cross validation function:

$$(4.1) \qquad \frac{1}{n_\delta} \sum_{t_k \in \mathfrak{I}(\hat{\lambda})} \left[\frac{Y_k - f_k\ (t_k)}{1 - A_{kk}(\lambda)} \right]^2$$

and the natural spline estimate is taken to be $\hat{f} = f_{\hat{\lambda}_\delta}$. $\hat{f}$ is computed using the full set of data, however, cross validation is only applied to estimates where $t_k \in \mathfrak{I}\ (\hat{\lambda})$. Besides this restriction, (4.1) is the usual

Table 2 Comparison of the distribution of $\mathfrak{U}_\delta$ with a standard normal for a natural spline estimate restricted to a subinterval.

True function	n	$\sigma = .0125$ %K-S			$\sigma = .05$ %K-S			$\sigma = .2$ %K-S		
		%K-S	$\frac{\mathrm{var}(b)}{\mathrm{var}(v)}$	$\frac{\%n_\delta}{n}$	%K-S	$\frac{\mathrm{var}(b)}{\mathrm{var}(v)}$	$\frac{\%n_\delta}{n}$	%K-S	$\frac{\mathrm{var}(b)}{\mathrm{var}(v)}$	$\frac{\%n_\delta}{n}$
Case I	64	2.34	.11	93.7	2.37	.12	93.8	2.71	.15	87.5
	128	2.20	.12	95.3	2.70	.12	92.2	2.78	.14	89.1
	256	2.26	.11	95.3	2.52	.11	93.8	2.77	.13	90.6
Case II	64	.78	.14	96.9	.77	.13	96.9	.98	.13	93.8
	128	.78	.13	98.4	.71	.12	96.9	.91	.13	95.3
	256	.75	.13	97.7	.70	.12	96.9	.89	.13	95.3
Case III	64	.60	.12	96.9	.79	.12	96.9	1.38	.14	93.8
	128	.74	.12	96.9	.86	.12	96.9	1.26	.13	95.3
	256	.62	.11	97.7	.91	.12	96.9	1.28	.13	95.3
Case III (Shifted)	64	1.67	.14	96.9	1.58	.15	96.9	.59	.21	93.8
	128	1.76	.14	98.4	1.35	.14	96.9	.50	.20	93.8
	256	1.57	.14	97.7	1.30	.14	96.9	.73	.18	94.5
Case III (shifted with no boundary adjustment)	64	.74	.14	100.	.87	.15	100.	.84	.22	100.
	128	1.44	.15	100.	1.17	.15	100.	1.03	.21	100.
	256	1.68	.15	100.	1.31	.15	100.	1.08	.22	100.

Note: $\%\text{K-S} = 100 \times \sup_z \left| P(\mathfrak{U}_\delta < z) - \Phi(z) \right|$

Case III shifted, refers to the function in Case III translated so that the maximum value of the second derivative is at 0.
No boundary adjustment means the spline was evaluated on the entire interval.

"leave one out" procedure that incorporates a shortcut for splines (Craven and Wahba (1979)). $E\,T\,(\lambda^{o}_{\delta})$ can be estimated by generalizing (2.4). First to simplify notation, let W denote a diagonal matrix such that $W_{kk} = 1$ for $t_k \varepsilon\ \mathcal{D}(\hat{\lambda})$ and zero otherwise and let $\mu(\lambda) = \frac{1}{n_\delta}\,\mathrm{tr}\,(W\,A(\lambda))$. With this notation the generalized cross validation function for the interval $\mathcal{D}(\hat{\lambda})$ is:

$$V_\delta(\lambda) = \frac{\frac{1}{n_\delta}\ \| W(I-A(\lambda))Y\|^2}{(1-\mu(\lambda))^2}$$

the estimate of S^2_n is

$$\hat{S}^2_n = \frac{\frac{1}{n_\delta}\ \| W\left(I-A(\hat{\lambda}_\delta)\right)Y \|}{1 - \mathfrak{C}\mu\ (\hat{\lambda}_\delta)}\ ,\ \mathfrak{C} = \frac{37}{27} \tag{4.2}$$

and we have $\hat{T}_\delta = \left(V_\delta\left(\hat{\lambda}_\delta\right) - \hat{S}^2_n\right)\ \varphi\left(\hat{\lambda}_\delta\right)$

The constant $\mathfrak{C}$ in (4.2) is the appropriate adjustment when f has four continuous derivatives. φ is a slight bias correction based on the assumption that $\hat{\lambda}_\delta$ is a consistent estimate of λ^{o}_{δ}. The consistency of $\hat{T}_\delta$ is discussed at the end of the next section while the form for φ is derived in Appendix B.

Table 3 summarizes the results for a simulation study based on the same cases considered in evaluating the distribution of $\mathfrak{U}_\delta$. Each case involved 200 repetitions and 95% confidence intervals were computed of the form:

$$\hat{f}(t_k) \pm 1.96\sqrt{\hat{T}_\delta}\ ,\ t_k \varepsilon\ \mathcal{D}(\hat{\lambda})\ .$$

The first number in this table is the ACP and the

second value is the ratio: $\dfrac{E\left(\hat{T}_\delta\right)}{E\ T_\delta\left(\lambda_\delta^o\right)}$ and the third is the sample

standard deviation for this ratio from the 200 trials. Over the range of cases the ACP remained close to the nomial level, .95 while $\hat{T}_\delta$ appears to have little bias. The smoothing spline and the diagonal elements of $A(\lambda)$ were calculated using the order N algorithm in Hutchinson and de Hoog (1985). The minimizations of $V(\lambda)$ and (4.1) were carried out with respect to log (λ) using a coarse grid search followed by a golden section search.

The first number in this table is the percent ACP, the

second value is the ratio: $\dfrac{E\left(\hat{T}_\delta\right)}{E\ T_\delta\left(\lambda_\delta^o\right)}$ and the third is the sample

standard deviation for this ratio from the 200 trials. Over the range of cases the ACP remained close to the nomial level, .95, while $\hat{T}_\delta$ appears to have little bias. The smoothing spline and the diagonal elements of $A(\lambda)$ were calculated using the order N algorithm in Hutchinson and de Hoog (1985). The minimizations of $V(\lambda)$ and (4.1) were carried out with respect to log (λ) using a coarse grid search followed by a golden section search. Normal errors were simulated by Knuth's version of the Kinderman-Monahan ratio of uniforms.

Table 3 **Average coverage probabilities for a natural spline estimate restricted to an interior interval.**

True function	n	$\sigma = .0125$ %ACP	$\sigma = .0125$ $\frac{\text{mean}(\hat{T})}{ET(\lambda^o)}$	$\sigma = .0125$ $\frac{SD(\hat{T})}{ET(\lambda^o)}$	$\sigma = .05$ %ACP	$\sigma = .05$ $\frac{\text{mean}(\hat{T})}{ET(\lambda^o)}$	$\sigma = .05$ $\frac{SD(\hat{T})}{ET(\lambda^o)}$	$\sigma = .2$ %ACP	$\sigma = .2$ $\frac{\text{mean}(\hat{T})}{ET(\lambda^o)}$	$\sigma = .2$ $\frac{SD(\hat{T})}{ET(\lambda^o)}$
	64	94.6	1.01	.16	94.1	1.02	.21	92.4	1.03	.38
Case I	128	94.6	1.01	.14	93.8	1.02	.19	93.5	1.03	.39
	256	94.3	1.02	.11	94.4	1.04	.26	93.7	1.02	.26
	64	94.3	1.04	.20	94.1	1.03	.20	94.2	1.01	.19
Case II	128	94.7	1.00	.14	94.3	1.00	.13	94.2	1.00	.17
	256	94.3	1.00	.09	94.8	1.00	.09	94.2	1.00	.10
	64	94.2	1.01	.19	93.9	1.03	.20	93.6	1.00	.18
Case III	128	94.7	1.01	.14	94.2	1.00	.13	94.2	1.00	.20
	256	94.6	1.01	.09	94.7	1.00	.10	94.3	1.00	.11
	64	94.1	1.01	.19	93.8	1.02	.19	93.3	.98	.19
Case III	128	94.6	1.00	.14	93.9	1.00	.13	92.9	.93	.21
(shifted)	256	94.1	1.00	.09	94.5	.98	1.00	92.9	.93	.14

Note: The standard errors for these estimated %ACPs range from .22 to .57 and the sample statistics for $\hat{T}$ were computed from 200 trials.

SECTION 5

In this section we will establish (1.1) showing that the ACP when λ^o and $E\,T_n(\lambda^o)$ are estimated is asymptotically equivalent to the ACP when both of these quantities are known and the measurement errors are normally distributed.

A proof is given for periodic splines and at the end of this section we discuss the modifications necessary to show this equivalence for a natural spline evaluated on a restricted interval. If we limited the scope to periodic splines with equally spaced knots this proof could be greatly simplified by expanding $\hat{f}$ and f in Fourier series. We do not take this approach because it will not generalize easily to the non-periodic case or to higher dimensional thin plate splines.

Let G_n denote the empirical distribution function for $\{t_k\}_{k=1,n}$.

Although other cases can be analyzed, we will assume that the knots are chosen by one of two methods:

A: designed knots:

$$\sup_{u \in [0,1]} \left| G_n - u \right| = \mathcal{O}\left(\frac{1}{n}\right)$$

B: random knots: $\{t_k\}_{k=1,n}$

is a random sample from the uniform distribution on [0,1].

We will make the following assumptions:

F1: $E\left(\left|e_k\right|^9\right) < \infty$

F2: $\hat{\lambda}$ is the minimizer of $V(\lambda)$ over the interval $[\lambda_n, \infty]$ where $\lambda_n \sim n^{-\frac{a}{s}}$

F3: f is such that there is a $\gamma > 0$ and

$$\frac{1}{n}\sum_{k=1} b_k^2 = \gamma\lambda^2\left[1 + o(1)\right]$$

uniformly for $\lambda \in [\lambda_n, \infty)$.

A more detailed set of assumptions that separate the two different types of knot selection is given in Nychka (1986).

F1 is necessary to guarantee asymptotic normality of the spline estimate. Unfortunately, uniform asymptotic approximations to f_λ are only possible for $\lambda \in [\lambda_n, \infty)$ and F2 insures that $\hat{\lambda}$ will be in this range. The last assumption concerns the asymptotic behavior of the mean squared bias for f.

F3 will hold for periodic splines provided that

$f^{iv} \in L^2[0,1]$ and $f^{(k)}(0) = f^{(k)}(1) = 0 \quad k = 0,3$.

Theorem 5.1

(5.1) Let $C(t,\alpha) = \hat{f}(t) \pm Z_{\frac{\alpha}{2}}\sqrt{\hat{T}}$ where $\hat{T}$ is given by (2.4) .

Under F1-F3

(5.2) $P\left[f(\tau_n) \in C(\tau_n,\alpha)\right] - P\left[|\mathcal{U}| \le Z_{\frac{\alpha}{2}}\right] \to 0$

uniformly in α as $n \to \infty$.

Corollary 5.1

If $\hat{T}$ is replaced by $\frac{\hat{\sigma}^2}{n}$ tr $A(\hat{\lambda})$ in (5.1) then (5.2) holds provided $\mathcal{U}$ is replaced by $\left(\frac{27}{32}\right)\mathcal{U}$ (c.f. (2.3)).

An outline of the proof is given while the technical details are developed by the lemmas included in Appendix A. To simplify notation, for $\underset{\sim}{a} \varepsilon \mathbb{R}^n$, $c > 0$ let

$$\varphi\left(\underset{\sim}{a} , c\right) = \frac{1}{n} \sum_{k=1}^{n} I\left(a_k\right)_{\left\{\left|a_k\right| < c\right\}}$$

$$\hat{\underset{\sim}{w}} = \frac{\hat{\underset{\sim}{f}} - \underset{\sim}{f}}{\sqrt{\hat{T}_n}} \quad \text{and } \underset{\sim}{w}^o = \frac{\underset{\sim}{f}^o - \underset{\sim}{f}}{\sqrt{E\ T_n(\lambda^o)}} \quad \text{, and set } Z_{\frac{\alpha}{2}} = z$$

For $\underset{\sim}{a}$, $\underset{\sim}{b} \varepsilon \mathbb{R}^n$ and $\varepsilon > 0$, φ satisfies the elementary inequality:

$$(5.3) \quad \left|\varphi\ (\underset{\sim}{a} , c) - \varphi\ (\underset{\sim}{b} , c)\ \right| \leq \frac{1}{n} \frac{\|\ \underset{\sim}{a} - \underset{\sim}{b}\ \|^2}{\varepsilon^2} + \left|\varphi\left(\underset{\sim}{b} , c\right) - \varphi\left(\underset{\sim}{b} , c+\varepsilon\right)\right|$$

(see Lemma A.3) and is the main device in the proof for bounding expressions. Also, we will use a Gaussian approximation to a smoothing spline developed in Cox(1986). Under assumptions F1-F3 it is possible to construct a probability space containing the processes, $\hat{w}(t)$, $w^o(t)$ and $\bar{w}^o(t)$ such that Law $(\hat{\underset{\sim}{w}})$ = Law $(\hat{\underset{\sim}{w}})$, Law $(\underset{\sim}{w}^o)$ = Law $(\underset{\sim}{w}^o)$, $\frac{1}{n} \left\|\underset{\sim}{w}^o - \bar{\underset{\sim}{w}}^o\right\|^2 \overset{P}{\to} 0$

and $E\varphi\left(\bar{\underset{\sim}{w}}^o, z\right) = P\left(\left|\mathfrak{U}\right| < z\right)$. This last property follows from the fact $\bar{w}^o(t)$ is the approximation to a spline estimate based on normal errors. (See Appendix A for details.)

Proof of Theorem 1.

Reexpressing (5.2) for the probability space that contains the Gaussian approximation to $\hat{f}_\lambda$.

$$\left|E\varphi\ \left(\hat{\underset{\sim}{w}} , z\right) - P\ \left(\left|\mathfrak{U}\right| < z\right)\right| = \left|E\ \varphi\left(\hat{\underset{\sim}{w}} , z\right) - E\ \varphi\left(\bar{\underset{\sim}{w}}^o , z\right)\right|$$

$$\leq E\left|\ \varphi\left(\hat{\underset{\sim}{w}} , z\right) - \varphi\left(\bar{\underset{\sim}{w}}^o, z\right)\right| \quad .$$

Adding and subtracting $\varphi\left(\underset{\sim}{w}^o, z\right)$ and using (5.3), for any $\varepsilon > 0$

$$\left|\varphi\left(\hat{\underset{\sim}{w}} , z\right) - \varphi\ \left(\bar{\underset{\sim}{w}}^o, z\right)\right|$$

$$\le \frac{1}{n\varepsilon^2} \left\| \hat{w} - w^o \right\|^2 + \left| \varphi\left(\bar{w}^o, z + \varepsilon\right) - \varphi\left(w^o, z\right) \right|$$

Adding and subtracting $\varphi\left(\bar{w}^o, z\right)$ within the second expression and using (5.3) twice more

$$\le \frac{1}{\varepsilon^2} \left\{ \frac{1}{n} \left\| \hat{w} - w^o \right\|^2 + \frac{3}{n} \left\| w^o - \bar{w}^o \right\|^2 \right\}$$
$$+ \left\{ 2 \left| \varphi\left(\bar{w}^o, z + \varepsilon\right) - \varphi\left(\bar{w}^o, z\right) \right| \right\}$$
$$\le \frac{1}{\varepsilon^2} \left\{ \alpha_n \right\} + \left\{ \beta_n \right\}$$

$\alpha_n \overset{P}{\to} 0$ as $n \to \infty$ by Lemma A.1 and the construction of $\bar{w}^o$. By Lemma A.2, $E(\beta_n) = \mathcal{O}(\varepsilon)$ uniformly for $z \varepsilon \mathbb{R}$. With these results we have for any $\delta > 0$

$$P\left(\left| \varphi\left(\hat{w}, z\right) - \varphi\left(\bar{w}^o, z\right) \right| > \delta \right) \le P\left(\alpha_n \ge \varepsilon^2 \delta\right) + \mathcal{O}\left(\frac{\varepsilon}{\delta}\right)$$

and it must follow that $\left| \varphi\left(\bar{w}, z\right) - \varphi\left(w^o, z\right) \right| \overset{P}{\to} 0$ as $n \to \infty$.

Moreover, because this random variable is bounded

$$E \left| \varphi\left(\hat{w}, z\right) - \varphi\left(w^o, z\right) \right| \to 0 \text{ as } n \to \infty .$$

The corollary follows from the remarks at the end of Sections and by a straight forward modification of the definitions of $\hat{w}$ and w^o in Lemma A.1.

The last part of this section describes how Theorem 5.1 can be extended to a natural spline estimator. In the statement of the Theorem $\mathfrak{U}$ and $\hat{T}$ need to be replaced by $\mathfrak{U}_\delta$ and $\hat{T}_\delta$ respectively. Also, τ_n should be restricted to $\mathfrak{I}$. The corollary will also require similar modifications. To simplify this discussion, we will assume that $\mathfrak{I}$ is a fixed interval rather than the adaptive interval $\mathfrak{I}(\hat{\lambda})$ described in Section 4. For assumption F2 the interval $[\lambda_{n,\infty}]$ should be replaced by $[\lambda_n, \zeta_n]$ where $\zeta_n \to 0$ as $n \to \infty$. F3 needs to be changed to read

$$\frac{1}{n_\delta} \sum_{t_k \varepsilon \mathcal{I}} b_k^2 = \gamma \lambda^2 (1 + o(1))$$

uniformly for $\lambda \ \varepsilon \ [\lambda_n, \zeta_n]$ as $n \to \infty$.

This later condition will hold provided $f^{i\nu}$ is continuous and not identically zero.

The last requirement for this modification of Theorem 5.2 is the consistency of $\hat{\lambda}_\delta$ and $\hat{T}_\delta$. The convergence of both of these estimates hinges on the uniform convergence of T_δ and V_δ to their expected values:

(5.4) $$T_\delta(\lambda) - E\ T_\delta(\lambda) = \mathcal{O}_p\left[E\ T_\delta(\lambda)\right]$$

(5.5) $$\left[V_\delta(\lambda) - S_n^2\right] - \left[E\ V_\delta(\lambda) - \sigma^2\right] = \mathcal{O}p\left[E\ T_\delta(\lambda)\right]\ ,\ S_n^2 = \frac{1}{n} \sum_{k=1}^{n} e_k^2\ .$$

uniformly for $\lambda \ \varepsilon \ [\lambda_n, \infty)$ as $n \to \infty$.

Given F1–F3, (5.4) and (5.5) it will follow that

$\hat{\lambda}_\delta \Big/ \lambda_\delta^o \overset{P}{\to} 1$ and $\hat{T}_\delta \Big/ E\ T_\delta(\lambda_\delta^o) \overset{P}{\to} 1$ as $n \to \infty$ and this restatement of Theorem 5.2 will also hold. Unfortunately at present, there is no theory to establish (5.4) and (5.5) when $\mathcal{I}$ is a sub interval of [0,1]. Using the kernel approximation to a smoothing spline and the asymptotic theory for kernel estimators (Härdle and Marron (1985)), however, we believe it is possible to prove that these relationships are true.

REFERENCES

Bates, P., Lindstrom, M., Wahba, G. and Yandell, B. (1986). GCVPACK - Routines for generalized cross validation. Technical Report No. 775, Department of Statistics, University of Wisconsin/Madison.

Cox, D. D. (1983). Approximation of method of regularization estimators. Technical Report 723, University of Wisconsin/Madison, Department of Statistics.

Cox, D. D. (1984). Gaussian approximation of smoothing splines. Technical Report 743, University of Wisconsin/Madison, Department of Statistics.

Cox, D. D. (1985). Approximation theory of method of regularization estimators: applications. unpublished manuscript.

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

Hutchinson, M. F. and de Hoog, F. R. (1985). Smoothing noisy data with spline functions. Numer. Math. 47, 107-122.

Marron, J. S. and Hardle, W. (1984). Random approximations to some measures of accuracy in nonparametric curve estimation. N. C. Inst. of Statist. Mimeo Series No. 1566.

Messer, K. (1986). A comparison of a spline estimate to its "Equivalent" kernel estimate. unpublished manuscript.

Nychka, D. W. (1986). The average posterior variance of a smoothing spline and a consistent estimate of the mean squared error. North Carolina Institute of Statistics Mimeo Series No. 1688 .

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. J. R. Statist. Soc. B. 47 1-52.

Speckman, P. (1983). Efficient nonparametric regression with cross validated smoothing splines. unpublished manuscript.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. J. R. Statist. Soc. B. 45, 133-150.

Wahba, G. (1985). Discussion of "Some aspects of the spline smoothing approach to nonparametric regression curve fitting", J. R. Statist. Soc. B. 47, 44.

Douglas Nychka
Department of Statistics
North Carolina State University - Campus Box 8203
Raleigh, North Carolina 27695-8203

APPENDIX A

The following 3 Lemmas form the basis for the proof of Theorem 1. For each of these results we will assume that the conditions F1-F3 are in force. Before presenting the lemmas, we first detail a slight difference between Cox's presentation of the Gaussian approximation to f_λ and that described in the previous section.

Cox (1984) (subsequently abbreviated GA) expresses the convergence of ω^o to $\bar{\omega}$ in terms of a continuous norm rather than the discrete one used to define the mean squared error. We note that assumption 4.1 (f) of Cox (1983) (subsequently abbreviated AMORE) gives the basic relationship between these two norms, and using arguments similar to those in the proof of Lemma 4.2 of AMORE and the results in Lemma 4.2 of GA it is possible to show that the results are also valid when discrete norms are used.

Lemma A.1

$$\text{Let} \quad \hat{\underset{\sim}{w}} = \frac{\hat{\underset{\sim}{f}} - \underset{\sim}{f}}{\hat{T}} \qquad \underset{\sim}{w}^o = \frac{\underset{\sim}{f}^o - \underset{\sim}{f}}{E\,T\,(\lambda^o)}$$

and suppose that $\hat{T}\big/_{ET(\lambda^o)} \xrightarrow{P} 1$ as $n \to \infty$

then $\frac{1}{n} \left\| \hat{\underset{\sim}{w}} - \underset{\sim}{w}^o \right\|^2 = \mathcal{O}_p(1)$ as $n \to \infty$

Proof

First, we will argue that:

$$\text{(A.1)} \qquad \frac{1}{n} \left\| \hat{\underset{\sim}{f}} - \underset{\sim}{f}^o \right\|^2 = \mathcal{O}_p\left(E\,T\left(\lambda^o\right)\right) \text{ as } n \to \infty .$$

We will use the fact that the eigenvalues of $A(\lambda)$ have that form:

$$\frac{1}{1 + \lambda\gamma_{kn}} \qquad \text{where } 0 \le \gamma_{1n} \le \gamma_{2n} \cdots \le \gamma_{nn} \quad \text{and the}$$

orthogonal matrix, U, that diagonalizes $A(\lambda)$ does not depend on λ.

(A.2) $$\frac{1}{n}\left\|\hat{f} - f^o\right\|^2 = \frac{1}{n}\left\|\left(A(\hat{\lambda}) - A(\lambda^o)\right)Y\right\|^2$$

$$= \frac{1}{n}\left\|\left(A(\hat{\lambda}) - A(\lambda^o)\right)f\right\|^2 + \frac{1}{n}\left\|\left(A(\hat{\lambda}) - A(\lambda^o)\right)e\right\|^2 - \frac{2}{n} e'\left(A(\hat{\lambda}) - A(\lambda^o)\right)^2 f$$

Now

$$\frac{1}{n}\left\|\left(A(\hat{\lambda}) - A(\lambda^o)\right)f\right\|^2 = \frac{1}{n}\left(\frac{\hat{\lambda} - \lambda^o}{\lambda^o}\right)^2 \sum_{k=1}^{n} \left(\frac{\lambda^o\gamma_{kn}}{1+\lambda^o\gamma_{kn}}\right)^2 \left(\frac{1}{1+\hat{\lambda}\gamma_{kn}}\right)^2 \beta_{kn}^2$$

where $\beta_n = U f$

$$\leq \frac{1}{n}\left(\frac{\hat{\lambda} - \lambda^o}{\lambda^o}\right)^2 \sum_{k=1}^{n} \left(\frac{\lambda^o \gamma_{kn}}{1 + \lambda^o\gamma_{kn}}\right)^2 \beta_{kn}^2$$

$$= \left(\frac{\hat{\lambda} - \lambda^o}{\lambda^o}\right)^2 \frac{1}{n}\left\|\left(I - A(\lambda^o)\right)f\right\|^2 \leq \left(\frac{\hat{\lambda} - \lambda^o}{\lambda^o}\right)^2 E\,T_n(\lambda^o)$$

and by similar arguments we have

$$\frac{1}{n}\left\|\left(A(\hat{\lambda}) - A(\lambda^o)\right)e\right\|^2 \leq \left(\frac{\hat{\lambda} - \lambda^o}{\hat{\lambda}}\right)^2 \frac{1}{n}\left\|A(\lambda^o)\, e\right\|^2$$

because $\left(\frac{\hat{\lambda}\gamma_{kn}}{1+\hat{\lambda}\gamma_{kn}}\right) \leq 1$

From Lemma 3.2 Nychka (1986), $\left(\frac{\hat{\lambda} - \lambda^o}{\lambda^o}\right) \xrightarrow{P} 0$ and combining Lemma 3.1 and Lemma 3.3 from the same paper $\frac{1}{n}\left\|A(\lambda^o)\, e\right\|^2 = O_p(E\,T_n)$.

Thus, the first two terms of (A.2) converge at the appropriate rate.

Setting $\xi = Ue$ the last term in (A.2) can be rewritten as

$$\frac{1}{n}\,\underset{\sim}{e}'\left[A(\hat{\lambda}) - A(\lambda^o)\right]^2 \underset{\sim}{f}$$

$$= \left(\frac{\lambda^o - \hat{\lambda}}{\lambda^o}\right)^2 \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\gamma_{kn}\lambda^o}{\gamma_{kn}\lambda^o + 1}\right)^2 \left(\frac{1}{1+\gamma_{kn}\hat{\lambda}}\right)^2 \xi_{kn}\,\beta_{kn}$$

$$= \left(\frac{\lambda^o - \hat{\lambda}}{\lambda^o}\right)^2 \sum_{k=1}^{n} c_{kn}X_{kn}\ ,$$

where $c_{kn} = \dfrac{1}{\left(1+\gamma_{kn}\hat{\lambda}\right)^2}$ and $X_{kn} = \dfrac{1}{n}\left(\dfrac{\gamma_{kn}\lambda^o}{1+\gamma_{kn}\lambda^o}\right)^2 \xi_{kn}\beta_{kn}$

First, we will assume that $\underset{\sim}{e} \sim N(0,\sigma^2 I)$. In this case, X_{kn} are independent, mean zero random variables because $\underset{\sim}{\xi}$ will also be distributed $N(0,\sigma^2 I)$.

Let $\mathcal{C} = \left\{\underset{\sim}{u} \varepsilon \mathbb{R}^n\colon 1 \geq u_1 > u_2 > \ldots > u_n > 0\right\}$ and applying Lemma 4.3 from Speckman (1982) it follows that

$$P\left(\sup_{\underset{\sim}{v}\,\varepsilon\,\mathcal{C}} \left|\Sigma\, v_k\, X_{kn}\right| > \varepsilon\right) \leq \frac{\mathrm{VAR}\left(\Sigma\, X_{kn}\right)}{\varepsilon^2}$$

Note that $\underset{\sim}{c}_n \varepsilon\, \mathcal{C}$ and thus

$$P\left(\left|\sum_{k=1}^{n} c_{kn}X_{kn}\right| > \varepsilon\right) \leq \frac{1}{n^2}\sum_{k=1}^{n}\frac{\left(\frac{\gamma_{kn}\lambda^o}{1+\gamma_{kn}\lambda^o}\right)^4 \beta_{kn}^2}{\varepsilon^2}$$

$$\leq \frac{1}{n^2}\,\frac{\left\|\left(I-A(\lambda^o)\right)\underset{\sim}{f}\right\|^2}{\varepsilon^2}$$

or,

$$P\left(\left|\sum_{k=1}^{n}\frac{C_{kn}X_{kn}}{E\ T_n(\lambda^o)}\right| > \varepsilon\right) \leq \frac{1}{n\ E\ T_n(\lambda^o)}$$

It is well known that $\dfrac{1}{n\ E\ T_n(\lambda^o)} \longrightarrow 0$ as $n \longrightarrow \infty$ and in view of (A.2), (A.1) must also go to zero at the necessary rate.

Now consider the case when the errors are not normal but satisfy F1 and F2. Following the development on pg. 53 of GA there is a probability space with two random vectors $\tilde{e}$ and $\bar{e}$ such that $\text{Law}(\tilde{e}) = \text{Law}(e)$, $\bar{e} \sim N(0,\sigma^2 I)$ and $\bar{e}$ is a suitable approximation to $\tilde{e}$ in the following sense:

(A.3) If $H_2(n,\lambda) = \tilde{e}'\left[I-A(\lambda)\right]^2 f$ and $\bar{H}_2(n,\lambda) = \bar{e}'\left[I-A(\lambda)\right]^2 f$

(A.4) then $\left|H_2(n,\lambda) - \bar{H}_2(n,\lambda)\right| = o_p\left[E\, T_n(\lambda)\right]$

uniformly for $\lambda \in \left[\lambda_n, \infty\right]$.

From the discussion above the claim (A.1) will follow for nonnormal errors, provided we can replace f in (A.3) by $A(\hat{\lambda})^2 f$. This substitution can be shown to be valid by examining Cox's proof of (A.4) on pages 59-60 of GA. The upper bounds obtained by Cox actually hold uniformly for all $f \in B_M$ where $B_M = \left\{h \in W_2^2 : \|h\|_{W_2^2} \le M\right\}$. (see also Theorem 4.3 of AMORE). Now let $M = 2\|f\|_{W_2^2}$ where f is the true function in the model. Because $f \in W_2^4$ by assumption F3, we have $\|\bar{f}\|_{W_2^2} - \|f\|_{W_2^2} \to 0$ as $n \to \infty$. Where $\bar{f}$ is the spline interpolant of $A(\hat{\lambda})^2 f$. Thus, for n sufficiently large, $\bar{f} \in B_M$ and by uniformity of the convergence in (A.4)

$$\frac{1}{n} e' \left[I-A(\hat{\lambda})\right]^2 A(\hat{\lambda})^2 f = \frac{1}{n} e' \left[I-A(\hat{\lambda})\right]^2 \bar{f} = o_p\left[E\, T_n(\lambda^o)\right] \text{ as } n \to \infty .$$

For convenience let $m_o = \sqrt{E\ T_n(\lambda^o)}$ and $\hat{m} = \sqrt{\hat{T}_n}$.

$$\| \hat{\underset{\sim}{w}} - \underset{\sim}{w}_o \| = \| \frac{\hat{\underset{\sim}{f}} - \underset{\sim}{f}}{\hat{m}} - \frac{\underset{\sim}{f}_o - \underset{\sim}{f}}{m_o} \|$$

$$\leq \left(\frac{m_o}{\hat{m}}\right) \| \frac{\hat{\underset{\sim}{f}} - f_o}{m_o} \| + \| \frac{\underset{\sim}{f}_o - \underset{\sim}{f}}{m_o} \| \left| 1 - \frac{m_o}{\hat{m}} \right|$$

and the Lemma now follows from the consistency of $\hat{m}$.

Lemma A.2

for $\underset{\sim}{a}$, $\underset{\sim}{b} \varepsilon \mathbb{R}^n$, $c > 0$, $\varepsilon > 0$

$$\left|\varphi(\underset{\sim}{a}\ ,\ c) - \varphi\ (\underset{\sim}{b}\ ,\ c)\right| \leq \frac{1}{n} \| \frac{\underset{\sim}{a} - \underset{\sim}{b}}{\varepsilon^2} \|^2 + \left| \varphi\ (\underset{\sim}{b}\ ,\ c) - \varphi\ (\underset{\sim}{b}\ ,\ c+\varepsilon) \right|$$

Proof

$$\left|\varphi(\underset{\sim}{a}\ ,\ c) - \varphi(\underset{\sim}{b}\ ,\ c)\right| \leq \left|\varphi(\underset{\sim}{a}\ ,\ c) - \varphi(\underset{\sim}{b}\ ,\ c+\varepsilon)\right| + \left|\varphi(\underset{\sim}{b}\ ,\ c) - \varphi(\underset{\sim}{b}\ ,\ c+\varepsilon)\right|$$

Now

$$\left|\varphi(\underset{\sim}{a}\ ,\ c) - \varphi(\underset{\sim}{b}\ ,\ c-\varepsilon)\right| \leq \frac{1}{n} \sum_{k=1}^{n} \left| \underset{\{|a_k| < c\}}{I} (a_k) - \underset{\{|b_k| < c+\varepsilon\}}{I} (b_k) \right|$$

$$\leq \frac{1}{n} \sum_{k=1}^{n} \underset{\{|a_k - b_k| > \varepsilon\}}{I} (a_k - b_k)$$

and thus

$$\left|\varphi(\underset{\sim}{a}\ ,\ c) - \varphi\ (\underset{\sim}{b}\ ,\ c+\varepsilon)\right| \leq \frac{1}{n} \Sigma \left(\frac{a_k - b_k}{\varepsilon}\right)^2$$

by Chebyshev's inequality.

Lemma A.3

For $\varepsilon > 0$ there is a $M < \infty$ independent of z such that:

$$\text{(A.5)} \qquad E\left[\varphi\left(\bar{\underset{\sim}{w}}^{o}, z + \varepsilon\right) - \varphi\left(\bar{\underset{\sim}{w}}^{o}, z\right)\right] \leq \varepsilon M .$$

Proof

(A.5) can be reexpressed as

$$\frac{1}{n}\sum_{k=1}^{n} P\left[z < \bar{\omega}_k^{o} < z + \varepsilon\right]$$

and $\bar{\omega}_k \sim N\left(\mu_k, \rho_k^2\right)$, $\mu_k = \dfrac{b(t_k)}{\sqrt{E\, T_n(\lambda^o)}}$

$$\rho_k^2 = \frac{A_{kk}^2(\lambda^o)}{E\, T_n(\lambda^o)}$$

Setting $\alpha_k = \left(c-\mu_k\right)/\rho_k$ and $\beta_k = \left(c+\mu_k\right)/\rho_k$ and taking Φ to be the standard normal c.d.f.

$$P\left[c < \left|\bar{\omega}_k\right| < c + \varepsilon\right]$$

$$= \Phi\left(\alpha_k + \varepsilon/\rho_k\right) - \Phi\left(\alpha_k\right) + \Phi\left(\beta_k + \varepsilon/\rho_k\right) - \Phi\left(\beta_k\right) \leq \frac{1}{\sqrt{2\pi}}\, \varepsilon/\left[\min\left\{\rho_k\right\}\right]^{-1}$$

by the mean value theorem and because ρ_k is bounded away from 0 the lemma follows.

Appendix B

In this appendix we derive a bias correction to the cross validation function to improve the estimates of the mean squared error.

Let $\mu_j = \frac{1}{n_\delta}$ tr $(WA^j(\lambda))$ for $j = 1,2$

and $b^2 = \frac{1}{n_\varepsilon} \sum_{t_k \varepsilon \mathcal{D}(\lambda)} \left[E f_\lambda(t_k) - f(t_k)\right]^2$

$$E\ V_\delta(\lambda) - \left[E\ T_\delta(\lambda) + \sigma^2\right] = \frac{b^2 + \sigma^2 \left[1 - 2m_1 + m_2\right] - \left[b^2 + \sigma^2\left[1 + m_2\right]\right]}{\left[1 - m_1\right]^2}$$

and after some algebra

$$\mathcal{E}(\lambda) = \frac{E\ V_\delta(\lambda) - \left[E\ T_\delta(\lambda) + \sigma^2\right]}{E\ T_\delta(\lambda)} = \frac{m_1\left[2 - \beta_1\beta_2 - m_1\right]}{\left[1 - m_1\right]^2}$$

where $\beta_1 = \frac{m_1}{m_2}$ and $\beta_2 = \frac{b^2}{b^2 + \sigma^2 m_2}$.

Thus, we take $\varphi(\lambda) = \frac{1}{1 + \mathcal{E}(\lambda)}$ as the bias correction.

Note that all the quantities can be computed directly from the diagonal elements of $A(\lambda)$ and $A^2(\lambda)$ except for β_2 . Under the assumption that $\hat{\lambda}_\delta$ is a consistent estimate of λ^o_δ and that b^2 satisfies condition F3, it is straight forward to show: $\beta_2 = \frac{8}{9}$. In general, if f_λ is an m^{th} order smoothing spline $\beta_2 = \frac{4m}{4m+1}$ (see Section 3).