Bayesian Confidence Intervals for Smoothing Splines

DOUGLAS NYCHKA*

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 are averaged across the observation points, the average coverage probability (ACP) should be close to the nominal level. From a frequency point of view, this agreement occurs because the average posterior variance for the spline is similar to a consistent estimate of the average squared error and because the average squared bias is a modest fraction of the total average squared error. These properties are independent of the Bayesian assumptions used to derive this confidence procedure, and they explain why the ACP is accurate for functions that are much smoother than the sample paths prescribed by the prior. This analysis accounts for the choice of the smoothing parameter (bandwidth) using cross-validation. In the case of natural splines an adaptive method for avoiding boundary effects is considered. The main disadvantage of 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.

KEY WORDS: Boundary effects; Cross-validation; Nonparametric regression; Smoothing parameter.

1. INTRODUCTION

Consider the additive model $Y_k = f(t_k) + e_k$ $(1 \le k \le n)$, where the observation vector $\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_n)$ depends on a smooth, unknown function f evaluated at the points $0 \le t_1 \le t_2 \le \cdots \le t_n \le 1$ and a vector of independent and identically distributed errors: $\mathbf{e}' = (e_1, e_2, \ldots, e_n)$ with $E(\mathbf{e}) = 0$ and $E(\mathbf{ee}') = \sigma^2 I$. The statistical problem posed by this model is to estimate f from $\mathbf{Y}$ without having 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 [see Silverman (1985) for a review]. One limitation in applying spline methods in practice, however, is the difficulty in constructing confidence intervals or specifying other measures of the estimate's accuracy. Wahba (1983), using the interpretation of a smoothing spline as a posterior mean, 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)

$$\frac{1}{n} \sum_{k=1}^{n} \Pr\{f(t_k) \in C(\alpha, t_k)\}$$

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 Gaussian process prescribed by the prior, she used only several different fixed functions.

Figures 1 and 2 contrast the difference between these two choices for f. Figure 1 is a plot of three realizations of the stochastic process corresponding to Wahba's prior distribution for f. Figure 2 shows three of the cases used in her simulations. These two groups of curves clearly exhibit 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.

My goal in this article 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 consider a confidence interval for $f(\tau)$, where $f(\cdot)$ is the realization of a stochastic process and τ is fixed, I 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 $\Pr[f(\tau_n) \in C(\alpha, \tau_n)]$. Moreover,

$$\Pr[f(\tau_n) \in C(\alpha, \tau_n)] = \Pr(|\mathcal{U}| \le Z_{\alpha/2}) + o(1) \quad \text{as } n \to \infty, \qquad (1.1)$$

where $\mathcal{U}$ is close to a standardized random variable that is the sum of a normal random variable and a discrete random variable related to the bias of the spline estimate at the observation points. Z_α is the $(1 - \alpha)100$ normal percentile. 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. 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

* Douglas Nychka is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203. The author used equipment provided by National Science Foundation Grant DMS-8404511.

Journal of the American Statistical Association
December 1988, Vol. 83, No. 404, Theory and Methods

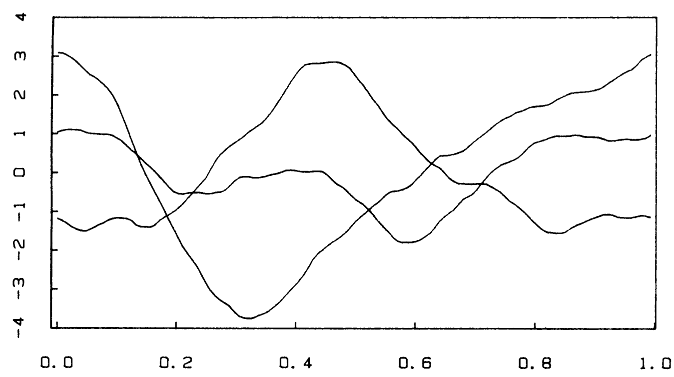

Figure 1. Three Realizations From the Bayesian Prior. For a periodic second-order smoothing spline the prior considered by Wahba is an integrated Brownian bridge adjusted to be a periodic function. To facilitate graphing the sample paths, they have been scaled to come from a Brownian bridge with unit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

are two factors that contribute to the accuracy of the average coverage probability. The average posterior variance of the cross-validated spline estimate is proportional to a consistent estimate of the expected average squared error, and the average squared bias is a small fraction of the total average squared error. Suppose that $\hat{f}$ denotes a spline estimate where the smoothing parameter has been determined by cross-validation, and let $\hat{T}$ be a consistent estimate of the expected average squared error. These 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 $\hat{f}(t_k) \pm Z_{\alpha/2}\sqrt{\hat{T}}$ will have an ACP close to $1 - \alpha$. Moreover, because this analysis does not specifically depend on $\hat{f}$ being a smoothing spline, it suggests that these results may hold in general for other nonparametric regression estimators.

This frequency interpretation also 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 across the observation points (see Wahba 1985). In particular, points where the biases are large will result in a coverage below $1 - \alpha$. Figures 3 and 4 illustrate this relationship for one of the cases from Wahba's Monte Carlo study. This problem is compounded because the bias of $\hat{f}$ tends to be large precisely at the points where f has a more complicated and interesting structure. Note that the points where the bias is large correspond to sharp peaks or kinks in the function (see Fig. 2). Thus one would not expect this confidence procedure to be reliable if these intervals were only computed at points where $\hat{f}$ has a peak. Despite this difficulty, this type of interval appears to provide a reasonable measure of the spline estimate's accuracy provided that the point for evaluation is chosen independently of the shape of f.

The large biases in Figure 4 are a consequence of the spline estimate not adapting to the local behavior of f. The estimate in this example uses a single, global value of the smoothing parameter (bandwidth). This value is appropriate at most points, but places where f changes rapidly require a smaller value of the smoothing parameter. One possible way of reducing the bias is to consider a spline estimate where the smoothing parameter can vary. Härdle and Bowman (1988) and Staniswallis (1986) proposed adjusting the bandwidth for a kernel estimate based on estimates of the bias, and a similar approach can be taken for splines. Another strategy is to estimate the smoothing parameter at a particular point, t_k, by cross-validating only on observations in a neighborhood of t_k. The size of this neighborhood could be inferred from the value of the global smoothing parameter and a suitable prior for f.

Another approach to confidence intervals (Cox 1986) is to estimate a worst-case bias for f to obtain a conservative interval. This method has the advantage of being reliable at all points; however, the width of these intervals may be unacceptably large.

Section 2 gives the details of the stochastic interpretation of a smoothing spline and discusses the connection between the average posterior variance and the expected

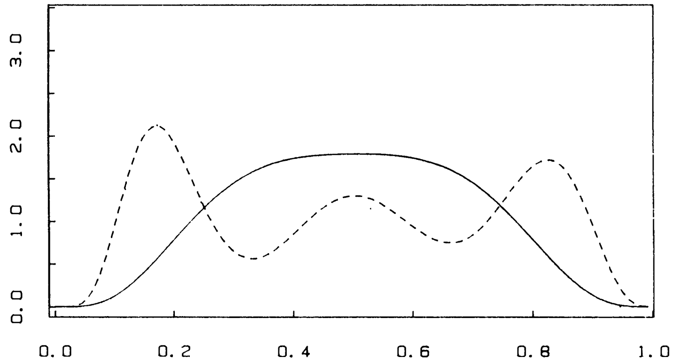

Figure 2. Three Smooth Test Functions Considered by Wahba (1983). These three cases all have at least one continuous derivative satisfying periodic boundary conditions and are mixtures of beta densities: Case 1, $\frac{1}{3}\beta_{10,5} + \frac{1}{3}\beta_{7,7} + \frac{1}{3}\beta_{5,10}$ (——); Case 2, $\frac{6}{10}\beta_{30,17} + \frac{4}{10}\beta_{3,11}$ (· · ·); Case 3, $\frac{1}{3}\beta_{20,5} + \frac{1}{3}\beta_{12,12} + \frac{1}{3}\beta_{7,30}$ (- - -), where $\beta_{m,n}$ is the standard beta density function on [0, 1].

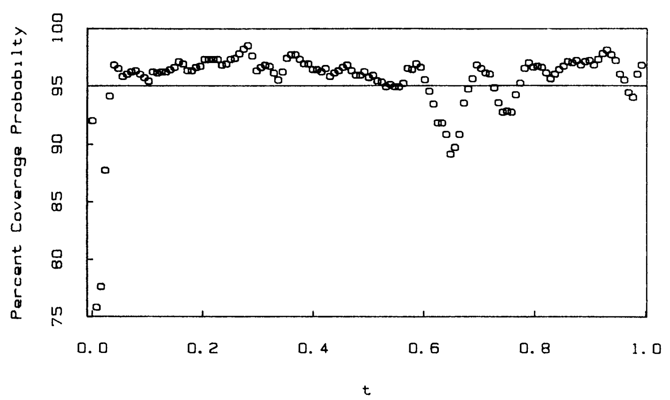

Figure 3. An Example of the Individual Coverage Probabilities of Wahba's "Bayesian" Confidence Intervals. Plotted are the pointwise coverage probabilities estimated from a simulation of 1,000 trials. The true function is Case 2 with n = 128, σ = .05, and using a normal distribution for the errors. The smoothing for each estimate was determined by generalized cross-validation. The estimated ACP for this case is .952 (horizontal line) with a standard error of .001.

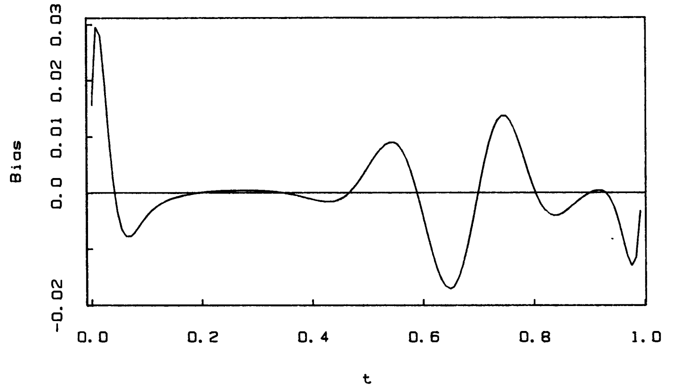

Figure 4. Bias of a Spline Estimate. The same parameters are used as in Figure 3, except λ is taken to be the minimizer of the expected average squared error (λ^o = 6.99E–5). Points where the estimate is significantly biased correspond to lower coverage probabilities and match sharp peaks or kinks in the true function (see Fig. 2).

average squared error. Section 3 defines the random variable, $\mathcal{U}$, that is related to the ACP. The distribution of $\mathcal{U}$ was 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 Section 3 generalize to natural smoothing splines, and gives some simulation results. As part of this discussion, a simple adaptive method is suggested for avoiding large biases in the estimate at the endpoints. Section 5 outlines a proof of the relationship in (1.1). This analysis accounts for the fact that the smoothing parameter is determined adaptively by cross-validation and the errors may not be normally distributed. At least in this limited context, inference based on this estimated smoothing parameter is asymptotically equivalent to the case when the "optimal" value is known.

2. DEFINITION OF A SPLINE ESTIMATE

One property of a spline that distinguishes it from other nonparametric regression estimators is that it is the solution to a variational problem. Let $\mathcal{H} = W_2^2[0, 1) = \{h : h, h'$ are absolutely continuous and $h'' \in L^2[0, 1]\}$. Then, for $\lambda > 0$ the spline estimate, $\hat{f}_\lambda$, is the minimizer of

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

for all $h \in \mathcal{H}$. The integral in (2.1) may be interpreted as a roughness penalty because it increases as the curvature of h increases. Thus the smoothing parameter, λ, controls the relative weight given to the smoothness of h versus its fit to the data. When λ is fixed, $\hat{f}_\lambda$ is a linear function of $\mathbf{Y}$ and it is convenient to define an $n \times n$ "hat" matrix, $A(\lambda)$, depending only on $\{t_k\}$ such that $\hat{\mathbf{f}}_\lambda = A(\lambda)\,\mathbf{Y}$, where $\hat{\mathbf{f}}_\lambda' = \{\hat{f}_\lambda(t_1), \hat{f}_\lambda(t_2), \ldots, \hat{f}_\lambda(t_n)\}$. For details concerning the form of $A(\lambda)$ and its computation see Bates, Lindstrom, Wahba, and Yandell (1986) and Hutchinson and De Hoog (1985). The spline estimate $\hat{f}_\lambda$ is referred to as a natural spline because it satisfies the "natural" boundary conditions $\hat{f}_\lambda^{(2)}(0) = \hat{f}_\lambda^{(2)}(1) = 0$ and $\hat{f}_\lambda^{(3)}(0) = \hat{f}_\lambda^{(3)}(1) = 0$. If $\mathcal{H}$ is taken to be the periodic set of functions $\{h \in W_2^2[0, 1]$: $h^{(j)}(0) = h^{(j)}(1), j = 0, 1\}$, then $\hat{f}_\lambda$ will be a periodic spline satisfying the periodic boundary conditions $\hat{f}_\lambda^{(j)}(0) = \hat{f}_\lambda^{(j)}(1)$, $0 \le j \le 3$. It is important to distinguish between these two estimates, because Wahba's simulation results pertain to periodic splines whereas natural splines are more likely to be used in applications.

The smoothing parameter plays the same role as the bandwidth in a nonparametric regression kernel estimate, and this article concentrates on the statistical properties of $\hat{f}_\lambda$ when λ is selected by a data-based procedure. Specifically, let

$$V(\lambda) = \frac{1}{n}\|(I - A(\lambda))\mathbf{Y}\|^2 \Big/ \left(\frac{1}{n}\operatorname{tr}(I - A(\lambda))\right)^2$$

be the generalized cross-validation function and let $\hat{\lambda}$ denote the global minimum of V. The spline estimate $\hat{f} \equiv \hat{f}_{\hat{\lambda}}$ is the article's focus. This form of cross-validation has worked well in practice; one can show that, asymptotically, $\hat{\lambda}$ will also minimize the average squared error for f (Cox 1984; Craven and Wahba 1979; Härdle, Hall, and Marron 1988; Li 1986).

The rationale for Wahba's confidence intervals comes from the correspondence between $\hat{f}_\lambda$ and the mean of a posterior distribution. Suppose that f is a sample path from the Gaussian process

$$f(t) = \alpha_0 + \alpha_1 t + (\sigma^2/n\lambda)\int_0^t (t - s)\, dW(s), \qquad (2.2)$$

where $W(\cdot)$ is the standard Weiner process and $\boldsymbol{\alpha} \sim N(0, \zeta I)$. With this model for f, a natural spline satisfies the relationships $\hat{f}_\lambda(t) = \lim_{\zeta\to\infty} E(f(t) \mid \mathbf{Y})$ and $\sigma^2 A(\lambda) = \lim_{\zeta\to\infty} \operatorname{cov}(\mathbf{f} \mid \mathbf{Y})$, with $\mathbf{f}' = [f(t_1), f(t_2), \ldots, f(t_n)]$. 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 between a smoothing spline and a posterior mean led Wahba to propose

$$\hat{f}(t_k) \pm Z_{\alpha/2}\sqrt{\hat{\sigma}^2[A(\hat{\lambda})]_{kk}},$$

where $\hat{\sigma}^2 = \|(I - A(\hat{\lambda}))\mathbf{Y}\|^2/\operatorname{tr}(I - A(\hat{\lambda}))$ 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 is that the posterior variance, $\sigma^2 A(\lambda)_{kk}$, is close to the expected average squared error. This observation was first made by Wahba (1983) and is used in Section 3 to explain the accuracy of the ACP for these intervals. This section ends by discussing this connection in more detail.

Let $T_n(\lambda) = (1/n)\sum(\hat{f}(t_k) - f(t_k))^2$, and let λ^o denote the value of the smoothing parameter that minimizes $ET_n(\lambda)$. If f satisfies the conditions of Assumptions 1–3 in Section 5, then Nychka (1986) shows that

$$\frac{\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})/n}{ET_n(\lambda^o)} \xrightarrow{p} K \quad \text{as } n \to \infty, \qquad (2.3)$$

where $K = (32/27)$. Thus, in the limit the average posterior variance will be proportional to the expected average squared error. Now if $\hat{f}$ is a periodic spline and if the observation points are equally spaced, then tr $A(\hat{\lambda})/n = A_{kk}(\hat{\lambda})$ because $A(\lambda)$ is a circulant matrix. For a natural spline, the diagonal elements of $A(\lambda)$ will not be equal. But the variation in these elements over most of the interval will be very small (see Fig. 5) and in the limit these terms will converge uniformly to their average value. Therefore, the asymptotic limit in (2.3) suggests that the standard error of Wahba's intervals will be only about 10% larger $[(32/27)^{1/2} \approx 1.089]$ than $[ET_n(\lambda^o)]^{1/2}$. Surprisingly, the limiting constant in (2.3) depends only on the order of the spline and the smoothness of f. This value differs slightly from 1 because the true function is assumed to have more derivatives than the sample paths implied by the Bayesian prior. A brief derivation of this value is given at the end of the Appendix; see Hall and Titterington (1986) for more background on this topic.

One way of proving the result in (2.3) is by the observation that the average posterior variance is proportional to a consistent estimate of $ET_n(\lambda^o)$ based on the cross-validation function. Also, because the standard error of the spline is related to the expected average squared error, any estimate of this quantity is of interest in its own right. Under Assumptions 1–3 in Section 5, $EV(\lambda) = (ET_n(\lambda) + \sigma^2)(1 + o(1))$ as $n \to \infty$. Rearranging these terms, we have $ET_n(\lambda) = (EV(\lambda) - \sigma^2)(1 + o(1))$ as $n \to \infty$. This asymptotic equivalence suggests that one might estimate $ET_n(\lambda)$ by subtracting an estimate of σ^2 from $V(\lambda)$. One possibility is to let

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

where $\hat{S}_n^2 = \|(I - A(\hat{\lambda}))\mathbf{Y}\|^2/\text{tr}(I - \mathcal{C}A(\hat{\lambda}))$ and $\mathcal{C} = 2 - (1/K)$. Under Assumptions 1–3 in Section 5 it follows that $\hat{T}_n/ET(\lambda_o)$ will converge to 1 in probability as $n \to \infty$. [The value for $\mathcal{C}$ is chosen so that terms of order $ET_n(\lambda_o)$ in the difference between $\hat{T}_n$ and $ET_n(\lambda_o)$ are eliminated.] In general, $\hat{T}_n$ will be a consistent estimate of $ET(\lambda^o)$ provided the average squared bias of the spline estimate is not dominated by effects at the boundaries [see Nychka (1986) for details]. Section 4 describes a method for eliminating boundary effects by restricting the evaluation of the spline estimate to an interval smaller than the range of $\{t_k\}_{k=1,n}$. This estimate of the expected average squared error has worked well in simulations (see Table 3 in Sec. 4), and in Section 5 I outline an argument for its consistency.

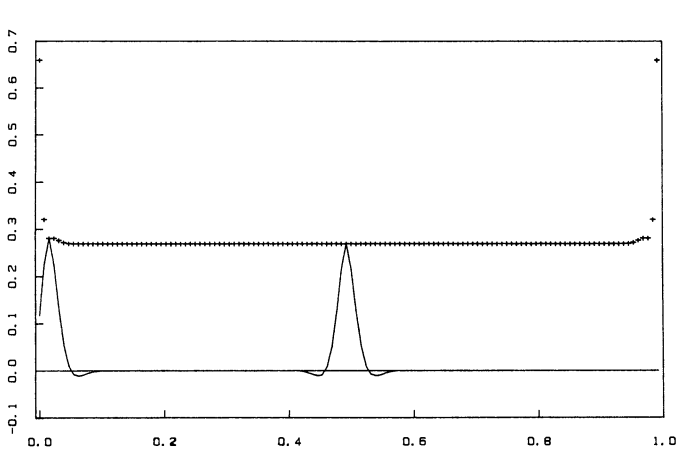

Figure 5. Influence of the Boundaries on the Matrix A(λ). Plotted crosses are $A_{kk}(\lambda)$ versus $t_k = k/n$ when $n = 128$ and $\lambda = 6.99E-4$. Over most of the interval the diagonal elements are essentially constant. Plotted using broken lines are the rows $A_{3,k}(\lambda)$ and $A_{64,k}(\lambda)$. These two curves suggest the shape of the kernel at $t = 3/128$ and $t = 64/128$ when the spline estimate is interpreted as a locally weighted average.

3. FREQUENCY INTERPRETATION FOR PERIODIC SPLINES

This section covers the case when $\hat{f}_\lambda$ is a periodic spline and $\{t_k\}_{k=1,n}$ are equally spaced. This case is considered for two reasons. First, it is the estimate computed by Wahba in her Monte Carlo study. Second, in this simple case the frequency interpretation of these confidence intervals is the least complicated. A general discussion for the natural spline estimate is given in Section 4.

The random variable $\mathcal{U}$ in (1.1) is defined for the case when the measurement errors are normally distributed and the spline estimate is evaluated at an optimal value of the smoothing parameter. Recall that λ^o minimizes the expected average squared error, and let $\hat{f}^o \equiv \hat{f}_{\lambda^o}$. Because λ^o is not a random quantity, it is easy to decompose $\hat{f}^o$ into fixed and random components related to the bias and variance of the estimate. Also, because this estimate is a linear function of the data, the random component will be normally distributed. With this decomposition, it is straightforward to interpret the ACP for this estimate.

In practice, the measurement errors may not follow a normal distribution and λ^o must be estimated. The main mathematical result of this article shows that the difference between the ACP for $\hat{f}$ when λ is determined by generalized cross-validation and the ACP for $\hat{f}^o$ converges to 0 as the sample size increases. This result suggests that the intuition gained from studying the idealized estimate, $\hat{f}^o$, may be useful in interpreting the ACP for the more practical case.

Recall that for a periodic spline with equally spaced knots, the diagonal elements of $A(\lambda)$ are equal. In this case all of the pointwise confidence intervals should have the same width. For now they are represented as

$$\hat{f}^o(t_k) \pm Z_{\alpha/2}D. \tag{3.1}$$

Now let $b(t) = E\hat{f}^o(t) - f(t)$ and $v(t) = \hat{f}^o(t) - E\hat{f}^o(t)$. Recall that τ_n has a distribution independent of $\mathbf{e}$ putting equal mass on the points $\{t_k\}_{1\le k\le n}$. Finally, set $b = b(\tau_n)$ and $v = v(\tau_n)$. It will be argued that the correct choice for D is $[ET(\lambda^o)]^{1/2}$ and $\mathcal{U} = [b + v]/[ET_n(\lambda^o)]^{1/2}$.

The ACP has the form

$$E\frac{1}{n}\sum_{k=1}^{n} I(|\hat{f}^o(t_k) - f(t_k)| \le Z_{\alpha/2}D)$$

$$= \Pr[|b(\tau_n) + v(\tau_n)| \le Z_{\alpha/2}D]$$

$$= \Pr[|(b + v)/D| \le Z_{\alpha/2}].$$

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

Lemma 3.1. If $\hat{f}_\lambda$ is a periodic spline, $t_k = k/n$ ($k = 0, n - 1$), and $\mathbf{e} \sim N(0, \sigma^2 I)$, then (a) $E(b) = 0$, (b) $v \sim N(0, (\sigma^2/n)\text{tr}[A(\lambda^o)^2])$, and (c) b and v are independent.

Proof. (a) Because the constant function is in the null space of the roughness penalty, $(I - A(\lambda^o))\mathbf{1} = 0$. Using the symmetry of A we have $E(b) = (1/n) \sum_{k=1}^{n} (E\hat{f}_\lambda(t_k) - f(t_k)) = (1/n)\mathbf{1}'(I - A(\lambda^o))\,\mathbf{f} = 0$. (b) $\mathbf{v}' = (v(t_1), \ldots, v(t_n)) = (A(\lambda^o)\mathbf{e})'$. Because $\mathbf{e}$ is normally distributed, $\mathbf{v} \sim N(0, \sigma^2 A^2(\lambda_o))$. $A^2(\lambda)$ will also be a circulant matrix, and the diagonal elements can be written as $(1/n)\text{tr}\, A^2(\lambda)$. Thus $v(t_k) = E(v \mid \tau_n = t_k) \sim N(0, \sigma^2 \text{tr}\, A^2(\lambda^o)/n)$. Because this distribution does not depend on k, the unconditional distribution is the same. (c) From the results in (b) it follows that $E(\varphi(v) \mid b)$ will not depend on b for any measurable function φ.

By definition, $\text{var}(b + v) = ET_n(\lambda^o)$, so by setting $D = [ET_n(\lambda^o)]^{1/2}$ it follows that $E(\mathcal{U}) = 0$ and $\text{var}(\mathcal{U}) = 1$. Also, for all of the cases considered by Wahba, $\text{var}(b)/\text{var}(v) < .25$. The distribution of $\mathcal{U}$ should be close to a normal, since it is the convolution of two independent random variables, one normal and the other with a variance that is small relative to the normal component. Figure 6 is a histogram of the distribution of b for Case 2 with $n = 128$ and $\sigma = .05$ (see Fig. 4) along with an appropriately scaled version of the normal density for v. The convolution of these two distributions is close to 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. For all of these cases the distribution of $\mathcal{U}$ differs from a standard normal by less than 1%. The ratio $\text{var}(b)/\text{var}(v)$ is also reported for each case. Note that these values are in rough agreement with asymptotic theory. If f is a second-order smoothing spline (second-derivative roughness penalty) where $(1/n) \sum_{k=1}^{n} (f(t_k) - E\hat{f}_\lambda(t_k))^2 = \gamma\lambda^\alpha(1 + o(1))$ for some $\gamma > 0$ as $\lambda \to 0$, then a simple calculation yields $\text{var}(b)/\text{var}(v) = (1/4\alpha)(1 + o(1))$ (see Nychka 1986, lemma 3.1). For the smooth test functions considered by Wahba, $\alpha = 2$ for Cases 1 and 3 because these functions have three periodically continuous derivatives and $f^{(4)} \in L^2[0, 1]$. Thus we would expect $\text{var}(b)/\text{var}(v) \simeq .125$. For Case 2, $\alpha = 1$ because this function does not have a periodic second derivative and the ratio of variances should be compared with .25. The difference between these actual fractions in Table 1 and the asymptotic values may be due to the fact that the beginning Fourier coefficients for these functions decrease at a rate different from what would be expected in the limit.

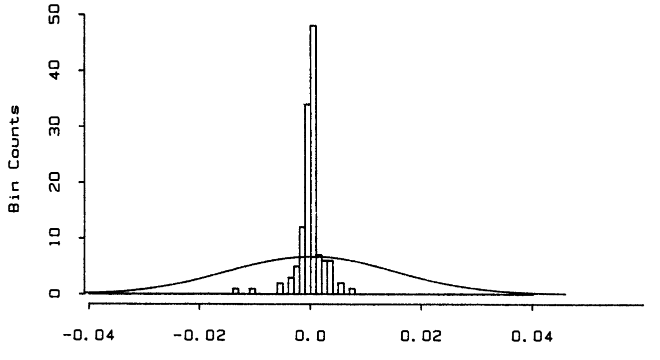

Figure 6. Comparison of the Densities of b and v. The histogram represents the discrete distribution of b for the same parameters as in Figure 4. The random variable v is distributed as N(0, 3.58E−4). This normal density has been superimposed on the plot by scaling by the histogram bin width and the number of observations. The convolution of these two distributions yields a distribution very close to a normal (see Table 1). It is this similarity that helps to explain the accuracy of the ACP.

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

4. EXTENSION TO NATURAL SPLINES

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 I generalize the frequency interpretation of these Bayesian confidence intervals to a natural spline estimate and give a straightforward solution to the problem of boundary effects. To make the presentation less complicated assume 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 article.

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 $v(t_k)$ must now depend on k and hence will not be independent of b. For uniformly distributed $\{t_k\}$ the variation in $A_{kk}(\lambda)$ reflects the fact that f can be estimated more accurately in the middle of the interval compared with the endpoints (see Fig. 5). (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 $\{t_k\}$.)

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

Table 1. Comparison of the Distribution of $\mathcal{U}$ With a Standard Normal for a Periodic Spline

True function	n	$\sigma = .0125$ %K-S	$\sigma = .0125$ var(b)/var(v)	$\sigma = .05$ %K-S	$\sigma = .05$ var(b)/var(v)	$\sigma = .2$ %K-S	$\sigma = .2$ var(b)/var(v)
Case 1	64	.06	.14	.10	.17	.16	.22
	128	.09	.14	.06	.16	.17	.21
	256	.11	.14	.04	.15	.16	.19
Case 2	64	.44	.16	.16	.15	.07	.16
	128	.70	.17	.27	.16	.05	.15
	256	.87	.18	.37	.16	.07	.15
Case 3	64	.13	.13	.12	.15	.09	.19
	128	.14	.13	.13	.14	.08	.18
	256	.14	.13	.13	.14	.10	.17

NOTE: The three functions considered here are plotted in Figure 2. %K-S = 100 × ($\sup_z|\Pr(\mathcal{U} < z) - \Phi(z)|$), where Φ is the standard normal cdf. Also, $\mathcal{U} = (b + v)/[\mathrm{var}(b + v)]^{1/2}$, where $v \sim N(0, (\sigma^2/n)\mathrm{tr}\, A^2(\lambda^o))$ and b has a discrete distribution taking on the values $\{E\hat{f}^o(t_k) - f(t_k)\}$, $1 \le k \le n$, with equal probability. $\hat{f}^o$ is the smoothing spline estimate using an optimal choice for the smoothing parameter. With these definitions var($b + v$) is equal to the expected average squared error of the spline estimate.

A simple way to deal with both of these problems is to restrict the evaluation of the spline to a subinterval of [0, 1]: $\mathcal{I} = [\delta_1, 1 - \delta_2]$ for $\delta_1, \delta_2 > 0$. If $\mathcal{I}$ remains fixed as the sample size increases, then $\hat{f}_\lambda(t)$ for $t \in \mathcal{I}$ will not be influenced by boundary effects (Nychka 1986). 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 $t_k \in \mathcal{I}$. Thus, at least for large sample sizes, a natural spline estimate restricted to an interior interval will have properties similar to those of a periodic spline.

The frequency interpretation for confidence intervals in this setting is similar to that in Section 2. Because $\mathcal{I}$ may be chosen so that the diagonal elements of $A(\lambda)$ for $t_k \in \mathcal{I}$ are essentially constant, consider pointwise confidence intervals of equal width. Let $n_\delta = \#\{t_k \in \mathcal{I}\}$ and

$$T_\delta(\lambda) = \frac{1}{n_\delta} \sum_{t_k \in \mathcal{I}} (\hat{f}_\lambda(t_k) - f(t_k))^2.$$

Let λ_δ^o be the minimizer of $ET_\delta(\lambda)$. Following the development in Section 3, the ACP for the intervals $\hat{f}_{\lambda_\delta^o}(t_k) \pm Z_{\alpha/2}D$ ($t_k \in \mathcal{I}$) is given by $\Pr(|\mathcal{U}_\delta| \le Z_{\alpha/2})$, where $\mathcal{U}_\delta = [\hat{f}_{\lambda_\delta^o}(\tau_n) - f(\tau_n)]/D$ and τ_n assumes the values $\{t_k \in \mathcal{I}\}$ with equal probability. As in 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 $\mathcal{U}_\delta$ is not identically equal to 0 because of the restriction of $\hat{f}$ to points in $\mathcal{I}$. Note that $E(\mathcal{U}_\delta^2) = ET_\delta(\lambda_\delta^o)$. Therefore, to standardize $\mathcal{U}_\delta$ a natural choice for D is $[ET_\delta(\lambda_\delta^o)]^{1/2}$. Despite these departures from the exact relationships in Section 3, the distribution of $\mathcal{U}_\delta$ is close to a standard normal with this choice for D.

Before discussing these results, however, a method for determining $\mathcal{I}$ is presented. For $\rho > 0$ let $\mathcal{I}(\lambda)$ be the smallest interval containing $\{t_j\colon \alpha_j(\lambda) \le 1 + \rho\}$, where $\alpha_j(\lambda) = A_{jj}(\lambda)/\inf_{1\le k\le n} A_{kk}(\lambda)$. By this construction, decreasing ρ will decrease the variation in the diagonal elements of $A(\lambda)$ for $t_k \in \mathcal{I}(\lambda)$. The interval $\mathcal{I}(\lambda^o)$ is used to evaluate the distribution of $\mathcal{U}_\delta$; therefore, two values of λ enter this computation. The interval for evaluation depends on λ^o, and the spline estimate on this interval uses λ_δ^o. Initially, a one-step procedure was considered, where the smoothing parameter minimized

$$E\left[\frac{1}{n_\delta} \sum_{t_k \in \mathcal{I}(\lambda)} (f(t_k) - \hat{f}_\lambda(t_k))^2\right].$$

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 later. In either case, note that the choice of $\mathcal{I}$ is not completely objective because ρ must be specified. In fact, when f has four continuous derivatives, for boundary effects to be asymptotically negligible we must have $\rho = o(n^{1/9}\exp\{-n^{1/9}\})$ as $n \to \infty$. This rate can be derived by relating the optimal choice for the smoothing parameter on an interior interval to an effective bandwidth (Nychka 1986; Silverman 1984). The rate on ρ should be chosen so that the distances between the left endpoint of $\mathcal{I}(\lambda_\delta^o)$ and 0 and the right endpoint of $\mathcal{I}(\lambda_\delta^o)$ and 1 converge to 0 at a slower rate than the optimal effective bandwidth.

Table 2 reports the properties of the ACP when a natural spline is restricted to an interior interval. This table compares the distribution of $\mathcal{U}_\delta$ when $D = [ET_\delta(\lambda_\delta^o)]^{1/2}$ to a standard normal in a format similar to Table 1. The column n_δ/n indicates the fraction of observations included in $\mathcal{I}(\lambda_o)$. In these computations $\rho = .1$; that is, the diagonal elements of $A(\lambda)$ in $\mathcal{I}(\lambda^o)$ differ by at most 10% from their minimum. Note that this restriction does not exclude many data points. To see how much boundary effects influence the distribution of $\mathcal{U}_\delta$, a version of Case 3 was considered, where the function was shifted periodically so that the maximum of the absolute value of the second derivative was at 0 [$f_{\text{shifted}}(t) = f(t + .164 \bmod 1)$]. From asymptotic theory, a large second derivative at 0 should induce large biases at this endpoint. Surprisingly, for the ranges of n and σ considered in these simulations 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

Table 2. Comparison of the Distribution of $\mathcal{U}_\delta$ With a Standard Normal for a Natural Spline Estimate Restricted to a Subinterval

True function	n	σ = .0125 %K-S	σ = .0125 var(b)/var(v)	σ = .0125 % n_δ/n	σ = .05 %K-S	σ = .05 var(b)/var(v)	σ = .05 % n_δ/n	σ = .2 %K-S	σ = .2 var(b)/var(v)	σ = .2 % n_δ/n
Case 1	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 2	64	1.20	.18	96.9	.78	.15	96.9	.25	.15	93.8
	128	.96	.16	96.8	.39	.14	95.3	.18	.15	93.8
	256	1.20	.17	97.7	.50	.14	96.9	.24	.14	94.5
Case 3	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 3 (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 3 (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: Case 3 (shifted) refers to the function in Case 3 translated so that the maximum value of the second derivative is 0. No boundary adjustment means that the spline and the cross-validation function were evaluated on the entire interval %K-S = 100 × $\sup_z|\Pr(\mathcal{U}_\delta < z) - \Phi(z)|$. In each of these cases the subinterval is chosen such that the diagonal elements of $A(\lambda^o)$ differ by less than 10% and n_δ is the number of observations included in the subinterval.

is, $\mathcal{I} = [0, 1]$. Again, there is little difference between the distribution of $\mathcal{U}_\delta$ and a standard normal.

This section ends by summarizing the results of a simulation study when λ^o, λ_δ^o, and $ET(\lambda_\delta^o)$ are estimated. These estimates require the use of cross-validation twice: first to determine $\mathcal{I}$ 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 $\mathcal{I}(\hat{\lambda})$. Now let $\hat{\lambda}_\delta$ be the minimizer of the restricted cross-validation function

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

The natural spline estimate is taken to be $\hat{f} \equiv f_{\hat{\lambda}_\delta}$. Note that $\hat{f}$ is computed using the *full* data set; however, cross-validation is only applied to estimates where $t_k \in \mathcal{I}(\hat{\lambda})$. Besides this restriction, (4.1) is the usual "leave one out" procedure incorporating a special shortcut for splines (Craven and Wahba 1979). $ET(\lambda_\delta^o)$ can be estimated by generalizing (2.4). First, to simplify notation let W denote a diagonal matrix such that $W_{kk} = 1$ when $t_k \in \mathcal{I}(\hat{\lambda})$ and $W_{kk} = 0$ otherwise, and let $\mu(\lambda) = (1/n_\delta)\text{tr}(WA(\lambda))$. With this notation the generalized cross-validation function for the interval $\mathcal{I}(\hat{\lambda})$ is

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

the estimate of S_n^2 is

$$\hat{S}_n^2 = \frac{(1/n_\delta)\|W(I - A(\hat{\lambda}))\mathbf{Y}\|}{1 - \mathcal{C}\mu(\hat{\lambda}_\delta)}, \qquad \mathcal{C} = \frac{37}{32}, \tag{4.2}$$

and we have $\hat{T}_\delta = (V_\delta(\hat{\lambda}_\delta) - \hat{S}_n^2)\varphi(\hat{\lambda}_\delta)$. The constant $\mathcal{C}$ 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_δ^o. The consistency of $\hat{T}_\delta$ is discussed at the end of Section 5. The form for φ is derived in the Appendix.

Table 3 summarizes the results for a simulation study based on the same cases considered in evaluating the distribution of $\mathcal{U}_\delta$. Each case involves 200 repetitions, and 95% confidence intervals were computed of the form $\hat{f}(t_k) \pm 1.96\sqrt{\hat{T}_\delta}$ $[t_k \in \mathcal{I}(\hat{\lambda})]$. The first number in this table is the ACP, the second value is the ratio $E(\hat{T}_\delta)/ET_\delta(\lambda_\delta^o)$, 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 .95; $\hat{T}_\delta$ appears to have little bias. The smoothing spline and the diagonal elements of $A(\lambda)$ were calculated using the order-N algorithm of Hutchinson and De Hoog (1985). The minimizations of $V(\lambda)$ and (4.1) were carried out with respect to $\log(\lambda)$ using a coarse grid search followed by a golden section search. The normal errors were generated by Knuth's version of the Kinderman–Monahan ratio of uniforms.

5. ASYMPTOTIC PROPERTIES OF THE AVERAGE COVERAGE PROBABILITY

This section establishes (1.1), showing that the ACP when λ^o and $ET_n(\lambda^o)$ 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. At the end of the section is a discussion of the modifications necessary to show this equivalence for a natural spline evaluated on a restricted interval. If this theorem was limited to periodic splines with equally spaced observation points, this proof could be greatly simplified by expanding $\hat{f}$ and f in Fourier series (e.g., see Rice and Rosenblatt 1983). This approach has not been taken because it will not generalize easily to the nonperiodic case or to higher-dimensional, thin-plate splines.

Let G_n denote the empirical distribution function for $\{t_k\}_{1\le k\le n}$. Although other cases can be considered, assume that the observation points satisfy one of two cases: (a) fixed design points, $\sup_{u\in[0,1]}|G_n - u| = O(1/n)$; (b) ran-

Table 3. Average Coverage Probabilities for a Natural Spline Estimate Restricted to an Interior Interval

True function	n	σ = .0125 %ACP	σ = .0125 Mean($\hat T$)/ET(λ^o)	σ = .0125 SD($\hat T$)/ET(λ^o)	σ = .05 %ACP	σ = .05 Mean($\hat T$)/ET(λ^o)	σ = .05 SD($\hat T$)/ET(λ^o)	σ = .2 %ACP	σ = .2 Mean($\hat T$)/ET(λ^o)	σ = .2 SD($\hat T$)/ET(λ^o)
Case 1	64	94.6	1.01	.16	94.1	1.02	.21	92.4	1.03	.38
	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
Case 2	64	93.4	1.01	.20	94.1	1.02	.20	94.1	1.04	.20
	128	94.8	1.00	.13	94.2	1.01	.15	94.9	1.02	.15
	256	94.7	.97	.09	94.6	1.02	.14	94.3	1.01	.14
Case 3	64	94.2	1.01	.19	93.9	1.03	.20	93.6	1.00	.18
	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
Case 3 (shifted)	64	94.1	1.01	.19	93.8	1.02	.19	93.3	.98	.19
	128	94.6	1.00	.14	93.9	1.00	.13	92.9	.93	.21
	256	94.1	1.00	.09	94.5	.98	.10	92.9	.93	.14

NOTE: The standard errors (SD's) for the estimated %ACP's range from .22 to .57. $ET(\lambda^o)$ is the minimum expected average squared error, and the estimate $\hat T$ is defined by (4.3). The sample statistics are based on 200 trials of a Monte Carlo simulation.

dom design points, $\{t_k\}_{1\le k\le n}$ is a random sample from the uniform distribution on [0, 1].

Assumption 1. $E(|e_k|^9) < \infty$.

Assumption 2. $\hat\lambda$ is the minimizer of $V(\lambda)$ over the interval $[\lambda_n, \infty]$, where $\lambda_n \sim n^{-8/5}$.

Assumption 3. f is such that for some $\gamma > 0$, $(1/n) \sum_{k=1} b_k^2 = \gamma\lambda^2(1 + o(1))$ uniformly for $\lambda \in [\lambda_n, \infty)$.

Assumption 1 is necessary to guarantee asymptotic normality of the spline estimate. Unfortunately, uniform asymptotic approximations to $\hat f_\lambda$ are only possible for $\lambda \in [\lambda_n, \infty)$, and Assumption 2 insures that $\hat\lambda$ will be in this range. Assumption 3 concerns the asymptotic behavior of the mean squared bias for f. It will hold for second-order periodic splines provided that $f^{(4)} \in L^2[0, 1]$ and $f^{(k)}(0) = f^{(k)}(1)$ for $0 \le k \le 3$.

Theorem 5.1. Let

$$C(t, \alpha) = \hat f(t) \pm Z_{\alpha/2}\sqrt{\hat T_n}, \qquad (5.1)$$

where $\hat f$ is defined in Section 2 and $\hat T_n$ is given by (2.4). Under Assumptions 1–3,

$$\Pr(f(\tau_n) \in C(\tau_n, \alpha)) - \Pr(|\mathcal{U}| \le Z_{\alpha/2}) \to 0 \qquad (5.2)$$

uniformly in α as $n \to \infty$.

Corollary 5.1. If $\hat T_n$ is replaced by $\hat\sigma^2 \operatorname{tr} A(\hat\lambda)/n$ in (5.1), then (5.2) holds provided $\mathcal{U}$ is replaced by $(27/32)\,\mathcal{U}$ [compare (2.3)].

An outline of the proof is given here [see Nychka (1987) for the technical details]. To simplify notation, for $\mathbf{a} \in \mathbf{R}^n$ let $\varphi(\mathbf{a}, c) = (1/n) \sum_{k=1,n} I(|a_k| < c)$, $\mathbf{f}' = (f(t_1), \ldots, f(t_n))$, $\hat{\boldsymbol\omega} = [A(\hat\lambda)\mathbf{Y} - \mathbf{f}]/\sqrt{\hat T_n}$, $\hat{\boldsymbol\omega}^o = [A(\lambda^o)\mathbf{Y} - \mathbf{f}]/[ET_n(\lambda^o)]^{1/2}$, and $z = Z_{\alpha/2}$. With this notation the ACP when λ is estimated by cross-validation is simply $E\varphi(\hat{\mathbf{w}}, z)$. For $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$ and $\varepsilon > 0$, φ satisfies the elementary inequality

$$|\varphi(\mathbf{a}, c) - \varphi(\mathbf{b}, c)| \le \frac{1}{n}\frac{\|\mathbf{a} - \mathbf{b}\|^2}{\varepsilon^2} + |\varphi(\mathbf{b}, c) - \varphi(\mathbf{b}, c + \varepsilon)|, \qquad (5.3)$$

and this is the main device in the proof for bounding expressions. Also, I use a Gaussian approximation to a smoothing spline developed by Cox (1984). Under Assumptions 1–3 it is possible to construct a probability space containing the processes $\hat w(t)$, $\hat w^o(t)$, and $\bar w^o(t)$ such that $\operatorname{law}(\hat{\boldsymbol\omega}) = \operatorname{law}(\hat{\mathbf{w}})$, $\operatorname{law}(\hat{\boldsymbol\omega}^o) = \operatorname{law}(\hat{\mathbf{w}}^o)$, $(1/n)\|\hat{\mathbf{w}}^o - \bar{\mathbf{w}}^o\|^2 \xrightarrow{P} 0$ as $n \to \infty$, and $E\varphi(\bar{\mathbf{w}}^o, z) = \Pr(|\mathcal{U}| < z)$. This last property follows from the fact that $\bar w^o(t)$ is the approximation to a spline estimate based on normal errors.

Outline for Proof of Theorem 5.1. Reexpressing (5.2) with respect to the probability space that contains the Gaussian approximation to $\hat f_\lambda$,

$$\begin{aligned}|E\varphi(\hat{\boldsymbol\omega}, z) - \Pr(|\mathcal{U}| < z)| &= |E\varphi(\hat{\mathbf{w}}, z) - E\varphi(\bar{\mathbf{w}}^o, z)| \\ &\le E|\varphi(\hat{\mathbf{w}}, z) - \varphi(\bar{\mathbf{w}}^o, z)| \\ &\le E|\varphi(\hat{\mathbf{w}}, z) - \varphi(\hat{\mathbf{w}}^o, z)| + E|\varphi(\hat{\mathbf{w}}^o, z) - \varphi(\bar{\mathbf{w}}^o, z)|.\end{aligned} \qquad (5.4)$$

Now applying (5.3) to the first term in (5.4), for any $\varepsilon > 0$

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

Adding and subtracting $\varphi(\bar{\mathbf{w}}^o, z + \varepsilon) - \varphi(\bar{\mathbf{w}}^o, z)$ within the second expression and using (5.3) twice more gives

$$\begin{aligned}&|\varphi(\hat{\mathbf{w}}, z) - \varphi(\hat{\mathbf{w}}^o, z)| \\ &\quad\le \frac{1}{\varepsilon^2}\left\{\frac{1}{n}\|\hat{\mathbf{w}} - \hat{\mathbf{w}}^o\|^2 + \frac{2}{n}\|\hat{\mathbf{w}}^o - \bar{\mathbf{w}}^o\|^2\right\} \\ &\qquad + \{|\varphi(\bar{\mathbf{w}}^o, z + 2\varepsilon) - \varphi(\bar{\mathbf{w}}^o, z + \varepsilon)| \\ &\qquad + |\varphi(\bar{\mathbf{w}}^o, z + \varepsilon) - \varphi(\bar{\mathbf{w}}^o, z)|\} \\ &\quad\le \frac{1}{\varepsilon^2}\{\alpha_n\} + \{\beta_n\}.\end{aligned}$$

Now, $\alpha_n \xrightarrow{P} 0$ as $n \to \infty$ by lemma A.1 of Nychka (1987), the construction of $\bar{\mathbf{w}}^o$, and the consistency of $\hat T_n$. Also, $E(\beta_n) = \mathcal{O}(\varepsilon)$ uniformly for $z \in \mathbf{R}$ (Nychka 1987, lemma A.2). With these results, for any $\delta > 0$ we have $\Pr(|\varphi(\hat{\mathbf{w}}, z) - \varphi(\hat{\mathbf{w}}^o, z)| > \delta) \le_P \Pr(\alpha_n \ge \varepsilon^2\delta) + \mathcal{O}(\varepsilon/\delta)$. Therefore,

$|\varphi(\hat{\mathbf{w}}, z) - \varphi(\hat{\mathbf{w}}^o, z)| \xrightarrow{\cdot} 0$ as $n \to \infty$. Moreover, because this random variable is bounded, $E|\varphi(\hat{\mathbf{w}}, z) - \varphi(\hat{\mathbf{w}}^o, z)| \to 0$ as $n \to \infty$. This takes care of the first term on the left side of (5.4), and the second term can be handled in a similar manner.

The corollary follows from the remarks at the end of Section 2 and by straightforward modifications of the definitions of $\hat{\omega}$ and $\hat{\omega}^o$.

The last part of this section describes how Theorem 5.1 can be extended to a natural spline estimator. To simplify this discussion, assume that $\mathcal{I}$ is a fixed interval rather than the adaptive interval $\mathcal{I}(\hat{\lambda})$ described in Section 4. In the statement of Theorem 5.1, $\mathcal{U}$ and $\hat{T}$ need to be replaced by $\mathcal{U}_\delta$ and $\hat{T}_\delta$, respectively. Also, τ_n should be restricted to $\mathcal{I}$. The corollary will require similar modifications. For Assumption 2 the interval $[\lambda_{n,\infty}]$ should be replaced by $[\lambda_n, \zeta_n]$, where $\zeta_n \to \infty$ as $n \to \infty$. Assumption 3 needs to be changed to read $(1/n_\delta) \sum_{t_k \in \mathcal{I}} b(t_k)^2 = \gamma\lambda^2(1 + o(1))$ uniformly for $\lambda \in [\lambda_n, \zeta_n]$ as $n \to \infty$. This condition will hold provided $f^{(4)}$ is continuous and not identically 0 (Nychka 1986).

The proof of this modified version of Theorem 5.1 depends on the consistency of $\hat{\lambda}_\delta$ and $\hat{T}_\delta$. The convergence of both of these estimates in turn hinges on the uniform convergence of T_δ and V_δ to their expected values. It is necessary that

$$T_\delta(\lambda) - ET_\delta(\lambda) = o_p(ET_\delta(\lambda)) \quad (5.5)$$

and

$$\left(V_\delta(\lambda) - \frac{1}{n_\delta} \sum_{t_k \in \mathcal{I}} e_k^2\right) - (EV_\delta(\lambda) - \sigma^2) = o_p(ET_\delta(\lambda)) \quad (5.6)$$

uniformly for $\lambda \in [\lambda_n, \zeta_n)$ as $n \to \infty$. Given Assumptions 1–3, (5.4), and (5.5), it follows that $\hat{\lambda}_\delta/\lambda_\delta^o \xrightarrow{P} 1$ and $\hat{T}_\delta/ET_\delta(\lambda_\delta^o) \xrightarrow{P} 1$ as $n \to \infty$. Unfortunately, at present there are no published results establishing (5.5) and (5.6) when $\mathcal{I}$ is a subinterval of [0, 1]. Using the kernel approximation to a smoothing spline and the asymptotic theory for kernel estimators (Marron and Härdle 1984), however, it should be possible to prove these relationships.

APPENDIX: DERIVATIONS OF THE BIAS CORRECTION FOR $\hat{T}_\delta$ AND THE CONSTANT K

A bias correction to the cross-validation function is derived to improve the estimates of the average squared error. An argument is given for the form of the constant, K, in (2.3).

Let $m_j = (1/n_\delta)\text{tr}(WA^j(\lambda))$ for $j = 1, 2$, and let $b^2 = (1/n_\delta) \sum_{t_k \in \mathcal{I}(\lambda)} (E\hat{f}_\lambda(t_k) - f(t_k))^2$. Then,

$$EV_\delta(\lambda) - (ET_\delta(\lambda) + \sigma^2) = \frac{b^2 + \sigma^2(1 - 2m_1 + m_2)}{(1 - m_2)^2} - (b^2 + \sigma^2(1 + m_2)).$$

After some algebra,

$$\mathcal{E}(\lambda) = \frac{EV_\delta(\lambda) - \sigma^2 - ET_\delta(\lambda)}{ET_\delta(\lambda)} = \frac{m_1(2 - \alpha_1\alpha_2 - m_1)}{(1 - m_1)^2},$$

where $\alpha_1 = m_1/m_2$ and $\alpha_2 = b^2/(b^2 + \sigma^2 m_2)$. Thus $\varphi(\lambda) = 1/[1 + \mathcal{E}(\lambda)]$ is taken to be the bias correction and $E[(V_\delta(\lambda) - \sigma^2)\varphi(\lambda)] = ET_\delta(\lambda)$. Note that all of the quantities in φ can be computed directly from the diagonal elements of $A(\lambda)$ and $A^2(\lambda)$ except for α_2. Under the assumption that $\hat{\lambda}_\delta$ is a consistent estimate of λ_δ^o and that b^2 satisfies Assumption 3, it is possible to show that α_2 converges to $\frac{1}{9}$.

Now I provide a heuristic derivation for the limit in (2.3). Assume that

$$\frac{\hat{\sigma}^2 \text{tr } A(\hat{\lambda})/n}{\sigma^2 \text{tr } A(\lambda^o)/n} \xrightarrow{P} 1 \quad \text{as } n \to \infty.$$

Thus, using the more general notation,

$$K = \frac{\sigma^2 A(\lambda^o)/n}{ET(\lambda^o)} = \frac{\sigma^2 m_1}{b^2 + \sigma^2 m_2}$$
$$= \left[\frac{m_1}{m_2}\right]\left[\frac{\sigma^2 m_2}{b^2 + \sigma^2 m_2}\right] = \left[\frac{1}{\alpha_1}\right][1 - \alpha_2].$$

For a second-order smoothing spline it is known that, under Assumption 2, α_1 converges to $\frac{9}{4}$ (e.g., see Nychka 1986, lemma 3.1). Also, from the previous discussion $1 - \alpha_2$ has the limit $\frac{8}{9}$. The value for K now follows.

[*Received January 1987. Revised May 1988.*]

REFERENCES

Bates, D., Lindstrom, M., Wahba, G., and Yandell, B. (1986), "GCVPACK-Routines for Generalized Cross-Validation," Technical Report 775, University of Wisconsin–Madison, Dept. of Statistics.

Cox, D. D. (1984), "Gaussian Approximation of Smoothing Splines," Technical Report 743, University of Wisconsin–Madison, Dept. of Statistics.

——— (1986), "Approximation Theory of Method of Regularization Estimators: Applications," unpublished manuscript.

Craven, P., and Wahba, G. (1979), "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik*, 31, 377–403.

Hall, P., and Titterington, M. (1986), "Common Structure of Techniques for Choosing Smoothing Parameters in Regression Problems," unpublished manuscript.

Härdle, W., and Bowman, A. (1988), "Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands," *Journal of the American Statistical Association*, 83, 102–110.

Härdle, W., Hall, P., and Marron, S. (1988), "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?" *Journal of the American Statistical Association*, 83, 86–95.

Hutchinson, M. F., and De Hoog, F. R. (1985), "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik*, 47, 107–122.

Li, K. (1986), "Asymptotic Optimality of C_L and Generalized Cross-Validation in Ridge Regression With an Application to Spline Smoothing," *The Annals of Statistics*, 14, 1101–1112.

Marron, J. S., and Härdle, W. (1984), "Random Approximations to Some Measures of Accuracy in Nonparametric Curve Estimation," North Carolina Institute of Statistics Mimeo 1566, North Carolina State University.

Messer, K. (1986), "A Comparison of a Spline Estimate to Its 'Equivalent' Kernel Estimate," unpublished manuscript submitted to *The Annals of Statistics*.

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 1688, North Carolina State University.

——— (1987), "A Frequency Interpretation of 'Bayesian' Confidence Intervals for a Smoothing Spline," North Carolina Institute of Statistics Mimeo 1699, North Carolina State University.

Rice, J., and Rosenblatt, M. (1983), "Smoothing Splines: Regression, Derivatives, and Deconvolution," *The Annals of Statistics*, 11, 141–156.

Silverman, B. W. (1984), "Spline Smoothing: The Equivalent Variable Kernel Method," *The Annals of Statistics*, 12, 898–916.

——— (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," *Journal of the Royal Statistical Society*, Ser. B, 47, 1–52.

Staniswallis, J. (1986), "Local Bandwidth Selection for Kernel Estimates," unpublished manuscript.

Wahba, G. (1983), "Bayesian 'Confidence Intervals' for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society*, Ser. B, 45, 133–150.

——— (1985), Discussion of "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," by B. W. Silverman, *Journal of the Royal Statistical Society*, Ser. B, 44, 47.