The Annals of Statistics
1990, Vol. 18, No. 1, 415–428

THE AVERAGE POSTERIOR VARIANCE OF A SMOOTHING SPLINE AND A CONSISTENT ESTIMATE OF THE AVERAGE SQUARED ERROR

BY DOUGLAS NYCHKA

North Carolina State University

A smoothing spline estimator can be interpreted in two ways: either as the solution to a variational problem or as the posterior mean when a particular Gaussian prior is placed on the unknown regression function. In order to explain the remarkable performance of her Bayesian "confidence intervals" in a simulation study, Wahba conjectured that the average posterior variance of a spline estimate evaluated at the observation points will be close to the expected average squared error. The estimate of the average posterior variance proposed by Wahba is shown to converge in probability to a quantity proportional to the expected average squared error. This result is established by relating this statistic to a consistent risk estimate based on generalized cross-validation.

1. Introduction. Consider the model

$$Y_{kn} = f(t_{kn}) + e_{kn}, \qquad 1 \le k \le n, \tag{1.1}$$

where f is a smooth, unknown function evaluated at the points $0 \le t_{1n} \le t_{2n} \le \cdots \le t_{nn} \le 1$ and $\{e_{kn}\}_{1 \le k \le n}$ are independent and identically distributed errors with $E(e) = 0$, $E(e^2) = \sigma^2$. One statistical problem posed by this observational model is to estimate f without assuming a particular parametric form for this unknown function. A promising nonparametric estimate of f that has been successful in a diverse range of applications is a smoothing spline. [See Silverman (1985) for a review.] One feature that distinguishes spline estimators from other methods of nonparametric regression is that under a suitable prior (partially improper) they are Bayes estimates [Wahba (1978)]. This paper will study the connection between Bayesian and frequentist interpretations of a spline estimate by showing that under suitable regularity conditions the average posterior variance is proportional to a consistent estimate of the expected average squared error. These results help in interpreting recent work on confidence procedures for spline estimates.

For $\lambda > 0$, f_λ, the mth order, natural smoothing spline estimate for f, is the minimizer of

$$\frac{1}{n}\sum_{k=1}^{n}\left(Y_{kn} - h(t_{kn})\right)^2 + \lambda\int_{[0,1]}\left(h^{(m)}(t)\right)^2 dt, \quad \text{for all } h \in W_2^m[0,1], \tag{1.2}$$

Received July 1986; revised December 1988.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G15.

Key words and phrases. Nonparametric regression, smoothing spline, confidence intervals, risk estimate, cross-validation.

where

$$W_2^m[0,1] = \{f\colon f^{(m)} \in L^2[0,1] \text{ and } f^{(k)},\ 1 \le k \le m-1, \text{ are absolutely continuous}\}.$$

When the smoothing parameter λ is fixed, f_λ is a linear function of $\mathbf{Y}$. Thus, there is a matrix $A(\lambda)$, such that $\mathbf{f}_\lambda = A(\lambda)\mathbf{Y}$ with $\mathbf{f}_\lambda' = (f_\lambda(t_{1n}), \dots, f_\lambda(t_{nn}))$. This nonparametric estimate for f has an intriguing stochastic interpretation. Suppose $\mathbf{e}_n \sim N(\mathbf{0}, \sigma^2\mathrm{I})$ and f is a realization from the Gaussian process

$$\sum_{j=1}^{m} \theta_j t^{j-1} + \frac{\sigma^2}{n\lambda} \int_0^t \frac{(t-s)^{m-1}}{(m-1)!}\, dW(s),$$

where $\theta \sim N(0, \xi I)$ and $W(\cdot)$ is the standard Wiener process with $W(0) = 0$. If f_λ is a natural spline estimate, then

$$f_\lambda(t) = \lim_{\xi \to \infty} E(f(t)|\mathbf{Y})$$

and the limiting posterior covariance is

$$\sigma^2 A(\lambda) = \lim_{\xi \to \infty} \mathrm{Var}(\mathbf{f}|\mathbf{Y}).$$

Wahba (1983) used this correspondence between a smoothing spline estimator and the posterior distribution of f to motivate 95% pointwise "confidence" intervals for $f(t_{kn})$ of the form

$$f_{\hat\lambda}(t_{kn}) \pm 1.96\hat\sigma\sqrt{A_{kk}(\hat\lambda)}\,.$$

Here $\hat\lambda$ is the minimizer of the generalized cross-validation function

$$V(\lambda) = \frac{(1/n)\,\|(I - A(\lambda))\mathbf{Y}\|^2}{((1/n)\mathrm{tr}(I - A(\lambda)))^2}$$

and

$$\hat\sigma^2 = \frac{\|(I - A(\hat\lambda))\mathbf{Y}\|^2}{\mathrm{tr}(I - A(\hat\lambda))}$$

is an estimator for σ^2. Although these intervals were derived within a Bayesian framework, the simulation results reported in Wahba (1983) indicate that they work well when evaluated by a frequentist criterion for fixed functions. (It is for this reason that we refer to these as confidence intervals rather than regions of high posterior probability.) Wahba's confidence procedure is one of the few approaches that has been developed for spline estimates and, given the growing interest in the use of smoothing splines for data analysis, it is important to understand the statistical properties of this method. To explain the success of this confidence procedure, Wahba hypothesized that the average posterior variance (APV) is close to the expectation of the average squared error (ASE). The

APV is given by

$$\frac{\sigma^2}{n}\sum_{k=1}^{n} A_{kk}(\lambda) = \frac{\sigma^2 \operatorname{tr} A(\lambda)}{n},$$

while the ASE is

$$T_n(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\left(f_\lambda(t_{kn}) - f(t_{kn})\right)^2.$$

Now let $ET_n(\lambda)$ denote the expected ASE (where now f is taken to be a fixed function) and let λ° be the minimizer of $ET_n(\lambda)$ for $\lambda \in [0, \infty)$. With this notation Wahba's conjecture is

$$\text{If } f \in W_2^m[0,1], \text{ then } \frac{\sigma^2 \operatorname{tr} A(\lambda^\circ)/n}{ET_n(\lambda^\circ)} = \kappa(1 + o(1)) \quad \text{as } n \to \infty \tag{1.3}$$
$$\text{for some } \kappa \in \left[1, \left(\frac{2m}{2m-1}\right)\left(\frac{4m}{4m+1}\right)\right].$$

Unfortunately, Wahba's argument has several places that are heuristic. Also, her analysis does not acknowledge the fact that the value for λ was not fixed in the simulations, but rather was determined by generalized cross-validation. Recently Hall and Titterington (1987) have given a proof for Wahba's conjecture in the context of periodic splines, but these authors also restrict themselves to a deterministic choice of the smoothing parameter.

In this paper we give a proof for a version of Wahba's conjecture that accounts for the adaptive choice of the smoothing parameter. To establish this result we consider an estimator of the expected ASE obtained by subtracting an estimate of $(1/n)\sum_{k=1}^{n} e_{kn}^2$ from the generalized cross-validation function. This estimator is consistent in the sense that the ratio of this estimator to the expected ASE converges to 1 in probability as $n \to \infty$. (In the rest of this paper, this property of the estimator will be referred to as simply being consistent.) Asymptotically, this estimator also turns out to be proportional to the estimated APV proposed by Wahba. Given this relationship, one version of Wahba's conjecture is simple to establish.

Wahba's conjecture is important because it links a frequency quantity with a functional of the posterior distribution. In fact, this correspondence provides a simple explanation for the accuracy of the confidence intervals in Wahba's Monte Carlo study. From now on it will be assumed that f is a fixed function. Consider the random variable $U = (f_{\lambda^\circ}(\tau) - f(\tau))$ where τ is a random variable independent of $\mathbf{e}_n$ taking on the values $\{t_{kn}\}_{1 \le k \le n}$ with equal probability. U has a mean of zero [because $A(\lambda)$ is symmetric and $A(\lambda)\mathbf{1} = \mathbf{1}$], has variance $ET_n(\lambda^\circ)$ and, for the test cases considered by Wahba, is approximately normally distributed [see Nychka (1988)]. Thus, if one knew $ET_n(\lambda^\circ)$, then one might consider the $(1 - \alpha)$ 100% confidence interval for $f(\tau)$:

$$f_{\lambda^\circ}(\tau) \pm Z_{\alpha/2}\sqrt{ET_n(\lambda^\circ)}. \tag{1.4}$$

The confidence procedure studied by Wahba yields intervals that are similar to (1.4). Because the periodic splines used in Wahba's study were applied to equally spaced data, $A(\lambda)$ is a circulant matrix and thus $A_{kk}(\lambda) = \operatorname{tr} A(\lambda)/n$. Moreover, Wahba evaluated her confidence procedure using the average coverage probability of the pointwise confidence intervals at $\{t_{kn}\}_{1 \le k \le n}$. This is equivalent to computing the confidence level for the interval at the random point τ. Thus we are left with interpreting the coverage probability of intervals of the special form

$$f_{\hat{\lambda}}(\tau) \pm Z_{\alpha/2} \sqrt{\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})/n}\,. \tag{1.5}$$

One possibility is to consider the difference in coverage probabilities between (1.4) and (1.5) as $n \to \infty$. In this situation Nychka (1988) shows that $f_{\hat{\lambda}}(\tau) - f(\tau) - U = o_P(ET_n(\lambda^\circ))$ and thus the distribution of U provides an approximation to the distribution $f_{\hat{\lambda}}(\tau) - f(\tau)$ as $n \to \infty$. The other ingredient to this analysis, which is the focus of this article, is the relationship between the standard errors for these intervals. Under suitable assumptions the ratio of these two quantities converges to a constant that is close to 1. This similarity between the standard errors implies that the coverage probability for (1.5) will not be very different from that of (1.4) [see Nychka (1988) for more details].

With this motivation, the main result of this article will be stated.

THEOREM 1.1. *Suppose that the observation points* $\{t_{kn}\}_{1 \le k \le n}$ *are a random sample from a distribution with density function* g *such that* g *is strictly positive on* $[0,1]$ *and* $g \in C^\infty[0,1]$. *If* $E|e_{kn}|^8 < \infty$, $\hat{\lambda}$ *is the minimizer of* $V(\lambda)$ *restricted to* $[\lambda_n, \infty)$ *with* $\lambda_n \sim n^{-4m/5}$, f_λ *is a natural spline with* $m \ge 2$, $f \in W_2^{2m}$, *and* f *satisfies the natural boundary conditions*

$$f^{(k)}(0) = f^{(k)}(1) = 0, \qquad m \le k \le 2m - 1, \tag{1.6}$$

then

$$\frac{\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})/n}{ET_n(\lambda^\circ)} \xrightarrow{P} K \quad \text{as } n \to \infty \tag{1.7}$$

where

$$K = \left(\frac{2m}{2m-1}\right)\left(\frac{4m}{4m+1}\right).$$

The hypotheses of this theorem differ in several significant ways from Wahba's original formulation. First, note that the frequency interpretation of the average coverage probability outlined above does not really depend on the errors being normally distributed. All that is required is that U be approximately normal. For this reason, the consistency of the estimated APV is considered when the normal assumption has been replaced by the boundedness of high order moments of the error distribution. By placing more restrictions on f, convergence is obtained to a particular constant that is actually the lower endpoint for the range of κ given in (1.3). More will be said about the boundary conditions required of f in the next section. Theorem 1.1 will also hold if f_λ is a periodic spline, provided that the natural boundary conditions in (1.6) are replaced by the periodic ones: $f^{(k)}(0) = f^{(k)}(1)$, for $0 \leq k \leq 2m - 1$. Finally, these results are not restricted to random sequences of observation points and in Section 2 conditions are also given for deterministic sequences.

Theorem 1.1 will be established by comparing $\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})/n$ to an estimator of $ET_n(\lambda°)$ based on the generalized cross-validation function. Besides giving a clear proof, this risk estimate may be of interest in its own right. From Speckman (1983) and Cox (1984), under the conditions of Theorem 1.1, if $S_n^2 = (1/n)\sum_{k=1}^n e_k^2$, then

$$\frac{V(\hat{\lambda}) - S_n^2}{ET_n(\lambda°)} \to_P 1 \quad \text{as } n \to \infty. \tag{1.8}$$

Now suppose $\hat{S}_n^2$ is an estimator of S_n^2 such that $S_n^2 - \hat{S}_n^2 = o_p(ET_n(\lambda°))$. Then a natural estimate of $ET_n(\lambda°)$ is

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

With this motivation, a key point of this paper is to establish the following:

THEOREM 1.2. *Under the same hypotheses as Theorem* 1.1, *if*

$$\hat{S}_n^2 = \frac{\|(I - A(\hat{\lambda}))\mathbf{Y}\|^2}{\operatorname{tr}(I - \mathscr{C}A(\hat{\lambda}))} \quad \text{with } \mathscr{C} = 2 - \frac{1}{K}, \tag{1.9}$$

then

$$\hat{S}_n^2 - S_n^2 = o_P(ET_n(\lambda°)) \quad \text{as } n \to \infty.$$

The proof is given in Sections 3 and 4.

Theorem 1.1 now follows easily from the conclusion of this theorem and the asymptotic behavior of $\operatorname{tr} A(\hat{\lambda})/n$. With some algebra one can show

$$\hat{T}_n = (2 - \mathscr{C})\left[\frac{\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})}{n}\right]\left[\frac{1 + \beta_n^2/(2 - \mathscr{C})}{(1 - \mathscr{C}\beta_n)(1 - \beta_n)^2}\right], \tag{1.10}$$

where $\beta_n = \operatorname{tr} A(\hat{\lambda})/n$. Thus, $\hat{T}_n$ is proportional to the estimated APV. Also, from Theorem 1.2 and (1.8), $\hat{T}_n/ET_n(\lambda°) \to_P 1$. From Lemma 3.1 and the choice of λ_n, the second bracketed term in (1.10) converges to 1 in probability as $n \to \infty$. Noting that $K = 1/(2 - \mathscr{C})$, Theorem 1.1 now follows.

The next section states general versions of Theorems 1.1 and 1.2, discusses the hypothesis concerning boundary conditions on f and suggests a method for eliminating this hypothesis using a semiparametric spline model. Section 3 develops some preliminary lemmas, while the proofs of the theorems are given in Section 4. We end this introduction by discussing some other estimators of the error variance and the expected ASE.

Arguing by analogy with ordinary linear regression, Wahba suggested $\hat{\sigma}^2$ as an estimator for σ^2. Although $\hat{\sigma}^2 - S_n^2 = o_P(1)$, under the hypotheses of Theorem

1.1, $(\hat{\sigma}^2 - S_n^2)/ET_n(\lambda^\circ) \to_P \mathscr{E}$ where $\mathscr{E} \neq 0$. For this reason, $V(\lambda) - \hat{\sigma}^2$ will not be a consistent estimator of $ET_n(\lambda^\circ)$. Note that the estimator of S_n^2 given in Theorem 1.2 is only a slight modification of $\hat{\sigma}^2$. In the denominator of $\hat{\sigma}^2$, $\operatorname{tr}(I - A(\hat{\lambda}))$ has been replaced by $\operatorname{tr}(I - \mathscr{C}A(\hat{\lambda}))$, where $\mathscr{C}$ is a constant depending only on m. There are other ways of modifying $\hat{\sigma}^2$ to yield consistent (in the sense of Theorem 1.2) estimators for S_n^2. One interesting approach, that can be inferred from Hall and Titterington (1987), is to replace $\hat{\lambda}$ in the definition of $\hat{\sigma}^2$ by a slightly smaller value, $\bar{\lambda}$, where

$$\bar{\lambda} = \hat{\lambda}\left(\frac{4m}{2m-1}\right)^{2m/(4m+1)}$$

The proof of Theorem 1.2 can be easily adapted to handle this alternative estimator for S_n^2. One estimate of σ^2 that has some interesting minimum mean squared error properties is given by Buckley, Eagleson and Silverman (1988).

Besides an estimate of σ^2 based on the residual sum of squares, estimates for σ^2 can be constructed using first (or second) differences of $\mathbf{Y}$ [Rice (1984)]. This latter estimate has the advantage that no assumptions need to be made about the functional form for the bias of f_λ. Unfortunately, estimators based on differencing $\mathbf{Y}$ will not yield suitable estimates of S_n^2.

Risk estimators for a smoothing spline have been developed through the connection between Stein's unbiased risk estimate and generalized cross-validation in Li (1986). One interesting feature of Li's work is that his risk estimate only requires an estimate of σ^2 rather than S_n^2. His results, however, are limited to the case of normal errors and involve the asymptotic approximation of f_λ by a Stein estimate. Also, the reader is referred to Rice (1984) because, in the special case when $t_{kn} = k/n$, a periodic smoothing spline is also a kernel estimate.

2. General theorem.

In this section a general theorem that includes the results of Theorems 1.1 and 1.2 will be stated. This theorem depends on the three conditions F1–F3. The last of these assumptions is fairly restrictive and a semiparametric method is suggested for avoiding some of its restrictions in practice.

Let G_n denote the empirical distribution for the design points, $\{t_{kn}\}_{1 \le k \le n}$. Following Cox (1984) (subsequently abbreviated GA) two cases of knot sequences will be considered.

CASE A (Designed knots). There is a distribution function G such that

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

CASE B (Random knots). $\{t_{kn}\}$ is a random from a distribution with c.d.f. G.

In either case we will assume that $g = (d/dv)G$ is strictly positive on $[0,1]$ and $g \in C^{\infty}[0,1]$.

(F1) $E|e_{nk}|^{2+\nu} < \infty$ with

Case A (Designed knots): $\nu > 4m - 1$.

Case B (Random knots): $\nu > 2(8m - 3)/5$.

Unfortunately, due to the difficulty in obtaining uniform asymptotic approximations for f_λ over all λ, we must restrict attention to values of the smoothing parameter in an interval $[\lambda_n, \infty)$ with $\lambda_n \to 0$ at a particular rate. Thus, the smoothing parameter estimate is redefined as

$$\hat{\lambda} = \operatorname*{arg\,min}_{\lambda \in [\lambda_n, \infty)} V(\lambda).$$

(F2)

Case A (Designed knots): $\lambda_n \approx n^{-4m/5} \log(n)$.

Case B (Random knots): $\lambda_n \approx n^{-2m/5} \log(n)^m$.

(F3) There is a $\gamma > 0$ such that

$$\frac{1}{n} \sum_{k=1}^{n} \left(Ef_\lambda(t_{kn}) - f(t_{kn})\right)^2 = \gamma\lambda^2(1 + o(1)) \quad \text{uniformly for } \lambda \in [\lambda_{n,\infty}).$$

THEOREM 2.1. *Under* (F1)–(F3) *if* $\hat{S}_n^2$ *is given by* (1.9) *and* $\hat{T}_n = V(\hat{\lambda}) - \hat{S}_n^2$, *then*

$$S_n^2 - \hat{S}_n^2 = o_P(ET_n(\lambda^\circ)), \tag{2.1a}$$

$$\frac{\hat{T}_n}{ET_n(\lambda^\circ)} \to_P 1, \tag{2.1b}$$

$$\frac{\hat{\sigma}^2 \operatorname{tr} A(\hat{\lambda})/n}{ET_n(\lambda^\circ)} \to_P K, \tag{2.1c}$$

where

$$K = \left(\frac{2m}{2m-1}\right)\left(\frac{4m}{4m+1}\right) \quad \text{as } n \to \infty.$$

The proof of this theorem is given in Section 4.

We end this section by discussing some of the more restrictive hypotheses required for this theorem. Perhaps the most stringent assumption is (F3), assuming a specific asymptotic functional form for the average squared bias. Let $\mathscr{S}$ denote the class of functions that satisfy (F3). Members of $\mathscr{S}$ are the "very smooth" functions defined in Wahba (1977) or the case $p = 2$ in the Appendix of Wahba (1983). If f_λ is an mth order, natural smoothing spline, then

$$\mathscr{S} = \Big\{ f \colon f \in W_2^{2m}[0,1],\ f^{(k)}(0) = f^{(k)}(1) = 0,\ m \le k \le 2m-1,\ \|f^{(m)}\|_{L_2[0,1]} > 0 \Big\}.$$

This class is considered because for $f \in \mathscr{S}$, $ET_n(\lambda^\circ)$ achieves its fastest rate of convergence and is not dominated by end effects. (The last condition of $f^{(m)}$ excludes polynomials up to degree $m - 1$ from being members of $\mathscr{S}$.) Also, this assumption leads to explicit asymptotic expressions for $ET_n(\lambda)$ and λ°. The crucial use of these expressions is in Lemma 3.5 where it is necessary to know the limiting ratio of the average squared bias to the expected ASE when $\lambda = \lambda^\circ$. [Under (F3) this ratio is simply $1/(4m + 1)$.]

Although members of $\mathscr{S}$ must have more derivatives than required by the penalty function in (1.2), the main objection to this class is that f must satisfy boundary conditions in its higher derivatives. Unfortunately these conditions are necessary for (F3) to hold. If $f \in W_2^{2m}[0,1]$ but violates one or more of the boundary conditions, then the average squared bias must converge to zero at a lower rate than λ^2. In addition, the average squared bias will be dominated by the bias of the estimator in neighborhoods of 0 and 1. Thus the average squared error (and the cross-validation function) may be influenced by the behavior of the estimate at the boundaries. The reader is referred to Rice and Rosenblatt (1983) and Messer (1989) for a detailed discussion of this phenomenon.

Although these theoretical results suggest that boundary conditions may be important, in practice there is usually no boundary information available. One possible solution to this problem is to use a semiparametric model to account for the behavior of f at the boundaries and a smoothing spline to estimate the remaining portion. For example, suppose $m = 2$. One can easily construct functions ψ_k, $1 \le k \le 4$, such for all $f \in W_2^{2m}[0,1]$, $f = \sum_{k=1}^4 \psi_k a_k + h$ where $h \in \mathscr{S}$. Using the boundary adjusted smoothing spline approach developed by P. Speckman and J. H. Shiau, the parameter vector $\mathbf{a}$ can be estimated such that $E\,\|\mathbf{a} - \hat{\mathbf{a}}\|^2 = O(\lambda^2)$. With these estimates one would set $y_k = Y_k - \sum_{k=1}^4 \hat{a}_k \psi(t_k)$ and then apply the usual spline smoothing techniques to this residual vector to obtain an estimate for h.

By construction, h will satisfy (F3) and it is conjectured that Theorem 2.1 will still hold when $\mathbf{y}$ replaces $\mathbf{Y}$ and h replaces f in the definitions of $\hat{S}_n^2$, $T_n(\lambda)$ and $V(\lambda)$. Due to the nonparametric nature of the spline estimate, $ET_n(\lambda^\circ)$ will have the same convergence rate as $E\,\|\mathbf{a} - \hat{\mathbf{a}}\|^2$. Using this fact and the mathematical tools developed in Speckman (1983) and GA, it is believed that the convergence of the above quantities can be proved. It is hoped that this brief discussion may motivate other researchers to study the properties of this type of boundary adjustment.

3. Preliminary results. The proof of Theorem 2.1 will depend on the six lemmas given below. The first lemma is a collection of some well known asymptotic properties of a smoothing spline, while the second establishes the consistency of $\hat{\lambda}$. The remaining lemmas can be motivated by examining the terms arising from the expansion $\hat{S}_n^2 - S_n^2$ at lines (4.1) and (4.2) in the proof of Theorem 2.1. For all of the following lemmas it should be assumed that (F1)–(F3) are in force and $A(\lambda)$ may be interpreted to be the "hat" matrix for either a natural or periodic spline.

To simplify notation the following definitions will be used:

$$m_k(\lambda) = \frac{1}{n}\operatorname{tr}\left[A(\lambda)^k\right],$$

$$\mu_k(\lambda) = \alpha l_k \frac{\lambda^{-1/2m}}{n},$$

where

$$l_k = \int_0^\infty \frac{dv}{(1+v^{2m})^k}, \qquad \alpha = \frac{\pi}{\int_0^1 (g(v))^{1/2m}\,dv} \quad \text{and} \quad k = 1,2.$$

LEMMA 3.1. *Under* (F1)–(F3), *as* $n \to \infty$,

$$m_k(\lambda) = \mu_k(\lambda)(1 + o(1)), \qquad k = 1,2, \tag{3.1}$$

uniformly for $\lambda \in [\lambda_n, \infty]$,

$$ET_n(\lambda) = \left(\gamma\lambda^2 + \sigma^2\mu_2(\lambda)\right)(1 + o(1)), \tag{3.2}$$

uniformly for $\lambda \in [\lambda_n, \infty]$,

$$\sup_{\lambda \in [\lambda_n, \infty)} \left| \frac{T_n(\lambda) - ET_n(\lambda)}{ET_n(\lambda)} \right| = o_P(1), \tag{3.3a}$$

$$\sup_{\lambda \in [\lambda_n, \infty)} \left| \frac{V(\lambda) - S_n^2 - T_n(\lambda)}{ET_n(\lambda)} \right| = o_P(1), \tag{3.3b}$$

$$\lambda^\circ = \left(\frac{\alpha l_2 \sigma^2}{n\gamma 4m} \right)^{2m/(1+4m)} (1 + o(1)), \tag{3.4}$$

$$\frac{T_n(\hat{\lambda})}{ET_n(\lambda^\circ)} = 1 + o_P(1) \tag{3.5}$$

and

$$\frac{\mu_k(\lambda^\circ)}{ET_n(\lambda^\circ)} = O(1) \quad \text{for } k = 1,2. \tag{3.6}$$

PROOF. Examining the proof of Lemma 4.4 from GA, Cox actually shows that $m_k(\lambda) = \sum_{j=1}^{\infty}(1 + \lambda\delta_j)^{-k}(1 + o(1))$ for $\lambda \in [\lambda_n, \infty)$. The eigenvalues δ_j are obtained from the differential operator $L\varphi = (-1)^m\varphi^{(2m)}/g$ where the domain for L consists of all members of W_2^{2m} that satisfy the natural (or periodic) boundary conditions. From Naimark [(1967), pages 78–79], $\delta_j = \alpha^{2m}j^{2m}(1 + o(1))$

and it follows that

$$\sum_{j=1}^{\infty}\left(1+\lambda\delta_j\right)^{-k}=\sum_{j=1}^{\infty}\left(1+\lambda(\alpha j)^{2m}\right)^{-k}(1+o(1))\quad \text{as } \lambda\to 0.$$

Applying arguments similar to those used in the proof of Lemma 2.1 in Cox (1988),

$$\sum_{j=1}^{\infty}\left(1+\lambda(\alpha j)^{2m}\right)^{-k}=\alpha l_k\lambda^{-1/2m}(1+o(1))\quad \text{as } \lambda\to 0.$$

The equivalence in (3.2) follows trivially from (3.1) and (F3). Both (3.3a) and (3.3b) are proved by Speckman (1983) for normally distributed errors and are generalized in the proof of Theorem 5.1 of GA for error distributions that satisfy the moment conditions in (F2).

The expression (3.4) is established in Wahba (1977) and can be verified by computing the minimum of the rhs of (3.2). Because $\lambda^{\circ}\in[\lambda_n,\infty)$ as $n\to\infty$, (3.5) can be easily derived from the preceding results. Finally, (3.6) may be verified by substituting the asymptotic form for λ° into the rhs of (3.2). □

LEMMA 3.2. *If $\hat{\lambda}$ is the minimizer of $V(\lambda)$ for $\lambda\in[\lambda_n,\infty)$, then*

$$\hat{\lambda}=\lambda^{\circ}(1+o_P(1))\quad \text{as } n\to\infty.$$

PROOF. Let $g(\lambda)=\gamma\lambda^2+\sigma^2\mu_2(\lambda)$ and from (3.2) and (3.3a) we have

$$\sup_{\lambda\in[\lambda_n,\infty)}\left|\frac{T_n(\lambda)-g(\lambda)}{g(\lambda)}\right|=o_P(1)$$

and thus $(T_n(\hat{\lambda})-g(\hat{\lambda}))/g(\hat{\lambda})=o_P(1)$ or $g(\hat{\lambda})/T_n(\hat{\lambda})=1+o_P(1)$. Combining this asymptotic equivalence with (3.2) and (3.5) it now follows that

$$\frac{g(\hat{\lambda})}{g(\lambda^{\circ})}=\frac{g(\hat{\lambda})}{T_n(\hat{\lambda})}\frac{T_n(\hat{\lambda})}{ET_n(\lambda^{\circ})}\frac{ET_n(\lambda^{\circ})}{g(\lambda^{\circ})}=1+o_P(1).$$

Set

$$\psi(u)=\frac{(au)^q+(au)^{-(1-q)}}{a^q+a^{-(1-q)}}.$$

Let $\delta=\sigma^2/\gamma$ and choose $\hat{\delta}$ such that

$$\hat{\lambda}=\left(\frac{\alpha l_2\hat{\delta}}{n4m}\right)^{2m/(1+4m)}$$

Then $g(\hat{\lambda})/g(\lambda^{\circ})=\psi(\hat{\delta}/\delta)$ where $q=4m/(1+4m)$, $a=1/4m$. Thus, $\psi(\hat{\delta}/\delta)=1+o_P(1)$. Note that $\psi(u)=1$ if and only if $u=1$. Because $\psi(u)$ is continuous at 1, $\hat{\delta}/\delta\to_P 1$; otherwise, a contradiction could be obtained. From the definition of $\hat{\delta}$, the consistency of $\hat{\lambda}$ now follows. □

LEMMA 3.3.

$$\hat{S}^2 - \frac{(1/n)\,\|(I - A(\lambda^\circ))\mathbf{Y}\|^2}{1 - \mathscr{C} m_1(\lambda^\circ)} = o_P\big(ET_n(\lambda^\circ)\big). \tag{3.7}$$

PROOF. The lhs of (3.7) can be rewritten as

$$V(\hat{\lambda})\varphi(\hat{\lambda}) - V(\lambda^\circ)\varphi(\lambda^\circ) \tag{3.8}$$

with $\varphi(\lambda) = (1 - m_1(\lambda))^2/(1 - \mathscr{C} m_1(\lambda))$. Expanding (3.8) using the identity

$$(ab - cd) = \tfrac{1}{2}(a - c)(b + d) + \tfrac{1}{2}(a + c)(b - d)$$

yields

(3.9)

$$\big(V(\hat{\lambda}) - V(\lambda^\circ)\big)\big(\varphi(\hat{\lambda}) + \varphi(\lambda^\circ)\big)/2 + \big(V(\hat{\lambda}) + V(\lambda^\circ)\big)\big(\varphi(\hat{\lambda}) - \varphi(\lambda^\circ)\big)/2.$$

Now from Lemma 3.1,

$$\begin{aligned}\big(V(\hat{\lambda}) - V(\lambda^\circ)\big) &= \big(T_n(\hat{\lambda}) - T_n(\lambda^\circ)\big)(1 + o_P(1)) \\ &= o_P\big(ET_n(\lambda^\circ)\big).\end{aligned}$$

It is clear that $\varphi(\lambda)$ is bounded and therefore the first term in (3.9) converges to zero at the necessary rate.

Now the second term will be dealt with. Note that $m_1(\lambda_n) \to 0$ as $n \to \infty$ and so by expanding the denominator of $\varphi(\lambda)$ in a power series one obtains $\varphi(\lambda) = 1 + (2 - \mathscr{C})m_1(\lambda) + o(m_1(\lambda))$ as $n \to \infty$ and $\lambda \in [\lambda_n, \infty)$. Thus,

$$\begin{aligned}\varphi(\hat{\lambda}) - \varphi(\lambda^\circ) &= (2 - \mathscr{C})\big(m_1(\hat{\lambda}) - m_1(\lambda^\circ)\big) \\ &\quad + o\big(m_1(\hat{\lambda})\big) + o\big(m_1(\lambda^\circ)\big) \quad \text{as } n \to \infty.\end{aligned}$$

Combining the results of (3.1) and Lemma 3.2, it follows that $m_1(\hat{\lambda}) - m_1(\lambda^\circ)$ is $o_P(\mu_1(\lambda^\circ))$. Thus from the relationship in (3.6), $\varphi(\hat{\lambda}) - \varphi(\lambda^\circ) = o_P(ET_n(\lambda^\circ))$. Clearly $V(\hat{\lambda}) + V(\lambda^\circ)$ will be bounded in probability and so the second term in (3.9) is $o_P(ET_n(\lambda^\circ))$ and the lemma now follows. □

LEMMA 3.4.

$$\frac{1}{n}\mathbf{e}_n' A(\lambda^\circ)^k \mathbf{e}_n = \sigma^2 \mu_k(\lambda^\circ)(1 + o_P(1)), \qquad k = 1, 2 \text{ as } n \to \infty.$$

PROOF. Let $X_n = \mathbf{e}_n' A(\lambda^\circ)^k \mathbf{e}_n / (\sigma^2 n m_k(\lambda^\circ))$. From the convergence in (3.1) it is sufficient to show that $X_n \to_P 1$ as $n \to \infty$. By definition $E(X_n) = 1$ and it will be argued that $\mathrm{Var}(X_n) \to 0$ as $n \to \infty$. Since $e_{1n}, \dots, e_{nn}$ are assumed to be i.i.d., if $\nu = E(e_{kn}^4)$, then

$$\mathrm{Var}\big(\mathbf{e}_n' A(\lambda^\circ)^k \mathbf{e}_n\big) = 2\sigma^4 \operatorname{tr} A(\lambda^\circ)^k + (\nu - 3\sigma^4)\, \|\mathrm{diag}\big(A(\lambda^\circ)^{2k}\big)\|^2.$$

Now

$$\|\mathrm{diag}\big(A(\lambda^\circ)^{2k}\big)\|^2 \le \mathrm{tr}\big(A(\lambda^\circ)^{4k}\big) \le \mathrm{tr}\big(A(\lambda^\circ)^k\big),$$

where the last inequality follows from the fact that the eigenvalues of $A(\lambda^\circ)$ are contained in $(0, 1]$. From these expressions it follows that $\mathrm{Var}(X_n) = o(1/(nm_k(\lambda^\circ)))$, and thus from (3.1) and (3.4), $\mathrm{Var}(X_n) = O(n^{-1/(1+4m)})$. □

For the next lemma, it is convenient to introduce the notation $\mathbf{f}' = (f(t_1), \dots, f(t_n))$.

LEMMA 3.5.

$$\frac{1}{n}\mathbf{f}'(I - A(\lambda^\circ))^2\mathbf{e}_n = o_P(ET_n(\lambda^\circ)).$$

PROOF. Let $X_n = (1/n)\mathbf{f}'(I - A(\lambda^\circ))^2\mathbf{e}_n/\sigma$. Note that $EX_n = 0$ and

$$\mathrm{Var}\, X_n = \frac{1}{n^2}\|(I - A(\lambda^\circ))^2\mathbf{f}\|^2 \le \frac{1}{n^2}\|(I - A(\lambda^\circ))\mathbf{f}\|^2 \le \frac{1}{n}ET_n(\lambda^\circ),$$

where the first inequality follows because the eigenvalues of $A(\lambda^\circ)$ lie in $(0, 1]$. Now in using the rates given in (3.2) and (3.4),

$$\mathrm{Var}(X_n/ET_n(\lambda^\circ)) \le (nET_n(\lambda^\circ))^{-1} = O(n^{1/(4m+1)})$$

and thus X_n converges in probability to zero at the appropriate rate. □

LEMMA 3.6.

$$\Gamma = \frac{1}{n}\|(I - A(\lambda^\circ))\mathbf{f}\|^2 - \sigma^2(2 - \mathscr{C})\mu_1(\lambda^\circ) + \sigma^2\mu_2(\lambda^\circ) \tag{3.10}$$
$$= o(ET_n(\lambda^\circ)) \quad \text{as } n \to \infty.$$

PROOF. Adding and subtracting $\sigma^2 m_2(\lambda^\circ)$ from Γ we have

$$\begin{aligned}\frac{\Gamma}{ET_n(\lambda^\circ)} &= 1 - \frac{\sigma^2(2 - \mathscr{C})\mu_1(\lambda^\circ)}{ET_n(\lambda^\circ)} - \frac{\sigma^2(\mu_2(\lambda^\circ) - m_2(\lambda^\circ))}{ET_n(\lambda^\circ)} \\ &= 1 - \frac{\sigma^2(2 - \mathscr{C})\mu_1(\lambda^\circ)}{\gamma(\lambda^\circ)^2 + \sigma^2\mu_2(\lambda^\circ)} - o(1) \quad \text{as } n \to \infty,\end{aligned} \tag{3.11}$$

by the asymptotic equivalences given in Lemma 3.1.

Substituting the asymptotic form for λ° into (3.11) yields

$$\frac{\Gamma}{ET_n(\lambda^\circ)} = 1 - \frac{(2 - \mathscr{C})l_1}{l_2(1/4m + 1)} - o(1). \tag{3.12}$$

Referring to Selby [(1979), page 461] $l_1/l_2 = 2m/(2m - 1)$. Also recall that

$$(2 - \mathscr{C}) = \frac{1}{K} = \left(\frac{2m - 1}{2m}\right)\left(\frac{4m + 1}{4m}\right)$$

and by this choice for $\mathscr{C}$, it is straightforward to verify that the middle term on the rhs of (3.12) above equals 1. □

4. Proof of Theorem 2.1. To further streamline notation, let $\mu_k = \mu_k(\lambda^\circ)$ and $m_k = m_k(\lambda^\circ)$ for $k = 1, 2$ and $A = A(\lambda^\circ)$ and $\mathbf{f}' = (f(t_{1n}), \dots, f(t_{nn}))$:

$$(4.1) \qquad \hat{S}_n^2 - S_n^2 = \left[\hat{S}_n^2 - \frac{(1/n)\,\|(I-A)Y\|^2}{1 - \mathscr{C}m_1}\right] + \left[\frac{(1/n)\,\|(I-A)Y\|^2 - (1-\mathscr{C}m_1)S_n^2}{1-\mathscr{C}m_1}\right].$$

The first term in (4.1) converges in probability at the necessary rate by Lemma 3.3. Recall that $\mathbf{Y} = \mathbf{f} + \mathbf{e}$ and the second term can be expanded by adding and subtracting $-\mathscr{C}\mu_1\sigma^2 + 2\mu_1\sigma^2 - \mu_2\sigma^2$ in the numerator to give

$$(4.2) \qquad \frac{R_1 + R_2 + R_3 + \Gamma}{1 - \mathscr{C}m_1}$$

where

$$R_1 = \mathscr{C}\left(m_1 S_n^2 - \mu_1\sigma^2\right),$$

$$R_2 = -2\left(\frac{1}{n}\mathbf{e}'A\mathbf{e} - \mu_1\sigma^2\right) + \left(\frac{1}{n}\mathbf{e}'A^2\mathbf{e} - \mu_2\sigma^2\right),$$

$$R_3 = \frac{1}{n}\mathbf{f}'(I-A)^2\mathbf{e}$$

and Γ is defined in the statement of Lemma 3.6.

Because $1 - \mathscr{C}m_1 \to 1$, to complete the proof it is sufficient to show that each term in the numerator of (4.2) is $o_P(ET_n(\lambda^\circ))$. Moreover, in view of the asymptotic relations between $ET_n(\lambda^\circ)$ and $\mu_k(\lambda^\circ)$ in Lemma 3.1, it is enough to establish a convergence rate to zero of the form $o_P(\mu_1 \vee \mu_2)$ for each term.

The convergence of R_1 follows from the equivalence in (3.1) and the fact that $S_n^2 \to_P \sigma^2$. R_2 is $o_P(\mu_1 \vee \mu_2)$ by Lemma 3.4 while the convergence of R_3 follows from Lemma 3.5. Finally, note that $\Gamma = o_P(ET_n(\lambda^\circ))$ by Lemma 3.6 and thus, having considered all the terms in (4.2), (2.1a) follows.

The second part of the theorem is a consequence of (2.1a) and (1.8). For (2.1c) we refer to the discussion in the introduction and note that from (3.1) and Lemma 3.2, $\operatorname{tr} A(\hat{\lambda})/n = \mu_1(\lambda^\circ)(1 + o_P(1))$. Thus $\operatorname{tr} A(\hat{\lambda})/n \to_P 0$ as $n \to \infty$ and the bracketed term in (1.10) converges to 1. □

REFERENCES

BUCKLEY, M. J., EAGLESON, G. W. and SILVERMAN, B. W. (1988). The estimation of residual variance in nonparametric regression. *Biometrika* **75** 189–200.

COX, D. D. (1984). Gaussian approximation of smoothing splines. Technical Report 743, Dept. Statist. University of Wisconsin/Madison.

COX, D. D. (1988). Approximation of method of regularization estimators. *Ann. Statist.* **16** 694–712.

HALL, P. and TITTERINGTON, D. M. (1987). Common structure of techniques for choosing smoothing parameters in regression problems. *J. Roy. Statist. Soc. Ser. B* **49** 184–198.

LI, K. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.* **14** 1101–1112.

MESSER, K. (1989). A comparison of a spline estimate to its equivalent kernel. Unpublished manuscript.

NAIMARK, M. A. (1967). *Linear Differential Operators*, Part I. Ungar, New York.

NYCHKA, D. W. (1988). Bayesian confidence intervals for a smoothing spline. *J. Amer. Statist. Assoc.* **83** 1134–1143.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

SELBY, S., ED. (1979). *CRC Standard Mathematical Tables*. Central Rubber Company, Cleveland, Ohio.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

SPECKMAN, P. (1983). Efficient nonparametric regression with cross-validated smoothing splines. Unpublished.

WAHBA, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14** 661–667.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45** 133–150.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
CAMPUS BOX 8203
RALEIGH, NORTH CAROLINA 27695-8203