Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://amstat.tandfonline.com/loi/uasa20

Choosing a Range for the Amount of Smoothing in Nonparametric Regression

Douglas Nychka

To cite this article: Douglas Nychka (1991) Choosing a Range for the Amount of Smoothing in Nonparametric Regression, Journal of the American Statistical Association, 86:415, 653-664

To link to this article: http://dx.doi.org/10.1080/01621459.1991.10475091

Choosing a Range for the Amount of Smoothing in Nonparametric Regression

DOUGLAS NYCHKA*

Most nonparametric regression estimates smooth the observed data in the process of generating a suitable estimate. The resulting curve estimate is sensitive to the amount of smoothing introduced, and it is important to examine the estimate over a range of smoothing choices. An objective way to specify such a range is to estimate the optimal value of the smoothing parameter (bandwidth) with respect to some loss function and then report a confidence interval for this parameter estimate. This interval suggests examining three regression estimates: an undersmoothed curve evaluated at the lower endpoint of the confidence interval, an oversmoothed curve evaluated at the upper endpoint, and also an optimal curve chosen by the data-based method. This article describes two strategies for constructing such confidence intervals using asymptotic approximations and simulation techniques. Suppose that data are observed from the model $Y_k = f(t_k) + e_k$ $(1 \le k \le n)$, where f is a function that is twice differentiable and $\{e_k\}$ are mean-zero, independent, random variables. A cubic smoothing spline estimate of f is considered where the smoothing parameter is chosen using generalized cross-validation. Confidence intervals are constructed for the smoothing parameter that minimizes average squared error. This is done using the asymptotic distribution of the cross-validation function and by a version of the bootstrap. Although this bootstrap method involves more computation, it yields confidence intervals that tend to have a shorter width. In general, this second method is easier to implement since it avoids the problem of deriving the complicated formulas for the asymptotic variance. Moreover, for spline estimates one can significantly reduce the amount of computation through a useful orthogonal decomposition. Although this work focuses on smoothing spline estimates where the smoothing parameter is found by generalized cross-validation, the basic ideas in this article are not limited to this pair of methods. Regardless of how the smoothing parameter is selected, it is useful to consider under- and oversmoothed curve estimates to investigate the sensitivity. A confidence interval for the smoothing parameter remains a useful way for specifying this range.

KEY WORDS: Bandwidth selection; Bootstrap; Confidence intervals; Cross-validation; Smoothing splines.

1. INTRODUCTION AND OVERVIEW

In many situations the goal of a statistical analysis is to recover a smooth function from noisy observations. One useful model is

$$Y_k = f(t_k) + e_k, \qquad 1 \le k \le n, \tag{1.1}$$

where f is a function that is twice differentiable and $\{e_k\}$ are mean-zero, independent, random variables with variance σ^2. Often a suitable parametric form for the function is not known, and for this reason it is important to study nonparametric techniques for estimating a function.

Most nonparametric curve estimates effect some form of smoothing of the data in the process of generating an estimate. The amount of smoothness is usually indexed by a parameter, and it is well known that the resulting estimate is sensitive to the value of this parameter. Figure 1 illustrates this sensitivity for a simulated data set and a cubic smoothing spline estimate [see Sec. 2, (2.1)]. The smoothing parameter will be denoted by λ and, to make the dependence on the smoothing parameter explicit, the curve estimate will be represented by $\hat{f}_\lambda$. Figure 2 is a perspective representation of the surface described by $\hat{f}_\lambda$ as a function of t and also of $\log \lambda$. As λ ranges from 0 to ∞, the estimate changes from a rapidly oscillating function that interpolates the data to a linear function. Cubic smoothing splines have the property that $\hat{f}_\infty$ will just be the line estimated by ordinary least squares, while the spline at $\lambda = 0$ will pass through all the data points.

One useful way of quantifying the accuracy of a nonparametric curve estimate is by the average squared error (ASE):

$$\mathrm{ASE}(\lambda) = \frac{1}{n}\sum_{k=1}^{n} [\hat{f}_\lambda(t_k) - f(t_k)]^2.$$

Figure 3 shows how the average squared error depends on λ, and the lettered points on this curve correspond to the different estimates plotted in Figure 1. Note that the point where the ASE is minimized corresponds to a visually pleasing estimate of f. Moreover, qualitatively the logarithm of the smoothing parameter seems to be a more sensible scale than the parameter itself. [This scale is suggested by the results in Silverman (1984).] It is interesting to relate the size of the loss function to the visual impression of the estimate. When $\log \lambda = -18.5$ the ASE has only been doubled with respect to the minimum, but $\hat{f}_\lambda$ has many more modes than the true curve (or the optimal spline estimate).

At first glance it may seem that nothing has been gained by considering a nonparametric approach because the choice of λ has the same subjectivity as the choice of a particular parametric model. Although one could use a loss function, such as the ASE, for choosing a value for λ, such a criterion will depend on f. Using methods based on unbiased risk estimates, cross-validation, or biased cross-validation, however, it is possible to determine the appropriate amount of smoothing objectively (Craven and Wahba 1979; Härdle, Hall, and Marron 1988; Li 1986; Scott 1988). These procedures do not depend on knowledge of the unknown func-

* Douglas Nychka is Associate Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203. This work was generously supported by National Science Foundation Grant DMS-8715756 and by the University of Bath, Bath, UK. The author thanks Jeff Hart, John Rice, and Steve Marron for their encouragement and insights and James Ward of GLAXO for making the rat-diet data available.

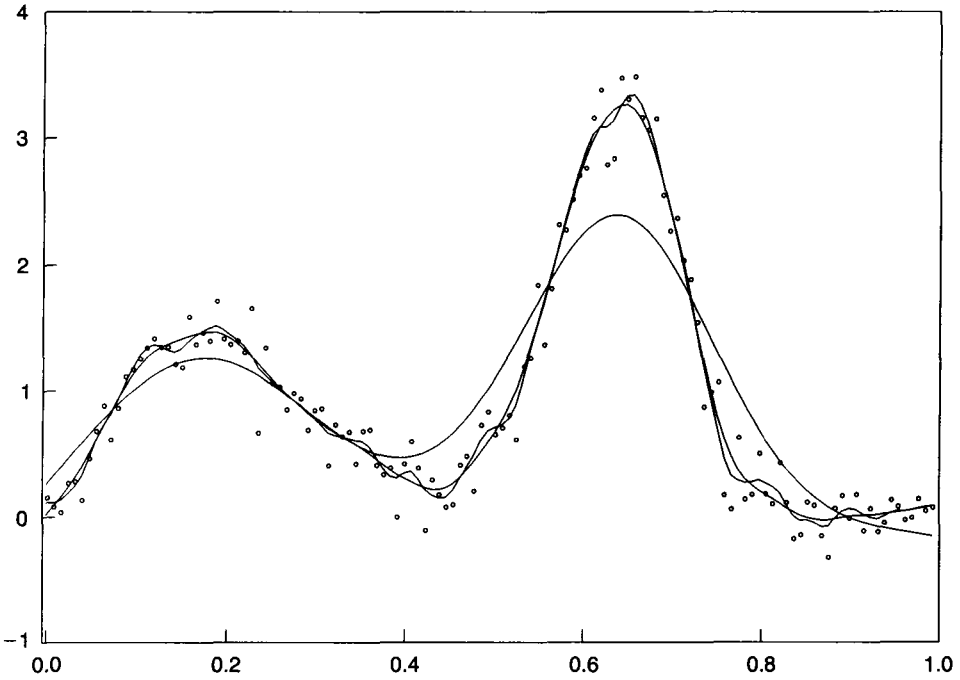


Figure 1. Simulated Data Illustrating the Dependence on the Curve Estimate to the Smoothing Parameter. Data have been simulated according to the additive model (1.1) with normal errors, $\sigma = .2$, $n = 128$, and f is function type II. The estimates are cubic smoothing splines with $\ln(\lambda) = (-18.5, -15.7, -11.1)$.

tion and have good asymptotic properties. In the limit they will minimize the average squared error of the estimate.

Since the curve estimate is sensitive to the value of λ, it is important to investigate how features in the curve depend on the choice of the smoothing parameter. In a nonparametric analysis it is tempting to determine the amount of smoothness using a data-based method and then just report the curve estimate evaluated at the estimated smoothing pa-

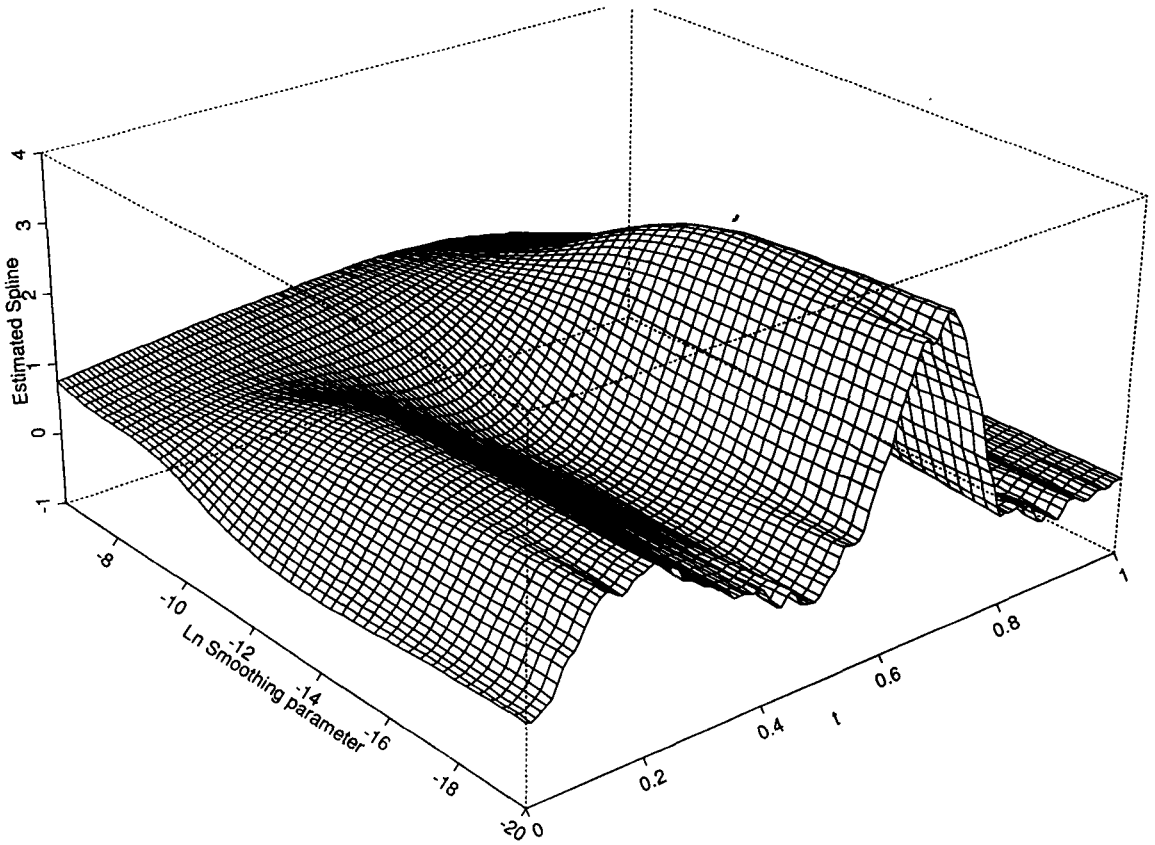


Figure 2. Perspective Surface Illustrating the Spline Estimate as a Function of t and the ln Smoothing Parameter. The simulated data are the same as plotted in Figure 1.

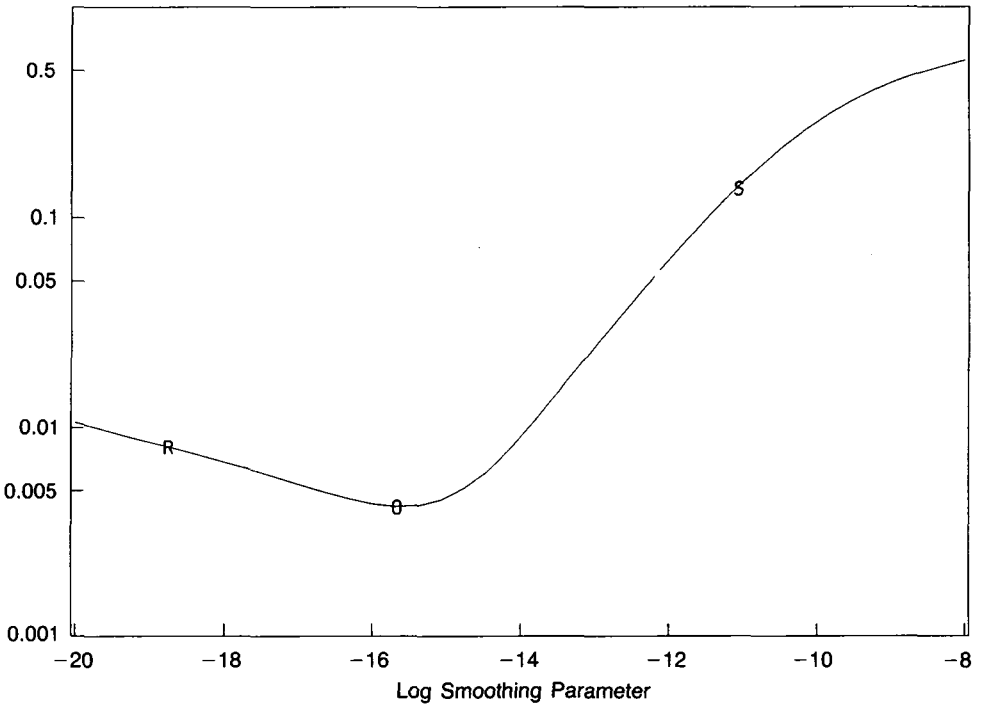


Figure 3. The Average Squared Error for a Spline Estimate as a Function of the ln Smoothing Parameter. Plotted points correspond to the rough, optimal, and oversmooth estimates in Figure 1. For this data set, the smoothing parameter found by generalized cross-validation is similar to the optimal value ($\hat{\lambda}_0$) using an average squared error loss function. The rough and oversmoothed estimates are actually based on the smoothing parameters from the endpoints of the simulation-based 95% confidence interval.

rameter. This is not good practice. One reason is the large amount of variability in the estimate of the smoothing parameter. Although one could examine curve estimates over a range of smoothing parameters, it is not obvious what range to choose. What is missing is some guidance in judging how much variability one can expect in the estimate of the smoothing parameter. A confidence interval for the estimated smoothing parameter is one way of specifying a reasonable range. This interval suggests reporting three estimates: an undersmoothed curve corresponding to the lower endpoint of the confidence interval, an oversmoothed curve corresponding to the upper endpoint, and also the curve chosen by the data-based method.

This article studies two different strategies for constructing confidence intervals for λ. One draws on the asymptotic theory that has been developed in Härdle et al. (1988), while the other uses a parametric version of bootstrapping. Although this second method involves more computation, it yields confidence intervals that tend to be narrower. In general, this second method is easier to implement since it avoids the problem of using the complicated formulas for the asymptotic variance. Although this work focuses on smoothing spline estimates when $\hat{\lambda}$ is found by generalized cross-validation, the basic ideas in this article are not limited to this pair of methods. Any nonparametric regression method will be sensitive to the choice of smoothing parameter. Regardless of how the smoothing parameter is selected, it is useful to consider both undersmoothed and oversmoothed curve estimates to investigate the sensitivity. A confidence interval for the smoothing parameter remains a useful way for specifying this range.

One possible criticism of this work is that there is nothing new: Data analysts always construct several curve estimates using a range of smoothing parameter values and have been doing so for years. One problem is that simple rules to specify a range for λ may fail to adapt to the sample size, to a nonuniform distribution of $\{t_k\}$, or to the shape of f. As an alternative to simple rules, one might develop a feel for the sensitivity through an interactive graphical analysis; however, besides being time consuming, the final results remain subjective. Certainly if one has the luxury to concentrate on a single regression problem, then an intensive graphical analysis is justified. In some practical situations, however, there are many separate data sets and thus many regression estimates to interpret. In this context it is reasonable to summarize the sensitivity of the curve estimates to the smoothing parameter in a concise and objective manner. While a thorough, graphical exploration of a data set is always recommended, statisticians should also be concerned about reliable techniques for data reduction when time is limited. Quantifying the range for smoothing using a confidence interval is one useful form of data reduction that has a clear interpretation. Moreover, at least in the case of smoothing splines, determining this range takes little more work than computing the actual estimate (see Sec. 5).

Before giving some background on this problem and describing the organization of this article, it is important to clarify the interpretation of a confidence interval for λ.

The estimated smoothing parameter and its corresponding confidence interval have no importance on their own. It is the changing shape of the estimated regression curve across this range that is important. Although most of this

article must necessarily focus on the details of smoothing parameter estimates it is always done with the ultimate intention of examining the curve estimates at the smoothing parameter values.

The range of regression curves based on a confidence interval for λ may be confused with a confidence band on the curve estimate itself. These two summaries of variability have very different interpretations. Let $\hat{\lambda}_0$ denote the minimizer of ASE(λ). If $\hat{f}$ denotes the nonparametric estimate of f, where λ has been estimated from the data and $\hat{f}_0$ is the optimal estimate using $\lambda = \hat{\lambda}_0$, then it is useful to consider the separation

$$\hat{f}(t) - f(t) = [\hat{f}(t) - \hat{f}_0(t)] + [\hat{f}_0(t) - f(t)].$$

Quantifying the variability of the first term is the subject of this article. An important limitation of a confidence interval for $\hat{\lambda}_0$ is that it only provides information about the deviation of the estimated curve from an estimate that employs the best value of the smoothing parameter. In contrast, pointwise or simultaneous confidence intervals for $\hat{f}(t)$ are based on the second term. To make inferences about f, one must also quantify the variability of the optimal estimate relative to the true function. Because $\hat{\lambda}$ is a consistent estimate of $\hat{\lambda}_0$, the contribution of the first term to the distribution of $\hat{f} - f$ is asymptotically negligible as $n \to \infty$ (Nychka 1988). What is important is the variability of the optimal estimate relative to the unknown function.

Figure 4 illustrates these two different ways of making an inference about f for some experimental data. The goal of this experiment was to study the effect of an appetite suppressant in a specific strain of rats, and the reported results are the daily median food intake for treatment and control groups at 39 different days. The appetite suppressant was removed from the treatment group diet after 63 days. Figure 4a plots the regression curves evaluated at an estimate of the optimal smoothing parameter and at the endpoints of a 90% confidence interval for $\hat{\lambda}_0$. The wide range in smoothness between the undersmoothed and oversmoothed curves is a sobering reminder of the large variability associated with smoothing parameter estimates and is partly due to the small number of data points in this example. The shaded regions in Figure 4b are approximate 90% confidence bands for f (see Sec. 2 for their description). Although the confidence limits for $\hat{\lambda}_0$ are more useful for assessing the *shape* of $\hat{f}_0$ (or f), the confidence band helps to set a *range* for f. Also, the confidence bands are constructed under the assumption that $\hat{\lambda}$ is close to the optimal value although the confidence interval for $\hat{\lambda}_0$ obviously takes this variability into account.

Besides the practical application of these intervals for data analysis, there are several aspects of this problem that are of more theoretical interest.

The work of Härdle et al. (1988) gives conditions under which a smoothing parameter estimated using cross-validation or other related methods will have a limiting normal distribution. One question posed by these researchers' work is whether this asymptotic distribution theory approximates situations where the sample size is in the range of a few hundred points. Moreover, the limiting distribution of the estimated smoothing parameter depends on the average squared bias of the unknown function and the error variance, σ. It is not clear how useful this asymptotic theory will be when these quantities are replaced by estimates. The asymptotic variance for $\hat{\lambda}$ has a complicated form, and to insure the accuracy of the results, the formula was computed symbolically using MACSYMA (Symbolics Inc. 1984).

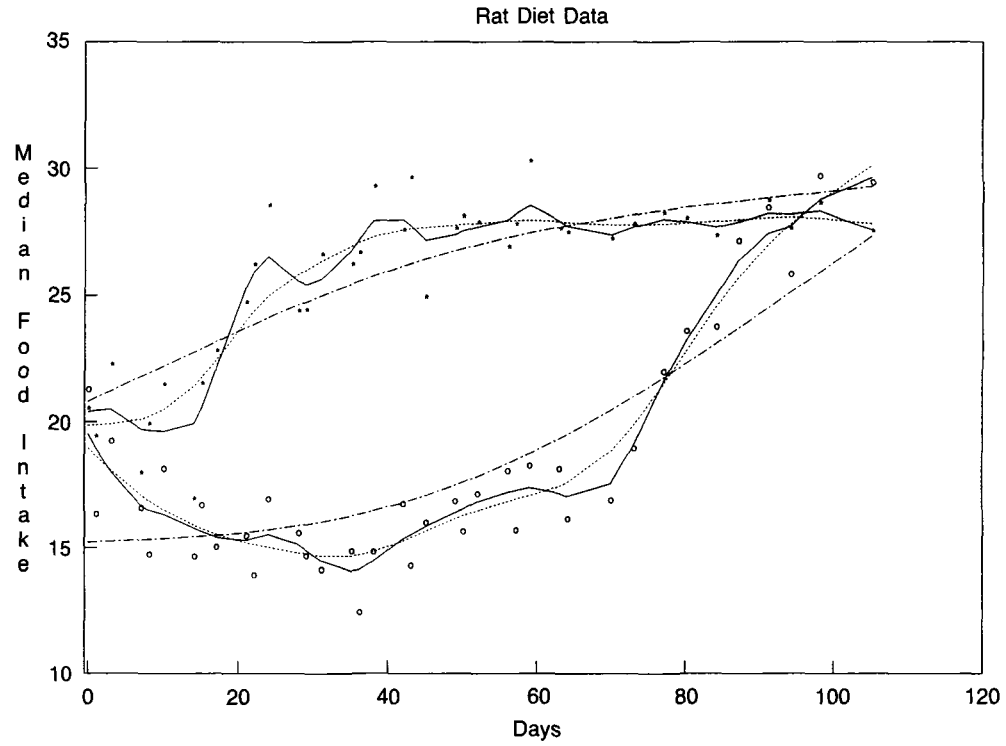
Recently Chiu (1990) has identified a much more accurate approximation to the asymptotic distribution for the estimated smoothing parameter using Mallow's criterion. Although the approximating distribution has a simple form, it is with respect to the optimal smoothing parameter for *expected* ASE. Chiu's analysis applied to cross-validation and centered about $\hat{\lambda}_0$ would yield a hybrid method that is intermediate between the asymptotic and simulation-based methods discussed here.

The version of parametric bootstrapping used in this article has some interesting differences from other applications of the bootstrap to nonparametric regression estimates. Bootstrapping has been applied to kernel estimates to improve the mean squared-error properties of the estimates and also has been suggested as a method for constructing confidence intervals for the estimated curve (Bowman and Härdle 1988). These researchers used one kernel estimate to find f and another, higher order kernel to compute the residuals used for resampling. In contrast, this article generates bootstrap replications from the model:

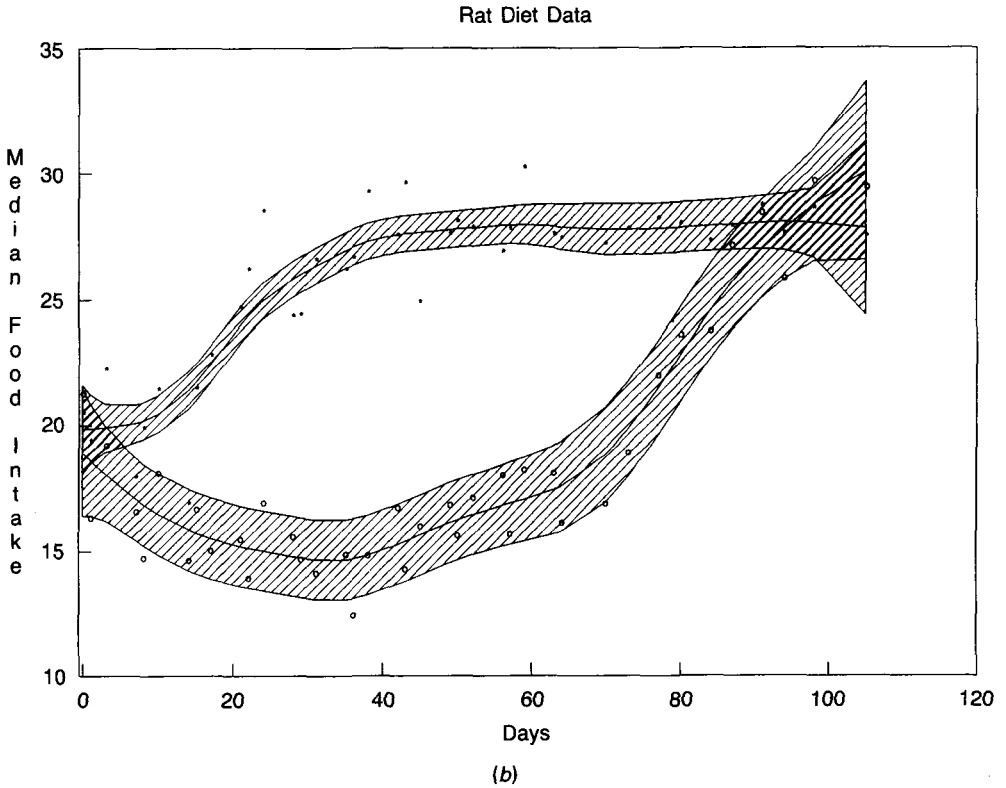
$$Y_k = \hat{f}(t_k) + \epsilon_k,$$

where the errors follow a normal distribution. This assumption on the error distribution may be justified from the asymptotic theory and may mean that this method is not robust to departures from normality for small sample sizes. One clear advantage of this approach is that it avoids the higher order kernel estimate (and the choice of a second bandwidth) to compute residuals.

Recent research has identified some problems with the method of cross-validation for determining the smoothing parameter. From Terrell and Scott (1986) and Härdle et al. (1988) it is known that the estimate of λ from cross-validation is negatively correlated with $\hat{\lambda}_0$. Let λ_0 denote the minimizer of the expected ASE. Cross-validation will tend to slightly undersmooth the data when $\hat{\lambda}_0$ is greater than λ_0 and to oversmooth when $\hat{\lambda}_0$ is less than λ_0. A heuristic explanation for this effect may be traced to the cross-validation function's sensitivity to autocorrelation in the errors. For example, spurious positive autocorrelation can occasionally yield smoothing parameter estimates that drastically undersmooth the data (see Diggle and Hutchinson 1989; Hart and Wehrly 1986). Some promising alternatives to cross-validation are biased cross-validation (Scott 1988) and more sophisticated "plug in" methods such as smoothed cross-validation (Hall, Marron, and Park 1989). With any of these estimates of λ it is still appropriate to use a confidence interval to quantify an acceptable range for the amount of smoothing, and the techniques discussed in this article would readily apply to these other methods. Because the confidence procedures work well for the highly variable esti-



(a)



(b)

Figure 4. Experimental Data Comparing the Effect Over Time of an Appetite Suppressant Drug on Two Groups of Rats. Each group consisted of nine animals, and the response is the median food intake measured on a daily basis (= control group, ○ = treatment group). The suppressant was removed from the treatment group's diet after 63 days. (a) The dotted lines are the spline estimates where the smoothing parameter has been found using generalized cross-validation. The solid (dashed) lines are undersmoothed (oversmoothed) estimates derived from 90% confidence intervals for the optimal smoothing parameter. The wide range of the estimates indicates the variability of cross-validation in small data sets. (b) Approximate 90% simultaneous confidence bands for median food intake based on a cubic smoothing spline where the smoothing parameter is found by generalized cross-validation. The construction of this band is described in Section 2.*

mates supplied by cross-validation, they are probably effective for more accurate methods of estimating λ.

This introduction ends with an outline of the rest of this article. Section 2 defines a smoothing spline estimate for the model (1.1) and motivates the use of generalized cross-validation to estimate the smoothing parameter. Section 2 also presents a smoothing spline as an orthogonal series estimate because that representation is useful for generating bootstrap replications efficiently. Section 3 reviews the asymptotic theory for $\hat{\lambda}$ in the context of spline estimates and describes the construction of confidence intervals based on this theory. Section 4 describes the confidence interval based on the bootstrap and discusses its computation. Section 5 reports the results of a simulation study. These results cover the three smooth test functions used by Wahba (1983), three levels of σ (.2, .4, .6), two sample sizes (n = 128, 256), and two error distributions (normal, T_6). Except for two cases, both types of confidence intervals hold their level across the two different error distributions. Moreover, the median width of these intervals (on a log scale) is comparable to the width of an interval obtained when the distribution of $\log(\hat{\lambda}_0) - \log(\hat{\lambda})$ is known. One significant difference is that the distribution of widths for the intervals based on the asymptotic theory has a much longer right tail (see Figure 5 for a quick graphical summary of these results). Section 6 states some theoretical results to support the simulation-based confidence intervals.

2. DEFINITION AND PROPERTIES OF A SMOOTHING SPLINE ESTIMATE

A spline is best defined as a solution to a minimization problem. Let $\mathcal{H} = W_2^2[0, 1] = \{h: h, h' \text{ absolutely continuous}, h'' \in L^2[0, 1]\}$. For $\lambda > 0$ and $h \in \mathcal{H}$ let

$$\mathcal{L}(h) = \frac{1}{n}\sum_{k=1}^{n} [Y_k - h(t_k)]^2 + \lambda \int_0^1 h''(t)^2\, dt. \quad (2.1)$$

A cubic smoothing spline (with natural boundary conditions) for the additive model (1.1) is defined as the func-

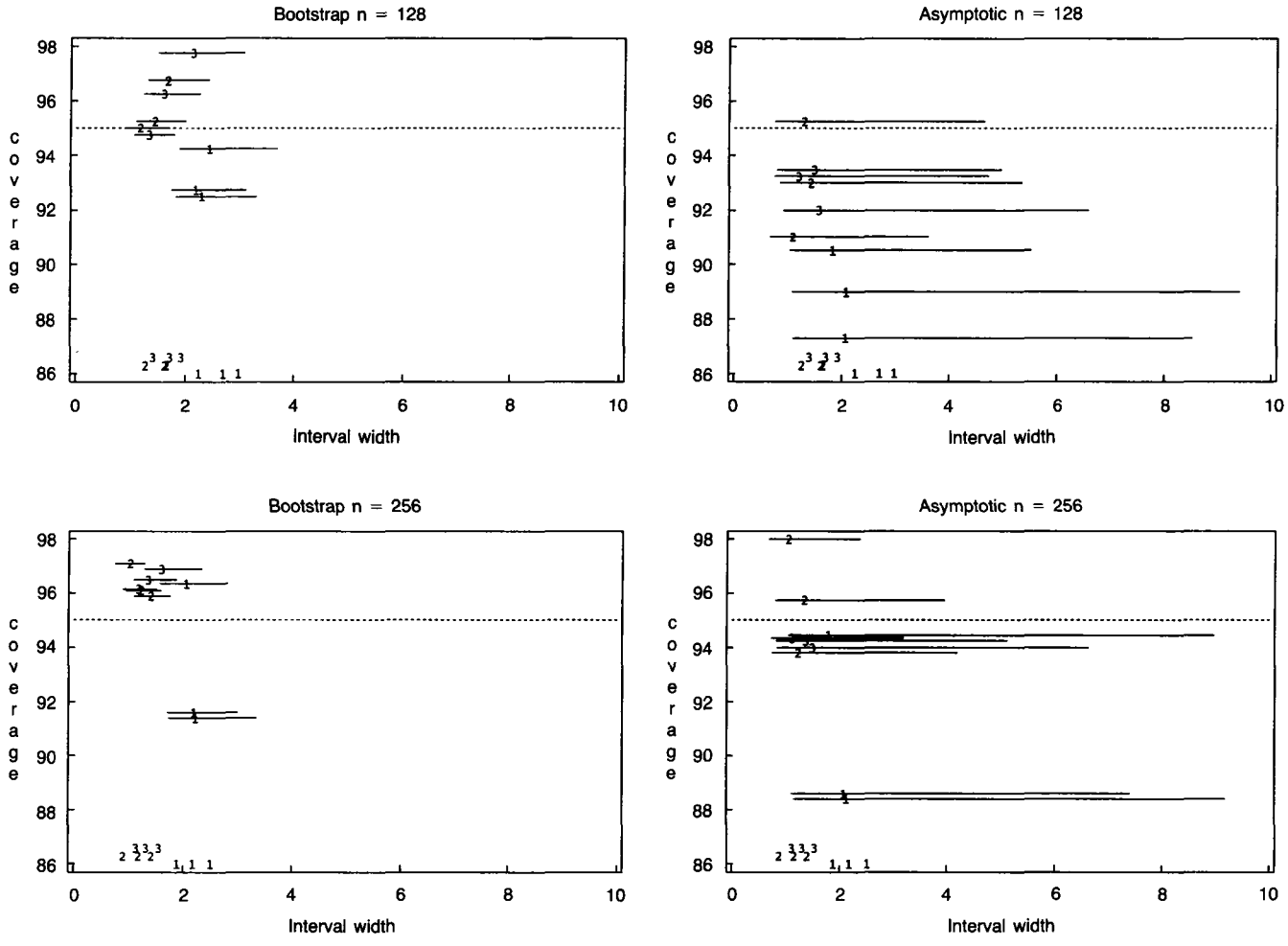


Figure 5. Simulation Results Studying Two Methods of Constructing 95% Confidence Intervals for the Smoothing Parameter. Values of the smoothing parameter are given in $\log_{10}$ scale. This panel of four plots reports the performance of a simulation-based method and one based on asymptotic normality of the log smoothing parameter for normal errors. Two sample sizes are considered (128, 256), and the nine cases in each plot correspond to three different levels of σ (.2, .4, .6) and three types of test functions. The level of each segment indicates the observed coverage probability. The endpoints of the line segments are the 2.5% and 97.5% percentiles of the distribution of confidence interval widths for a given case and a numeral identifying the function type locates the median of the distribution. As a benchmark, the confidence interval widths when the distribution of the smoothing parameter is known are also included. These serve as a reference for the two methods being studied and are located near the x axis on each plot. (Note that the actual coverage for these reference intervals will be exactly 95%.)

tion that minimizes $\mathcal{L}$ for all $h \in \mathcal{H}$. In this particular case, the solution will take the form of a piecewise cubic polynomial with join points (knots) at $\{t_k\}$. The estimate will be denoted by $\hat{f}_\lambda$ to emphasize its dependence on the smoothing parameter. For fixed λ, $\hat{f}_\lambda$ will be a linear function of $\mathbf{Y}$ and one can define an $n \times n$ "hat" matrix $A(\lambda)$ such that $\hat{\mathbf{Y}} = \hat{\mathbf{f}}_\lambda = A(\lambda)\mathbf{Y}$, where $\hat{\mathbf{f}}_\lambda^T = (\hat{f}_\lambda(t_1), \ldots, \hat{f}_\lambda(t_n))$. The generalized cross-validation function has a convenient form in terms of this hat matrix (Craven and Wahba 1979; Wahba 1989) and has the form

$$V(\lambda) = \frac{1}{n}\sum_{k=1}^{n} \frac{(Y_k - \hat{f}_\lambda(t_k))^2}{\left[\frac{1}{n}\operatorname{tr}(I - A(\lambda))\right]^2} = \frac{\frac{1}{n}\left\|(I - A(\lambda))\mathbf{Y}\right\|^2}{\left[\frac{1}{n}\operatorname{tr}(I - A(\lambda))\right]^2}.$$

We will take $\hat{\lambda}$ to be the minimizer of $V(\lambda)$. The reader is referred to Bates, Lindstrom, Wahba, and Yandell (1986) for a general description of the computation of $\hat{f}_\lambda$ and $V(\lambda)$.

Pointwise confidence intervals for $\hat{f}$ were proposed by Wahba (1983) based on a Bayesian interpretation of spline estimates. They have the form

$$\hat{f}(t_k) + Z_{\alpha/2}\,\hat{\sigma}\sqrt{A_{kk}(\hat{\lambda})} \qquad l \le k \le n \tag{2.2}$$

with $\hat{\sigma}^2 = \|(I - A(\hat{\lambda}))\mathbf{Y}\|^2/\operatorname{tr}(I - A(\hat{\lambda}))$. Although these intervals may not be reliable at specific points, their average coverage probability will be close to $1 - \alpha$ (Nychka 1988). An easy way of constructing n simultaneous intervals is to replace $Z_{\alpha/2}$ in (2.2) by $Z_{\alpha/2n}$. For the data in Figure 4, $n = 39$ and so $Z_{.025/39} = 3.02$. By the properties of these intervals, their average level will be close to $(1 - \alpha/n)$, and by Bonferroni's inequality we are guaranteed that the simultaneous confidence level is approximately $1 - \alpha$ or larger.

The last part of this section gives a representation of a smoothing spline in terms of an orthogonal series estimator. From the theory for splines (Eubank 1988) there is a set of functions $\{\phi_{\nu n}\}_{1\le\nu\le n} \subseteq \mathcal{H}$ with the following properties.

1. The solution to (2.1) is contained in the linear subspace spanned by $\{\phi_{\nu n}\}_{1\le\nu\le n}$.

2. $$\frac{1}{n}\sum_{k=1}^{n} \phi_{\nu n}(t_k)\phi_{\mu n}(t_k) = \delta_{\nu\mu}.$$

3. $$\int_{[0,1]} \phi''_{\nu n}(t)\phi''_{\mu n}(t)\,dt = \gamma_{\nu n}\delta_{\nu\mu}, \qquad 0 \le \gamma_{1n} \le \gamma_{2n} \cdots \le \gamma_{nn},$$

where $\delta_{\nu\mu}$ is the Kronecker delta.

Thus the spline estimate will have the explicit form $\hat{f}_\lambda(t) = \sum_{\nu=1}^{n} \alpha_\nu \phi_{\nu n}(t)$, for $\boldsymbol{\alpha} \in \mathbf{R}^n$, and because of the properties of this basis

$$\mathcal{L}\left(\sum_{\nu=1}^{n} \alpha_\nu\phi_{\nu n}\right) = \sum_{\nu=1}^{n} (y_\nu - \alpha_\nu)^2 + \lambda\sum_{\nu=1}^{n} \gamma_{\nu n}\alpha_\nu^2, \tag{2.3}$$

where $y_\nu = (1/n)\sum_{k=1}^{n} \phi_{\nu n}(t_k)Y_k$. Minimizing (2.3) over $\boldsymbol{\alpha}$ gives $\alpha_\nu = y_\nu/(1 + \lambda\gamma_{\nu n})$ and thus

$$\hat{f}_\lambda(t) = \sum_{\nu=1}^{n} \frac{y_\nu}{1 + \lambda\gamma_{\nu n}}\phi_{\nu n}(t).$$

The generalized cross-validation function and the ASE also have simple forms using this basis:

$$V(\lambda) = \frac{\sum_{j=1}^{n}\left[\frac{\lambda\gamma_{\nu n}y_\nu}{1 + \lambda\gamma_{\nu n}}\right]^2}{\left[\sum_{j=1}^{n}\frac{\lambda\gamma_{\nu n}}{1 + \lambda\gamma_{\nu n}}\right]^2} \tag{2.4}$$

$$\mathrm{ASE}(\lambda) = \sum_{j=1}^{n}\left[\frac{y_\nu}{1 + \lambda\gamma_{\nu n}} - f_\nu\right]^2, \tag{2.5}$$

where $f_\nu = (1/n)\sum_{k=1}^{n} f(t_k)\phi_{\nu n}(t_k)$. With this representation, once $\mathbf{y}$ and $\mathbf{f}$ have been found, the functions $V(\lambda)$ and $\mathrm{ASE}(\lambda)$ can be evaluated in $O(n)$ operations. This property makes it possible to minimize both of these functions rapidly.

Although this representation holds for many different variations on a spline estimate, it has a particularly appealing interpretation when $\mathcal{H}$ is restricted to the space of functions where h and h' are periodic and $t_k = k/n$. Under these conditions, if n is even, $\phi_1 = 1$, $\gamma_1 = 0$. For $1 \le \nu < n/2$, $\phi_{2\nu-1}(t) = \cos(2\pi\nu t)$, $\phi_{2\nu}(t) = \sin(2\pi\nu t)$ and $\gamma_{2\nu-1}$, $\gamma_\nu = 1/(2\pi\nu)^4$. The last term is $\phi_n(t) = \cos(\pi n t)$ and $\gamma_n = 1/(\pi n)^4$. The weight function, $1/(1 + \lambda\gamma_\nu)$, will be close to one for low-frequency components of the spectrum of $\mathbf{Y}$ and will down-weight high-frequency components. In this case a spline can be thought of as a low-pass filter where the frequency at the half-power point is approximately $\lambda^{1/4}/2\pi$. This periodic spline can be computed very efficiently by using the fast Fourier transform (FFT). One uses the FFT to calculate $\mathbf{y}$ and then applies the inverse FFT to the filtered spectrum to obtain $\hat{\mathbf{f}}_\lambda$.

3. CONFIDENCE INTERVALS FOR $\hat{\lambda}_0$ BASED ON ASYMPTOTIC THEORY

This section reviews the asymptotic theory for the distribution of the cross-validation function with the intent of constructing confidence intervals for the smoothing parameter. The theory in Härdle et al. (1988) deals specifically with positive kernel estimators, but analogous results may be derived for smoothing splines. A proof of these results is outlined in Nychka (1989a). The last part of this section gives the details for deriving the formula for the asymptotic variance and explains how the unknown quantities in this formula can be estimated.

To characterize the asymptotic distribution of a maximum likelihood estimator, one considers a Taylor series expansion of the score function. To study the distribution of a smoothing parameter, a similar approach is taken where the derivative of the cross-validation function plays the role of the score function. Accordingly, let $d_k(\lambda) = (\partial^k/\partial\lambda^k)V(\lambda)$, and let $I(\hat{\lambda}, \hat{\lambda}_0)$ denote the interval with endpoints at $\hat{\lambda}$ and $\hat{\lambda}_0$. By definition, $d_1(\hat{\lambda}) = 0$ and from the mean value theorem there is a $\lambda^* \in I(\hat{\lambda}, \hat{\lambda}_0)$ such that

$$-d_1(\hat{\lambda}_0) = d_2(\lambda^*)(\hat{\lambda} - \hat{\lambda}_0).$$

Recall that λ_0 is the minimizer of the expected average squared error ($\mathrm{EASE}(\lambda)$). It is conjectured that under Assumptions 1–4 in Section 6, the random variable $^9U_n(\lambda_0)$,

defined at (3.7), will have the following properties as $n \to \infty$:

$$d_1(\hat{\lambda}_0) - \mathcal{U}_n(\lambda_0) = o_p(\eta), \tag{3.1}$$

$$\mathcal{U}_n(\lambda_0)/\eta \xrightarrow{d} N(0, 1), \quad \eta^2 = \operatorname{var}(\mathcal{U}_n(\lambda_0)), \tag{3.2}$$

and

$$d_2(\lambda^*)/d_2(\hat{\lambda}) \xrightarrow{p} 1. \tag{3.3}$$

The limiting distribution of $\hat{\lambda}$

$$\frac{(\hat{\lambda} - \hat{\lambda}_0)}{\eta} d_2(\hat{\lambda}) \xrightarrow{d} N(0, 1) \tag{3.4}$$

is implied by relations (3.1)–(3.3). If η were known, (3.4) suggests the following $(1 - \alpha)100\%$ confidence interval for $\hat{\lambda}_0$

$$\hat{\lambda} \pm Z_{\alpha/2}\eta/d_2(\hat{\lambda}), \tag{3.5}$$

where Z_α is the $(1 - \alpha)100$ percentile of the standard normal distribution.

Although it is easy to differentiate $V(\lambda)$ with respect to λ, as mentioned in the introduction, $\log(\lambda)$ is a more suitable scale for quantifying the variability of the smoothing parameter. The asymptotic results mentioned previously can be translated to a log scale by applying the delta method. Let $\hat{\rho} = \ln(\hat{\lambda})$ and $\hat{\rho}_0 = \ln(\hat{\lambda}_0)$. A consistent estimate of $\hat{\rho}_0$ will be $\hat{\rho}$ (see, for example, Nychka 1990), and (3.4) suggests an interval for $\hat{\rho}_0$ of the form

$$\hat{\rho} \pm Z_{\alpha/2}\eta h_2(\hat{\rho})/\hat{\lambda}, \tag{3.6}$$

where $h_2(\rho) = (\partial^2/\partial\rho^2)V(e^\rho) = \lambda d_1(\lambda)^2 + d_2(\lambda)|_{\lambda = \ln(\rho)}$.

Readers familiar with Härdle et al. (1988) may be confused why the extensive asymptotic analysis by those authors does not resolve the conjectured relations in (3.1)–(3.3). The problem is that they only considered kernel estimators. Although one can identify a fourth-order kernel function that approximates a cubic smoothing spline (Silverman 1984; Messer 1991; Nychka 1989b), this correspondence is not exact. It appears that the martingale theory used in Härdle et al.'s proofs requires an exact kernel representation of the nonparametric estimator. One way to avoid this technical problem is to use the Gaussian approximation to a spline based on the strong approximation of a partial-sum process (Cox 1984). A brief description of this type of approximation is given in Section 4.

Despite these technical details, the limiting variance for $\hat{\lambda} - \hat{\lambda}_0$ agrees with what one would expect for a fourth-order kernel method. Using the form for η^2 reported in this section and the asymptotic relations from Nychka (1989a), it is possible to argue that

$$n^{1/18} \frac{(\hat{\lambda} - \hat{\lambda}_0)}{\hat{\lambda}_0} \zeta[\sigma^2/\gamma]^{-1/18} \xrightarrow{d} N(0, 1),$$

where $\gamma = \int_{[0,1]}(f^{(4)})^2\, dt$ and ζ is a constant that does not depend on f, σ, or n. The extremely slow rate of convergence is characteristic of smoothing parameter estimates of $\hat{\lambda}_0$ and is a reason to be skeptical of the accuracy of such asymptotic approximations for moderate sample sizes.

There are two practical problems in using the confidence intervals (3.5) or (3.6). First, η has a complicated form, and it is tedious to derive an explicit formula for this standard error. Second, even when this formula has been obtained, there is still the problem that η depends on unknown quantities such as λ_0 and σ. The last part of this section describes how η can be derived using MACSYMA and how the unknown quantities in this formula can be estimated.

Although the approach in Härdle et al. (1988) is to omit higher terms of η to obtain a simple expression, in this work an emphasis was placed on avoiding asymptotic approximations. For moderate sample sizes, higher order terms in η could easily be 10% or more of the size of the leading terms. For this reason approximations were avoided when possible. Moreover, because a symbol manipulation program was used to determine the form for η, little additional work is involved to derive an exact expression rather than an approximate one.

The basic idea behind using MACSYMA to compute η is to represent $\mathcal{U}_n(\lambda)$ as a quadratic form in $\mathbf{e}$ where the matrices involve powers of $A(\lambda)$. By making the correct identifications, one can differentiate and simplify these expressions as if they were polynomials.

First, the approximating random variable $\mathcal{U}_n$ will be defined. Note that $d_1(\hat{\lambda}_0)$ may be obtained by evaluating $\partial/\partial\lambda)(V(\lambda) - \mathrm{ASE}(\lambda))$ at $\lambda = \hat{\lambda}_0$. Accordingly, let

$$\mathcal{U}_n(\lambda) = (\partial/\partial\lambda)(V(\lambda) - \mathrm{ASE}(\lambda)). \tag{3.7}$$

The advantage of working with $\mathcal{U}_n(\lambda_0)$ is that the smoothing parameter no longer enters the expression as a random variable. This makes it much easier to analyze the variance of this quantity. Next $\mathcal{U}_n$ will be represented as a quadratic form with respect to $\mathbf{e}$.

Set $\mathbf{b} = (I - A(\lambda))\mathbf{f}$ and $m_k(\lambda) = \operatorname{tr}[A(\lambda)^k]/n$. Then

$$V(\lambda) = \frac{\frac{1}{n}(\mathbf{b} + (I - A(\lambda))\mathbf{e})^T(\mathbf{b} + (I - A(\lambda))\mathbf{e})}{(1 - m_1(\lambda))^2}$$

and

$$\mathrm{ASE}(\lambda) = \frac{1}{n}(\mathbf{b} + A(\lambda)\mathbf{e})^T(\mathbf{b} + A(\lambda)\mathbf{e}).$$

Because of the symmetry of $A(\lambda)$,

$$V(\lambda) - \mathrm{ASE}(\lambda) = \mathbf{b}^T G_1(\lambda)\mathbf{b} + \mathbf{b}^T G_2(\lambda)\mathbf{e} + \mathbf{e}^T G_3(\lambda)\mathbf{e}.$$

The matrices G_i will be polynomials in terms of $A(\lambda)$ with coefficients that may also include powers of m_1 and m_2. Using the diagonalization of $A(\lambda)$ given in Section 2 one can show that $(\partial/\partial\lambda)A(\lambda) = -1/\lambda(I - A(\lambda))A(\lambda)$ and $(\partial/\partial\lambda)m_k(\lambda) = -1/\lambda(m_k(\lambda) - m_{k+1}(\lambda))$. Therefore,

$$\mathcal{U}_n(\lambda) = \frac{1}{\lambda n}[\mathbf{b}^T H_1\mathbf{b} + \mathbf{b}^T H_2\mathbf{e} + \mathbf{e}^T H_3\mathbf{e}],$$

where once again H_k is a matrix polynomial in $A(\lambda)$ with coefficients that depend on powers of m_j, for $1 \le j \le 3$. The exact form for this expression was derived using MACSYMA and is reported in the Appendix.

The variance of $\mathcal{U}_n$ can be computed from the standard formulas for quadratic forms,

$$\eta^2 = \operatorname{var}(\mathcal{U}_n(\lambda)) = 2\sigma^4 \operatorname{tr}(H_3(\lambda)^2)/(\lambda n)^2 + (\kappa_4 - 3\sigma^4) \sum_{k=1}^{n} H_3(\lambda)_{kk}^2/(\lambda n)^2 - \rho_3 \sum_{k=1}^{n} H_3(\lambda)_{kk}(H_2(\lambda)\mathbf{b})_k/(\lambda n)^2 + \sigma^2 \mathbf{b}^T H_2(\lambda)^2 \mathbf{b}/(\lambda n)^2, \tag{3.8}$$

where $\rho_3 = E(e_1^3)$, $\kappa_4 = E(e_1^4)$.

The second and third terms on the right side of (3.8) are zero for normal errors, and even when the errors depart from normality, these terms will be asymptotically negligible. For this reason only the first and fourth terms in (3.8) were used to evaluate the variance. In order to evaluate these terms it is necessary to determine $H_2(\lambda)^2$ and $H_3(\lambda)^2$. These can also be computed symbolically and will just be polynomial expressions in $A(\lambda)$. Thus it follows that $(1/n)\mathbf{b}^T H_2(\lambda)^2 \mathbf{b} = \Sigma_{k=1}^{6} C_{2,k}\beta_k$ with $\beta_k = \mathbf{b}^T A(\lambda)^k \mathbf{b}^T/n$ and $(1/n)$ tr $H_3(\lambda)^2 = \Sigma_{k=1}^{4} C_{3,k} m_k(\lambda)$. Finally one can use MACSYMA to convert the algebraic expressions for the coefficients $C_{2,k}$ and $C_{3,k}$ into FORTRAN syntax. In this way these formulas can be incorporated into a subroutine without transcription error.

Clearly η depends on several unknown quantities. Even with an exact symbolic computation of η, it is necessary to substitute estimates for λ_0, σ, and β_k. Fortunately, under suitable conditions, these quantities can be estimated consistently using the smoothing parameter and spline from cross-validation.

To estimate η, λ_0 is replaced by $\hat{\lambda}$ and σ^2 is replaced by the estimate $\hat{S}_n^2 = \|(I - A(\hat{\lambda}))\mathbf{Y}\|^2/(n - \mathcal{C}m_1(\hat{\lambda}))$, where $\mathcal{C} = 37/32$. The estimates of $\beta_k(\lambda_0)$ used in this work are adapted from a consistent estimate of the average squared error described in Nychka (1990). From the properties of the cross-validation function, and under Assumptions 1–4 given in Section 6, $\widehat{\text{ASE}} = V(\hat{\lambda}) - \hat{S}_n^2$ is a consistent estimate of EASE(λ_0) in the sense that $\widehat{\text{ASE}}/\text{EASE}(\lambda_0) \xrightarrow{p} 1$ as $n \to \infty$. Now, since $\text{EASE}(\lambda_0) = \beta_0(\lambda_0) + \sigma^2 m_2(\lambda_0)$, from the consistency of $\hat{\lambda}$, one can estimate $\beta_0 \doteq \widehat{\text{ASE}} - \hat{S}_n^2 m_2(\hat{\lambda})$. Estimates for $\beta_k(\lambda_0)$ when $k > 0$ can be obtained by considering similar unbiased estimates, and their description can be found in Nychka (1989a).

4. SIMULATION–BASED CONFIDENCE INTERVALS

In contrast to confidence intervals based on asymptotic expansions, it is easy to construct intervals using simulation techniques. Recall that $\hat{\rho} = \ln(\hat{\lambda})$ and $\hat{\rho}_0 = \ln(\hat{\lambda}_0)$. The approach taken in this work estimates the critical points for $\hat{\rho} - \rho_0$ by simulating from a model where $\hat{f}$ is taken to be the true function and the measurement error is normally distributed with standard deviation $\hat{S}_n$. To be precise, for the model defined in (1.1), let F denote the distribution function for the measurement error, standardized to have unit variance. Let $T(f, \sigma, F)$ denote a random variable with the same distribution as $\hat{\rho} - \hat{\rho}_0$, and let Φ denote the standard normal distribution function. Then, for any data set, by Monte Carlo simulation one can compute the percentiles $\hat{C}_\beta$: $\Pr(T(\hat{f}, \hat{S}, \Phi) < \hat{C}_\beta) = \beta$. The basic assumption of this method is that the distribution of $T(\hat{f}, \hat{S}, \Phi)$ will be close to the distribution of $T(f, \sigma, F)$. If this is true, $\hat{C}_\beta$ is an estimate of the percentiles for $T(f, \sigma, F)$, and we are led to consider $(1 - \alpha)$ level confidence intervals for $\hat{\rho}_0$ of the form

$$(\hat{\rho} - \hat{C}_{\alpha/2}, \hat{\rho} + \hat{C}_{1-\alpha/2}).$$

An important feature of spline estimators is that simulating and minimizing the generalized cross-validation function can be greatly simplified by making use of the orthogonal decomposition described in Section 2. The distributions of $V(\lambda)$ and ASE(λ) depend on the transformed data, $\mathbf{y}$, rather than on the actual observations. For this reason it is more efficient to generate $\mathbf{y}$ directly and avoid performing a transformation for each bootstrap replication. Let $U_{iv} = \phi_v(t_i)$, then by the properties of this basis, $UU^T = nI$. If $\mathbf{Y} = \mathbf{f} + \mathbf{e}$, where $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, then $\mathbf{y} = U^T\mathbf{Y} = \mathbf{f} + \boldsymbol{\epsilon}$ with $\mathbf{f} = U^T\mathbf{f}$ and $\boldsymbol{\epsilon} = U^T\mathbf{e} \sim N(\mathbf{0}, n\sigma^2\mathbf{I})$. Since U only depends on the sequence $\{t_k\}$ for any problem, it need only be computed once. The computation of $V(\lambda)$ and ASE(λ) can be made even more efficient by truncating the sums in (2.4) and (2.5) for terms that are negligible.

One problem with this representation is when the error distribution departs from a normal. In this case the transformed errors will not be Gaussian or independent. The distribution of $V(\lambda)$ involves a weighted sum of the squared components of $\mathbf{e}$, however, and it is possible to approximate this by a sum depending on squared independent normal random variables. Cox (1984) has considered such a Gaussian approximation in his work. Let $V(\lambda, \mathbf{e})$ denote the generalized cross-validation function for data generated from the model (1.1) with the error vector, $\mathbf{e}$. Under Assumptions 1–4, Section 6, there is a probability space containing the random vectors $\bar{\mathbf{u}}$ and $\mathbf{u}$ such that Law($\mathbf{e}$) = Law($\mathbf{u}$), Law($\bar{\mathbf{u}}$) = $N(\mathbf{0}, \sigma^2\mathbf{I})$, and

$$|V(\lambda, \mathbf{u}) - V(\lambda, \bar{\mathbf{u}})| = o_p(\text{EASE}(\lambda)) \tag{4.1}$$

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

Although the convergence rate in (4.1) does not imply that $V(\lambda, \bar{\mathbf{u}})$ approximates the asymptotic distribution of $V(\lambda, \mathbf{u})$, it is conjectured that this is indeed true. Cox's technique relies on the strong approximation of a partial sum process by a Wiener process, and the convergence rate can be tied to the number of finite moments for the error distribution. Thus the convergence rate may be increased by strengthening the moment condition in Assumption 2. The derivatives of $V(\lambda)$ and ASE(λ) involve similar terms to $V(\lambda)$, and so it is reasonable to assume that this type of approximation can be extended to these random variables as well.

5. SIMULATION STUDY RESULTS AND DISCUSSION

A simulation study was conducted to investigate the reliability of these confidence intervals for λ. This study follows a 3 × 3 × 2 × 2 factorial design consisting of 3 different test functions, three levels of σ (.2, .4, .6), two sample sizes (128, 256), and two error distributions (Normal, T_6). Only 95% confidence levels were considered. To

simplify the programming, the observation points were taken to be equally spaced ($t_k = k/n$) and a periodic spline estimate was computed. With these restrictions, the orthogonal decomposition described in Section 2 can be readily found using the fast Fourier transform. The three test functions considered are mixtures of beta densities:

$$\text{Type I} \qquad \frac{1}{3}\beta_{10,5} + \frac{1}{3}\beta_{7,7} + \frac{1}{3}\beta_{5,10}$$

$$\text{Type II} \qquad \frac{6}{10}\beta_{30,17} + \frac{4}{10}\beta_{3,11}$$

$$\text{Type III} \qquad \frac{1}{3}\beta_{20,5} + \frac{1}{3}\beta_{12,12} + \frac{1}{3}\beta_{7,30},$$

where $\beta_{m,n}$ is the standard Beta density on [0, 1]. These are the same smooth test functions considered in Wahba (1983) and are plotted in Figure 2 of Nychka (1988).

For each of the 36 cases in this simulation study, 400 samples were generated from the additive model (1.1). The pseudo random normal deviates were generated using the Kinderman–Monahan ratio of two uniforms, and the T_6 random variables were constructed in the obvious way from 7 independent normals. The simulation-based confidence intervals were computed using 200 replications. It should be emphasized that, while T_6 errors were used to generate the data, the bootstrap replications assumed normally distributed errors.

The most important statistics from this study are those that quantify the level of the confidence procedure and the average width of the interval. Some summary statistics compiled from this study are reported in Table 1, however, a more efficient way of reviewing the results is by the panel of plots in Figure 5. The reader should keep in mind that all smoothing parameters have been transformed to a $\log_{10}$ scale.

It is useful to calibrate the width of these intervals with a confidence interval when the distribution of $\hat{\rho}_0 - \hat{\rho}$ is known. In this case, a suitable interval is $(\hat{\rho} - C_{.025}, \hat{\rho} + C_{.975})$, where $\Pr((\hat{\rho}_0 - \hat{\rho}) \leq C_\alpha) = \alpha$. As a reference, the width, $C_{.975} - C_{.025}$, for each of the 9 cases (and normal errors), is plotted near the x axis on each plot in Figure 5 and is listed in the last column of Table 1.

The main results of this study are illustrated clearly by Figure 5. In general the median interval widths for the two methods are comparable to the width when the distribution of $\hat{\rho}_0 - \hat{\rho}$ is known. The only instance where coverage seems to be poor is for asymptotic intervals applied to the first test function. The bootstrap-based confidence intervals tend to be shorter than those based on an asymptotic formula. Although the median widths for these two types of intervals agree, the 97.5 percentiles of the widths differ dramatically. At this percentile the bootstrap intervals are shorter by roughly a factor of two. Moreover, the level of the bootstrap intervals has a coverage probability that is closer to the nominal level of .95.

The simulation results for errors following a T_6 distribution are nearly the same as those given in Table 1, and for this reason they are not reported. The fact that these results are so close for different error distributions suggests that the Gaussian approximation described in Section 4 gives an accurate approximation to the distribution of the cross-validation function for the sample sizes considered in this study.

6. SOME ASYMPTOTIC PROPERTIES OF SIMULATING FROM A SPLINE ESTIMATE

This section discusses the large-sample properties of the simulation method for constructing confidence intervals. The validity of this method depends on approximating $T(f, \sigma, F)$ by $T(\hat{f}, \hat{S}, \Phi)$. Although a mathematically precise anal-

Table 1. A Comparison Between Two Types of Confidence Intervals for the $\log_{10}$ Smoothing Parameter

			Interval Type							
			Simulation				Asymptotic theory				
				Distribution of interval widths				Distribution of interval widths			Distribution known
Sample size	Test function	σ	Coverage probability	2.5%	50%	97.5%	Coverage probability	2.5%	50%	97.5%	Fixed width
$n = 128$	Type I	.2	.927	1.72	2.16	3.11	.905	1.05	1.81	5.52	2.22
		.4	.925	1.81	2.27	3.29	.890	1.10	2.06	9.41	1.86
		.6	.942	1.88	2.42	3.69	.873	1.11	2.06	8.53	2.67
	Type II	.2	.950	.90	1.13	1.67	.910	.67	1.07	3.61	2.14
		.4	.952	1.10	1.41	1.99	.952	.77	1.29	4.65	2.95
		.6	.967	1.29	1.64	2.42	.930	.86	1.41	5.36	2.48
	Type III	.2	.947	1.05	1.30	1.80	.932	.76	1.19	4.73	1.24
		.4	.962	1.22	1.57	2.26	.935	.81	1.47	4.98	.86
		.6	.977	1.48	2.11	3.06	.920	.92	1.55	6.58	1.60
$n = 256$	Type I	.2	.962	1.55	2.02	2.80	.942	1.07	1.77	8.97	1.14
		.4	.915	1.70	2.17	2.99	.885	1.12	2.05	7.40	1.62
		.6	.915	1.72	2.19	3.34	.885	1.17	2.10	9.18	1.37
	Type II	.2	.970	.74	.99	1.27	.980	.67	1.02	2.35	1.38
		.4	.960	.94	1.19	1.59	.940	.74	1.21	4.19	1.10
		.6	.960	1.09	1.37	1.75	.957	.81	1.32	3.95	1.68
	Type III	.2	.962	.88	1.15	1.50	.942	.73	1.10	3.17	1.29
		.4	.965	1.08	1.32	1.87	.942	.83	1.36	5.12	1.89
		.6	.970	1.27	1.56	2.33	.940	.85	1.47	6.36	1.51

NOTE: Normally distributed errors, nominal coverage is at 95%.

ysis of this correspondence is beyond the scope of this article it is worthwhile to compare the asymptotic properties of spline estimates based on f versus $\hat{f}$. For example, the limiting normal approximation to $T(f, \sigma, F)$ has a normalization factor that depends on the convergence rate of λ_0. It makes sense to approximate $T(f, \sigma, F)$ by $T(\hat{f}, \hat{S}, \Phi)$ only if the optimal value of the smoothing parameter for $\hat{f}$ converges to zero at the same rate as λ_0. It will be argued that this situation does hold. Moreover, it appears that the optimal smoothing parameter taking $\hat{f}$ as the true function is only slightly smaller than λ_0. This result is offered as a modest first step in a theoretical justification of the simulation method.

In order to compare the optimal choices of smoothing parameters, it is necessary to examine the properties of the EASE for the spline estimate when $\hat{f}$ is taken as the true function. Let $\mathbf{Y}_B$ denote the "data" generated from the model

$$Y_{B,k} = \hat{f}(t_k) + \epsilon_k, \qquad 1 \le k \le n,$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \hat{S}_n^2 \mathbf{I})$. Then the expected average squared error is

$$\mathrm{EASE}(\lambda)_B = \frac{1}{n}\|(I - A(\lambda))\hat{\mathbf{f}}\|^2 + \frac{\hat{S}_n^2 \operatorname{tr}[A(\lambda)^2]}{n}.$$

Here the B subscript indicates that this is under the model used for bootstrap replications.

The following assumptions are necessary.

Assumption 1. $\mathbf{Y}$ follows the model (1.1) where $\{e_k\}$ $1 \le k \le n$ are iid random variables $E(e_1) = 0$ and $\operatorname{var}(e_1) = \sigma^2$. Let t_k satisfy $G(t_k) = (k - 1)/(n - 1)$ where G is a cdf with a continuous, strictly positive density function g.

Assumption 2. $E((e_1)^8) < \infty$.

Assumption 3. $\hat{\lambda} \in [\lambda_n, \Lambda_n]$, $\lambda_n \sim n^{-4/5} \log(n)$, $\Lambda_n \to 0$ as $n \to \infty$.

Assumption 4. $f \in W_2^4[0, 1]$, f satisfies the "natural" boundary conditions, $f^{(2)}(0) = f^{(2)}(1) = f^{(3)}(0) = f^{(3)}(1) = 0$, and f is not a linear function.

These assumptions are standard for analyzing the asymptotic properties of splines. Assumption 2 is necessary to approximate the spline by a Gaussian process while Assumption 3 constrains λ to be a range where the asymptotic approximations are valid. The last assumption restricts f to be in the class of functions that achieves the best (nontrivial) rate of convergence for a second-order spline. The reader is referred to Rice and Rosenblatt (1983), Messer (1991), and Nychka (1990) for more discussion of this assumption and the related problem of boundary effects. Finally, it is necessary to define the integrals

$$L_k = \int_0^\infty \frac{dw}{(1 + w^4)^k}, \quad J(\kappa) = \int_0^\infty \frac{\kappa^2 w^8\, dw}{(1 + w^4)^2(1 + \kappa w^4)^2},$$

$$\text{and} \quad M(\kappa) = \frac{dJ(\kappa)}{d\kappa}.$$

Theorem 6.1. Let λ_{0B} be the minimizer of $\mathrm{EASE}_B(\lambda)$. Under Assumptions 1–4

$$\lambda_{0B} = K\lambda_0(1 + o_p(1)) \quad \text{as} \quad n \to \infty,$$

where $K \approx .92$ and is the solution to

$$K = [1 + 4K^{4/5}M(K)/L_2]^{-4/9}. \tag{6.1}$$

A proof of this theorem is outlined in Nychka (1989a).

The numerical results summarized in Section 5 suggest that the conclusion of this theorem may hold under weaker conditions than those stated here. Assumption 2 appears to be too conservative since the results for errors following a T_6 distribution were nearly the same as for normally distributed errors. Although the Type II test function does not satisfy Assumption 4, there is still good agreement between the bootstrap distribution and the actual one.

A parametric bootstrap method is not only convenient to compute but also appears to give useful approximations to the actual distribution of the estimated smoothing parameter. This approach may also be effective for other inference problems in nonparametric regression and deserves more study.

APPENDIX: MACSYMA SOURCE LISTING AND OUTPUT

This appendix includes a listing of the MACSYMA program used to compute the matrices H_1, H_2, and H_3 defined in Section 4. The representation of these expressions by MACSYMA is also given.

In these symbolic calculations the reader should make the following identifications:

Symbols in MACSYMA output	Notation used in this article
l	λ
b(l)	$(I - A(\lambda))\mathbf{f}$
a(l)	$A(\lambda)$
e	$\mathbf{e}$
m, iml	$\operatorname{tr} A(\lambda)/n$, $1 - \operatorname{tr} A(\lambda)/n$
m2	$\operatorname{tr} A(\lambda)^2/n$
con	H_1
lin	H_2
quad	H_3

MACSYMA Source Listing

```
phi:  ((b (l) + (1 - a(l)) * e)**2)/n + (1 - m(l))
        **2 * (-e**2/n - (b (l)**2 - 2 * b(l)
        * a(l) * e + e**2 * a(l)**2)/n);
phi:  phi/iml(l)**2;
phi:  combine(expand(phi))$
dphi: diff(phi, l)$
dphi: ev(dphi, diff(a(l),l) = -(1 - a(l)) * a(l)/
         l, diff(m(l),l) = -(m1 - m2)/l,
         m(l) = m1, diff(b(l),l) = -a(l) * b(l)/
         l, diff(iml(l),l) = (m1 - m2)/l)$
dphi: expand(dphi)$
dphi: ev(dphi, a(l)= A, b(l) = b, iml(l) = iml)$
dphi: ratsubst( im3, iml**3, dphi);
dphi: ratsubst( im2, iml**2, dphi)$
dphi: expand(ratsubst( 1 - m1, iml, dphi))$
con:  combine(coeff(dphi, b**2));
temp: expand( ratsubst(c, e * b, dphi) )$
lin:  ratsimp(combine(coeff(temp, c)), a);
```

```
quad: ratsimp(combine(coeff(dphi, e**2)), a, a**2,
      a**3);
```

MACSYMA Output

```
(c86)  con;

                  2 m2 - 2 a m1^3 + 6 a m1^2 - 4 a m1 - 2 m1
(d86)             ------------------------------------------
                                  im3 l n

(c87)  lin;

                    a (4 m2 - 2 m1^3 + 6 m1^2 - 10 m1 + 2) - 4 m2 + 4 m1
(d87)  - ------------------------------------------------------------------
                                      im3 l n

(c88)  quad;

                 (a^2 (2 m2 - 2 m1^3 + 6 m1^2 - 4 m1 - 2) + 2 m2
(d88)             + a (-4 m2 + 2 m1 + 2) + a^3 (2 m1^3 - 6 m1^2
                  + 4 m1) - 2 m1)/(im3 l n)
```

[Received July 1990. Revised October 1990.]

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.

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

Chiu, S. T. (1990), "On the Asymptotic Distribution of Bandwidth Estimates," *The Annals of Statistics,* 18, 1696–1711.

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

Craven, P., and Wahba, G. (1979), "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik,* 47, 107–122.

Diggle, P., and Hutchinson, M. F. (1989), "On Spline Smoothing With Autocorrelated Errors," *Australian Journal of Statistics,* 31, 166–182.

Eubank, E. (1988), *Spline Smoothing and Nonparametric Regression,* New York: Marcel Dekker.

Hall, P., Marron, J. S., and Park, B. U. (1989), "Smoothed Cross-Validation," unpublished manuscript.

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.

Hart, J., and Wehrly, T. (1986), "Kernel Regression Estimation Using Repeated Measurement Data," *Journal of the American Statistical Association,* 81, 1080–1088.

Li, K. C. (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.

Symbolics Incorporated (1984), MACSYMA Reference Manual (Beta Release 308.2), Cambridge, MA: Author.

Messer, K. (1991), "A Comparison of a Spline Estimate to Its 'Equivalent' Kernel Estimate," *The Annals of Statistics,* 19.

Nychka, D. W. (1988), "Bayesian Confidence Intervals for Smoothing Splines," *Journal of the American Statistical Association,* 83, 1134–1143.

——— (1989a), "Confidence Intervals for the Estimated Smoothing Parameter in Nonparametric Regression," North Carolina Institute of Statistics Mimeo Series, 1948.

——— (1989b), "Smoothing Splines as Locally Weighted Averages," North Carolina Institute of Statistics Mimeo Series, 1960.

——— (1990), "The Average Posterior Variance of a Smoothing Spline and a Consistent Estimate of the Average Squared Error," *The Annals of Statistics,* 18, 415–428.

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

Scott, D. (1988), Comment on "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?" by W. Härdle, P. Hall, and J. S. Marron, *Journal of the American Statistical Association,* 83, 96–98.

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.

Terrell, G., and Scott, D. (1986), "Biased and Unbiased Cross-Validation in Density Estimation," *Journal of the American Statistical Association,* 82, 1131–1146.

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

——— (1989), *Spline Models for Observational Data,* Philadelphia: SIAM.