APPROXIMATE REGRESSION MODELS AND SPLINES

by

R. L. Eubank

Technical Report No. 180
Department of Statistics ONR Contract

August 1983

Department of Statistics
Southern Methodist University
Dallas, Texas   75275

APPROXIMATE REGRESSION MODELS AND SPLINES

R. L. Eubank

Department of Statistics
Southern Methodist University
Dallas, Texas   75275

ABSTRACT

The literature pertaining to splines in regression analysis is reviewed. Spline regression is motivated as a simple extension of the basic polynomial regression model. Using this framework, the concepts of fixed and variable knot spline regression are developed and corresponding inferential procedures are considered. Smoothing splines are also seen to be an extension of polynomial regression and various optimality properties, as well as inferential and diagnostic methods, for these types of splines are discussed.

1.   INTRODUCTION

In recent years spline functions have enjoyed increased popularity as a tool in both theoretical and applied statistical research. They have been found useful for handling problems

such as nonparametric regression and data smoothing, analysis of
segmented regression models, nonparametric density estimation
and numerical differentiation of data. Splines have been uti-
lized for such diverse purposes as analysis of response curves
in Agriculture, Economics and Pharmacokinetics (Fuller (1969),
Poirer (1975) and Wold (1971)), estimation of the liquidity
trap in economics (Barth, Kraft and Kraft (1976)), determining
the base temperature in heat accumulation models (Gbur, Thomas
and Miller (1979)), calibration of nuclear materials processing
tanks (Lechner, Reeve and Spiegelman (1982)), analysis and
estimation of meteorological fields (e.g. Wahba and Wendelberger
(1980)), estimation of aerodynamic models for airplanes from
flight data (Smith and Klein (1982)) as well as a variety of
applications in Geophysics and Astrophysics (see e.g. Jupp and
Vozoff (1974) and Holt and Jupp (1978)) to name but a few.
This illustrated diversity of applications for splines is due,
in part, to their ability to provide simple approximate models
for complicated phenomena which are either difficult or impos-
sible to model precisely. In fact, spline regression models
are an extension of polynomial regression models which have long
been utilized for this purpose. In this paper an overview of
the role of splines in regression analysis will be provided
which focuses on the connection between these two types of
regression models.

Consider the classical input-response model where, for
given inputs $t_1, \ldots, t_n$ in some interval $[a,b]$, the corresponding
outputs satisfy,

$$y(t_i) = \eta(t_i) + \varepsilon(t_i) \ , \ i=1,\ldots,n, \tag{1.1}$$

for η some smooth response function and $\varepsilon(t_1),\ldots,\varepsilon(t_n)$ zero
mean uncorrelated random errors with common variance σ^2. If the
form of η is unknown or very complicated it is often feasible to
use an approximation to model (1.1) of the form

$$y(t_i) = p(t_i) + \varepsilon(t_i) \ , \tag{1.2}$$

where p is a polynomial of order m, i.e.,

$$p(t) = \sum_{j=0}^{m-1} \alpha_j t^j .$$

This has the advantage that the unknown parameters in the approximate model, namely $\alpha_0, \ldots, \alpha_{m-1}$, enter in a linear fashion and, hence, may be estimated by least squares. Moreover, p is known to be a good choice since, for example, polynomials provide a rich class of approximating functions which are, in fact, dense is the set of all continuous functions on [a,b] (see, e.g. Royden (1968, pg. 172)). If we are willing to assume that η is m times differentiable in [a,b] then, by Taylor's theorem with integral remainder, (1.1) can be written as

$$y(t_i) = p(t_i) + r(t_i) + \varepsilon(t_i) , \qquad (1.3)$$

where the α_j are constants which involve a, but not t,

$$r(t) = \int_a^t \frac{\eta^{(m)}(\xi)}{(m-1)!} (t-\xi)^{m-1} d\xi = \int_a^b \frac{\eta^{(m)}(\xi)}{(m-1)!} (t-\xi)_+^{m-1} d\xi,$$

$$(1.4)$$

and (1.4) uses the "+" function notation

$$x_+^k = \begin{cases} x^k & , \ x \geq 0 \\ 0 & , \ \text{otherwise.} \end{cases}$$

By further assuming that r(t) is so small on [a,b] as to be inconsequential one can then justify using model (1.2) in lieu of (1.1). These and other justifications for polynomial regression models are well known and, in many cases, such models will provide an entirely satisfactory description of a set of data. Such is not always the case, however, as will now be illustrated.

Figures 1 and 2 provide, respectively, plots of the best yearly times for the 800 meter Amateur Athletic Union (AAU) races from 1930-1972 (see Nougues and Sielken (1980, pg. 25)) and a

FIGURE 1. AAU Race Data
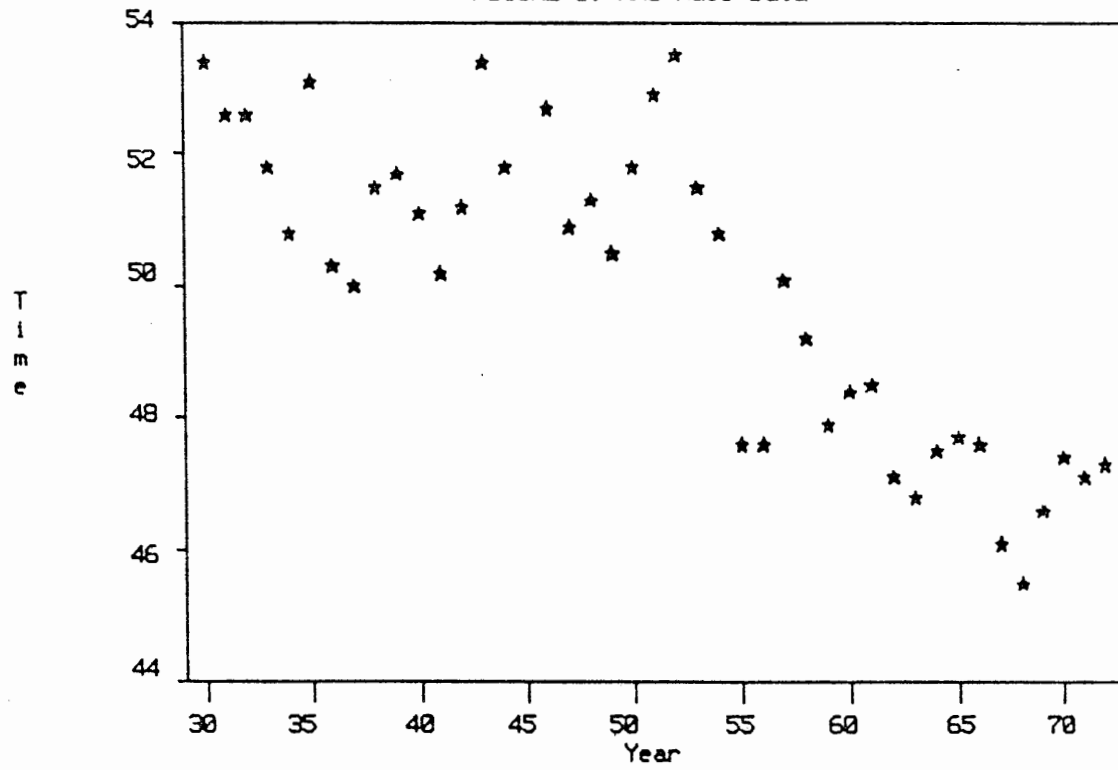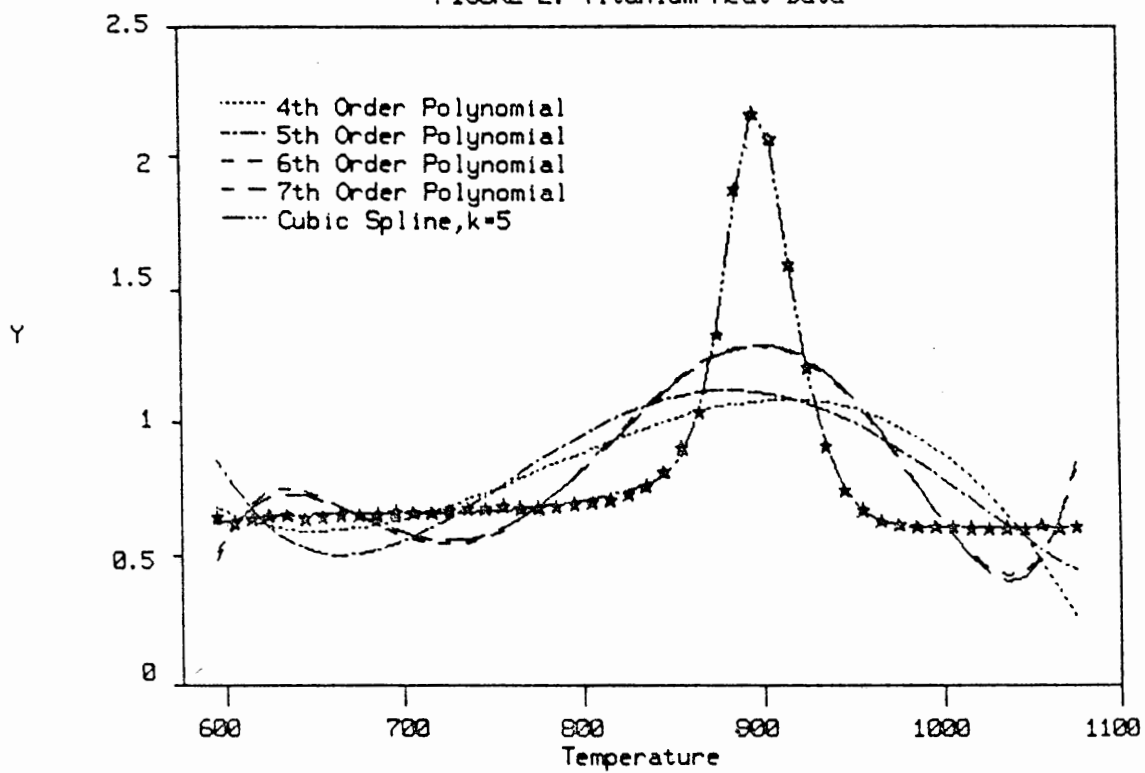
FIGURE 2. Titanium Heat Data

Legend:
- 4th Order Polynomial
- 5th Order Polynomial
- 6th Order Polynomial
- 7th Order Polynomial
- Cubic Spline, k=5

data set representing a property of titanium as a function of heat that will hereafter be referred to as the titanium heat data (see de Boor (1978, pg. 222)). Examination of the AAU race data suggests the possibility of a change in the structure of the underlying regression function occurring in the early 1950's. This possibility is made even more likely by the fact that new rubberized tracks were first utilized during this time period. Thus we might postulate a model for this data which allows for a structural change in the 1950's. Polynomials are inadequate for this purpose since they are globally determined by their values in any small interval and, hence, cannot adapt to model shifts. Furthermore, the use of a polynomial model for this data would make the hypothesis of most interest, namely the presence of a structural change in the early 1950's, difficult if not impossible to test. This example illustrates an important shortcoming of polynomial regression models; they are not well suited for modeling phenomena whose structure changes as a function of the independent variable.

Another difficulty frequently encountered in polynomial regression is illustrated by the titanium heat data in Figure 2. To obtain an adequate fit, successively higher order polynomials have been fitted to the data. Unfortunately, this results in estimators of η exhibiting oscillatory behavior that belies the smoothness visual perception would attribute to the underlying regression function. Such behavior can be explained, heuristically, by noting that the derivative of a polynomial is also a polynomial whose leading coefficient is a multiple of m. This implies the possibility of steep derivatives and, hence, increased oscillation for larger values of m.

A more general model, which alleviates many of the difficulties inherent in polynomial regression, can be obtained by dealing with the remainder term in (1.3). Rather than discarding it to obtain (1.2) let us instead approximate it by use of a quadrature formula. More specifically, for k suitably

chosen values, $\xi_1, \ldots, \xi_k$, satisfying

$$a < \xi_1 < \cdots < \xi_k < b \qquad\qquad (1.5)$$

and constants $\beta_1, \ldots, \beta_k$ which depend on $\xi_1, \ldots, \xi_k$, but not on t, we can write

$$r(t) \doteq \sum_{j=1}^{k} \beta_j (t-\xi_j)_+^{m-1} \quad,$$

where $\doteq$ means "approximately equal to." This provides a new approximate model

$$y(t_i) = s(t_i) + \varepsilon(t_i) \ , \ i=1,\ldots,n \qquad\qquad (1.6)$$

where

$$s(t) = \sum_{j=0}^{m-1} \alpha_j t^j + \sum_{j=0}^{k} \beta_j (t-\xi_j)_+^{m-1} \ . \qquad\qquad (1.7)$$

The function s is what is known as a <u>spline of order m with knots at</u> $\xi_1, \ldots, \xi_k$. For a fixed set of knots $\underline{\xi} = (\xi_1, \ldots, \xi_k)'$ the collection of all functions of the form (1.7) will be denoted by $S_{\underline{\xi}}^m$. It is well known (see de Boor (1978)) that $S_{\underline{\xi}}^m$ is a linear space of dimension m+k for which the functions, $1, t, \ldots, t^{m-1}$, $(t-\xi_1)_+^{m-1}, \ldots, (t-\xi_k)_+^{m-1}$ provide a basis.

Examination of model (1.6) reveals that i) the polynomial model (1.2) corresponds to the special case of $\beta_1 = \cdots = \beta_k = 0$, ii) for fixed $\underline{\xi}$ this is a linear model so that the unknown parameters $\alpha_0, \ldots, \alpha_{m-1}, \beta_1, \ldots, \beta_k$ can be estimated by least squares and iii) the regression function, s, consists of k+1 polynomial segments that are tied together at the knots, $\xi_1, \ldots, \xi_k$, in a fashion which insures the existence of m-2 continuous derivatives. The piecewise nature of s makes it a reasonable approximation for models with changing structure and, by providing more locally adaptive fits, the order of each segment can usually be kept small, often resulting in smoother estimates of η. This latter point is illustrated by comparing the fit of the cubic spline (m=4) with five knots to that of the polynomials in Figure 2.

The knots, $\xi_1,\ldots,\xi_k$, in model (1.7) may or may not be of interest in their own right. In certain instances, e.g., when η is postulated as having a segmented nature, they may be the parameters of principal interest. In these situations our objective will usually be to test hypotheses about their values or, perhaps, to construct point or interval estimates. On the other hand, for data summary or prediction purposes the precise values used for the ξ's may be unimportant as long as they result in a good fit to the data. Consequently, the ξ's in (1.7) might be i) specified a priori through hypotheses to be tested, ii) considered as unknown parameters in the model or, iii) candidates for their values might be selected through initial data analysis using some ad hoc methods. Both the first and last of these possibilities will be termed a case of _fixed knots_ since, in either event, when $\alpha_0,\ldots,\alpha_{m-1}$, $\beta_1,\ldots,\beta_k$ are estimated the ξ's are treated as being known. Fixed knot spline regression is the subject of Section 2. If the ξ's are included in the model specification as parameters to be estimated this is called _variable_ or _free knot spline regression_ and is discussed in Section 4.

A special class of splines, known as smoothing splines, also arises from model (1.3) as a result of certain smoothing and prediction considerations. An extensive literature has evolved on these types of splines which is discussed in Section 3.

Before proceeding it should be noted that this paper focuses entirely on splines which arise from or are used in regression problems. Whereas this limitation in scope provides for a more detailed development of the area it does not allow us to exhibit the breadth of applications for splines in statistics. For example, there is a voluminous literature on the use of splines in density estimation that will not be discussed here. For many of the important references in this area we refer the reader to the excellent review paper by Wegman and

Wright (1983). Some important discussions on the use of splines in the closely related problem of nonparametric regression can be found in comments by Wahba to Stone (1977) and the references she cites therein.

2. SPLINE REGRESSION: FIXED KNOTS

Throughout this section it is assumed that values to be utilized for the knots have been specified. Although these values will frequently arise from hypotheses to be tested, certain rules of thumb for knot placement, when this is not the case, can be found in Wold (1974) and Lenth (1977). Other ad hoc knot selection techniques will be discussed subsequently.

Since for specified $\underline{\xi} = (\xi_1, \ldots, \xi_k)'$ (1.7) is linear in $\alpha_0, \ldots, \alpha_{m-1}$, $\beta_1, \ldots, \beta_k$, the unknown parameters in model (1.6) can be estimated by ordinary least squares. Equation (1.7) is known as the "+" function or truncated power basis representation for s. Fuller (1969) was apparently one of the first to utilize "+" functions in connection with spline and other types of piecewise polynomial regression. As we will see, the "+" function basis has advantages from the point of view of statistical hypothesis testing. However, it is not the best suited for estimation and evaluation purposes since the design matrix for this form of the model can be poorly conditioned and the number of arithmetic operations required to evaluate s(t) depends on the location of t relative to the knots. These properties may lead to numeric inaccuracies, especially when many knots are use.

The problems inherent in the truncated power basis representation for s can be eliminated, in part, by using a piecewise polynomial basis for $S_{\underline{\xi}}^m$. It can be shown (see, e.g., de Boor (1978)) that any function of the form (1.7) can be written as

$$s(t) = \sum_{j=0}^{m-1} \delta_{jr} t^j , \ t \in [\xi_r, \xi_{r+1}), \ r = 0, \ldots, k , \qquad (2.1)$$

where $\xi_0 = a$, $\xi_{k+1} = b$ and the δ_{jr}'s satisfy the linear (continuity) constraints

$$s^{(\ell)}(\xi_r -) = s^{(\ell)}(\xi_r +) , \quad \ell = 0, \ldots, m-2, \quad r = 1, \ldots, k. \qquad (2.2)$$

The converse to this can also be established so that functions of the form (2.1), subject to (2.2), provide an alternative basis for $S_{\underline{\xi}}^m$. This segmented definition of s alleviates the evaluation difficulties associated with the "+" function representation.

The piecewise polynomial basis for splines was utilized for spline regression by Buse and Lim (1977) who related their approach to an alternative representation used in the pioneering work of Poirier (1973). In view of the constraints (2.2), it is seen that estimation using this representation for s must be accomplished by restricted least squares. Thus, explicit forms for the estimators as well as techniques for their computation can be obtained, for example, from Gerig and Gallant (1975) who also provide a spline oriented example. However, the restricted least squares procedure is cumbersome and a total of mk + m parameters must be estimated as compared to the "correct number," m+k, which is the dimension of $S_{\underline{\xi}}^m$.

A compromise between the "+" function and piecewise polynomial basis that provides for both ease of evaluation and a well conditioned design matrix is obtained through use of B-splines. Define 2m "additional knots"

$$\xi_{-(m-1)} = \cdots = \xi_0 = a$$

and

$$b = \xi_{k+1} = \cdots = \xi_{k+m}.$$

Then, the values for $B_{i,m}$, the B-spline of order m with knots at $\xi_i, \ldots, \xi_{i+m}$, can be obtained via the recurrence relation

$$B_{i,m}(t) = \frac{t-\xi_i}{\xi_{i+m-1}-\xi_i} B_{i,m-1}(t) + \frac{\xi_{i+m}-t}{\xi_{i+m}-\xi_{i+1}} B_{i+1,m-1}(t) \qquad (2.3)$$

with

$$B_{i,1}(t) = \begin{cases} 1 & , \quad \xi_i \leq t < \xi_{i+1} \ , \\ 0 & , \quad \text{otherwise} \ , \end{cases} \qquad (2.4)$$

for $i = 0, \ldots, k$. In using (2.3) we employ the fact that a B-spline of order q with $q+1$ coincident knots vanishes identically. It can be demonstrated (see de Boor (1978)) that $B_{-(m-1),m}, \ldots, B_{k,m}$ provide a basis for S_ξ^m on $[a,b]$ and, hence, there are constants $\gamma_{-(m-1)}, \ldots, \gamma_k$ such that

$$s(t) = \sum_{j=-(m-1)}^{k} \gamma_j B_{j,m}(t) \ , \quad t\epsilon[a,b]. \qquad (2.5)$$

It follows easily by induction from (2.3) and (2.4) that $B_{j,m}(t) = 0$ if $t \notin [\xi_i, \xi_{i+m}]$. Consequently, B-splines have local support and evaluation of $s(t)$ for any $t \ \epsilon \ [a,b]$ will involve only m of the B-splines. Perhaps of more importance is the fact that the design matrix for model (1.6), with s expressed in terms of B-splines, is well-conditioned for moderate m. In fact, since $\Sigma_{r=1}^{n} B_{i,m}(t_r) B_{j,m}(t_r)$ vanishes for $|i-j| \geq m$, the normal equations for this case are banded with only $m-1$ non-zero offdiagonal bands. An early example of the use of B-splines in spline regression can be found in Fuller (1969) who considered quadratic ($m=3$) cardinal (equally spaced ξ_j's) B-splines for trend removal from time series.

B-splines are most readily evaluated by means of computer, with FORTRAN code for this purpose available in de Boor (1978). They appear to be the basis best suited for computational purposes, particularly in large data sets or when many knots are utilized. The principal difficulty with their use is that they vanish outside of $[a,b]$ and, hence, provide a representation for s only on this interval. Consequently, to extrapolate beyond $[a,b]$ one must convert the fitted spline to either the piecewise polynomial or truncated power basis representation and use the first or last polynomial segment, as appropriate, for this purpose (see de Boor (1978) for the details involved in such conversions.) This is

only a minor drawback since extrapolation from a segmented model would seem ill advised in most cases.

For further discussion of the three bases for splines presented here see Cox (1971) who also considers another basis derived using Chebyshev polynomials. A more statistically oriented comparison of alternative spline bases can be found in Nougues and Sielken (1980).

Assuming normal errors in model (1.6), hypothesis tests and confidence intervals based on the least squares parameter estimates can be obtained using standard linear model methodology. However, some of these hypotheses for splines have special interpretations which merit further discussion.

A hypothesis that is frequently of interest in spline models stems from the belief that the true regression function, η, has a segmented nature with the ξ_j's being postulated points at which the different segments are connected. If s provides a close approximation to η a hypothesis of interest is, for example, that ξ_j is <u>active</u> in the sense that s has a jump in its (m-1)-st derivative at ξ_j. If this were not the case then s would consist of a single polynomial segment on $[\xi_{j-1}, \xi_{j+1})$. Consequently, failure to reject the hypothesis that ξ_j is not active would lead us to suspect that the same functional relationship governs input and output on both $[\xi_{j-1}, \xi_{j+1})$ and $[\xi_j, \xi_{j+1})$. This hypothesis is easily tested using the "+" function representation since it is equivalent to

$$H_0 : \beta_j = 0. \tag{2.6}$$

The required t-statistic for this test is standard output from linear models packages which, as noted by Smith (1979), is an important advantage of this basis. Generalizations of this idea to more than one knot follow immediately. For example, the hypothesis of a global model, i.e. that the model is (1.2), is tantamount to $H_0 : (\beta_1, \ldots, \beta_k)' = \underline{0}$.

To test the hypothesis that ξ_j is not active using piece-wise polynomials we must test $H_0 : \delta_{j-1,m} = \delta_{j,m}$. This can also be accomplished using a t-statistic, given the estimated δ_{jr}'s and their estimated variance-covariance matrix (see Gerig and Gallant (1975)). Since this requires a statistical package or programs that can perform restricted least squares, the required test statistics may be difficult to compute in practice.

For the B-spline basis (2.6) is equivalent to
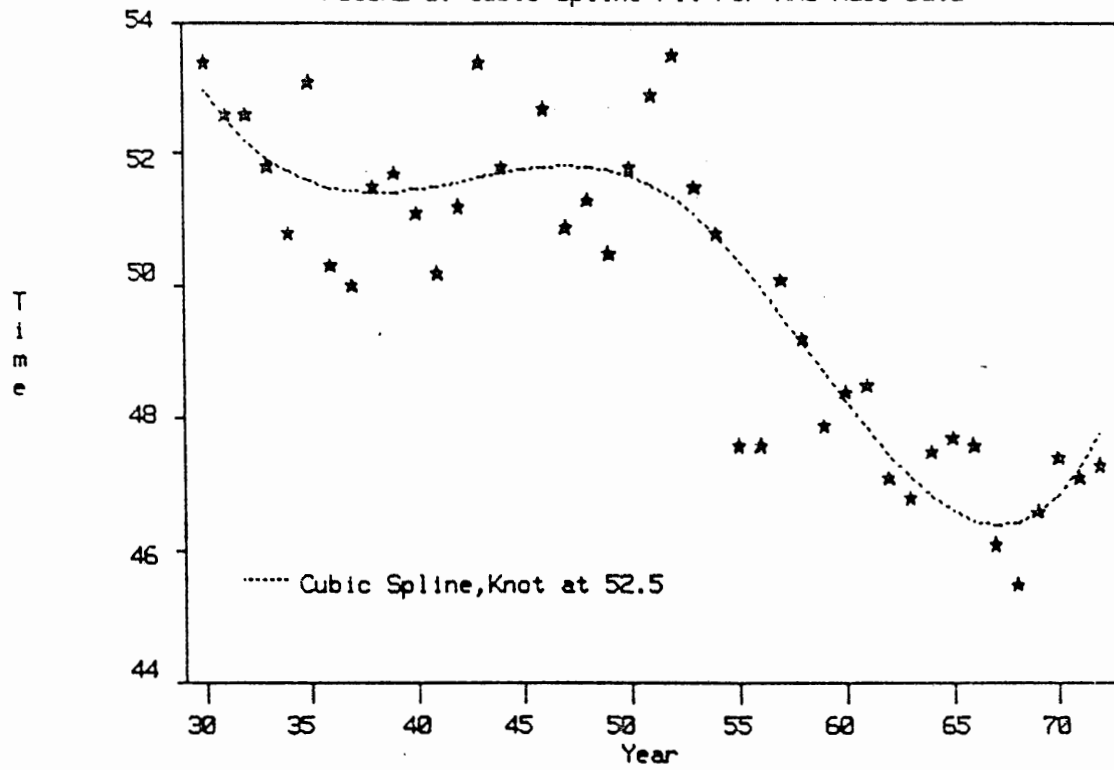
$$H_0 : -B_{i-m,m}^{(m-1)}(\xi_i-)\gamma_{i-m} + \sum_{j=i-m+1}^{i-1} [B_{j,m}^{(m-1)}(\xi_i+) - B_{j,m}^{(m-1)}(\xi_i-)]\gamma_j$$

$$+ B_{i,m}^{(m-1)}(\xi_i+)\gamma_i = 0. \qquad (2.7)$$

Smith (1982) has shown that the coefficients for the γ_j's in (2.7) form a contrast and provides an algorithm for computing the t-statistic required for the test.

As an illustration consider the AAU race data discussed in Section 1. Figure 3 shows the results of fitting a cubic spline (m=4) to this data with a knot at $\xi_1 = 52.5$. To test the hypothesis that the introduction of rubberized track in the early 1950's resulted in a structural change in the relationship governing yearly race times we might test the hypothesis $H_0 : \beta_1 = 0$. The resulting t-statistic is found to be significant at the .0003 level which leads us to believe such a change did, in fact, occur.

The diagnostics utilized for spline regression models include those available from standard regression analysis. Thus, one examines residuals, deleted residuals, Cook's distance measures, etc. to ascertain any modeling errors or overly influential observations (see Chapter 7 of Gunst and Mason (1980) for an account of these topics). An important modeling error, which is peculiar to spline models, results from the omission or misspecification of one or more knots. Such errors are frequently manifested through periodic or patterned behavior for the residuals (see Lechner, Reeve and Spiegleman (1982) for an example of this

FIGURE 3. Cubic Spline Fit For AAU Race Data

in the simple case of m=2). However, the problem of developing diagnostics tailored to the detection of knot specification errors has not, as yet, received much attention.

Several statistics that can be used, either to indicate reasonable ad hoc knot choices or as model diagnostics can be motivated from the work of Chow (1960) and Brown, Durbin and Evans (1975). The basic idea is as follows. Consider first the polynomial model (1.2) and, for $\ell \geq m+1$, let $\hat{p}_\ell(t)$ denote the estimate of $p(t)$ based on $y(t_1),\ldots,y(t_\ell)$. Then define the recursive residuals W_ℓ, $\ell=m+1,\ldots,n$ by

$$W_\ell = \frac{y(t_{\ell+1}) - \hat{p}_\ell(t_{\ell+1})}{\sqrt{1 + \underline{x}'_{\ell+1}[X'_\ell X_\ell]^{-1}\underline{x}_\ell}} . \tag{2.8}$$

where X_ℓ denotes the design matrix corresponding to $y(t_1),\ldots,y(t_\ell)$ and $\underline{x}_{\ell+1} = (1,t_{\ell+1},\ldots,t_{\ell+1}^{m-1})'$. If model (1.2) holds with normal errors it follows from Brown, Durbin and Evans (1975), that these will be independent $N(0,\sigma^2)$ random variables and can be utilized for detecting model shifts. They suggest using measures based on the W_ℓ such as

$$D_{j1} = \sum_{\ell=m+1}^{j} W_\ell/\hat{\sigma}_n, \quad j=m+1,\ldots,n \tag{2.9}$$

and

$$D_{j2} = (j-m)\hat{\sigma}_j^2/(n-m)\hat{\sigma}_n^2, \quad j=m+1,\ldots,n \tag{2.10}$$

where $\hat{\sigma}_j^2 = (j-m)^{-1}\sum_{i=1}^{j}(y(t_i)-\hat{p}_j(t_i))^2$. If the value of one of these statistics departs from within prescribed bounds (see Brown, Durbin and Evans (1975)) this would suggest the placement of a knot near the point of deviation. On the other hand, failure of this to occur would indicate that an overall polynomial model was acceptable. If ξ_1 is a knot selected using one of (2.9) or (2.10), then the previous type of procedure can now be repeated, assuming instead a spline model of order m with a knot at ξ_1, by computing the corresponding recursive residuals

and analogs of (2.9) - (2.10) to detect the presence of another knot, ξ_2, etc. A procedure similar to the one outlined here has been used effectively by Ertel and Fowlkes (1976) for the selection of breakpoints in piecewise linear regression. Related statistics which could also be utilized for knot placement can be found in the work of Chow (1960).

Recursive residuals should also be useful for the detection of model misspecifications such as the omission of a knot. This can be argued, heuristically, by first noting that, for example, the error from the $L_2[a,b]$ projection of an m-th order spline with knots at $\xi_1, \ldots, \xi_k$ onto splines with knots at $\xi_1, \ldots, \xi_{j-1}$, $\xi_{j+1}, \ldots, \xi_k$ is an m-th order spline with knots at $\xi_1, \ldots, \xi_k$. Consequently, the residuals from a spline fit, where a knot ξ_j has been omitted, should behave like a spline with knots at $\xi_1, \ldots, \xi_k$. Thus, by reasoning along these lines it should be possible to adapt the previous statistics for the detection of model misspecification as well.

The basic spline model can be generalized in any or all of several directions. For example, our assumptions regarding the random errors could be altered to include the case when σ^2 is not constant but varies between the different polynomial segments. Assuming that the ε's are still normal and independently distributed, the maximum likelihood equations for this altered model are readily derived and can be solved, in theory, to obtain variance and parameter estimates. Examples of equations of this type can be found in Quandt (1958), for piecewise linear regression models, and Whitten (1971), for the case of natural splines. In practice it is usually more convenient to employ some approximate solution such as using the average of a segment's squared residuals from the usual least squares fit to estimate that segment's variance. This approach was found to be quite successful by Whitten (1971). Another generaliza-

tion of this nature would be to assume the independent variable
is also measured with error. Gbur and Dahm (1982) have consi-
dered estimation for such models in the special case of m=2,
k=1.

The spline model can also, in some cases, be too smooth in
the sense of admitting more derivatives than η at the ξ_j's.
Thus, it is frequently more appropriate to use a more general
model with s in (1.6) now defined by

$$s(t) = \sum_{j=0}^{m-1} \alpha_j t^j + \sum_{r=1}^{k} \sum_{j=\nu_r}^{m-1} \beta_{j,r} (t-\xi_r)_+^j \ , \qquad (2.11)$$

for $\underline{\nu} = (\nu_1,\ldots,\nu_k)'$ some specified vector of integers between
0 and m-1. The parameter ν_j represents the number of continuity
constraints (number of continuous derivatives plus one) imposed
on s at ξ_j. The new model, (1.6) and (2.11), agrees with our
previous model, (1.6) and (1.7), when $\nu_j = m-1$, j=1,...,k, and
becomes a discontinuous piecewise polynomial of order m when
$\nu_j = 0$, j=1,...,k. Fuller (1969,1976), Gallant (1974) and
Gallant and Fuller (1973) consider models of this form with
$\nu_j = 2$, j=1,...,k, which they call grafted polynomials. Functions
of the form (2.11) are also frequently called splines and can
be expressed in terms of piecewise polynomials or analogs of
B-splines obtained by "stacking" $m-\nu_j$ knots at each ξ_j (see
de Boor (1978)). Important hypotheses for this more general
model include $H_0 : \beta_{\nu_j,j} = 0$, which is equivalent to s admitting
one more derivative than originally specified at ξ_j, and $H_0 :$
$(\beta_{\nu_j,j},\ldots,\beta_{m-1,j})' = \underline{0}$, which has the interpretation that ξ_j
is not active. Such hypotheses may be tested through standard
linear models methodology using the "+" function basis. Smith
(1982) has developed certain tests of this form using B-splines.
See Smith (1979) for an illuminating discussion of the relation-
ship between, and appropriate tests for comparing, models ob-
tained from (2.11) through alternative choices for m and/or the
ν_j's.

Fuller (1969), Gallant (1974), Gallant and Fuller (1973) and Smith (1979) have discussed the use of spline models (in the context of (2.11)) which have different order polynomial segments. This can be accomplished through constraining the coefficients and, hence, will usually require the use of restricted least squares for inferential purposes. An illustration of the use of restricted least squares for such models can be found in Gerig and Gallant (1975).
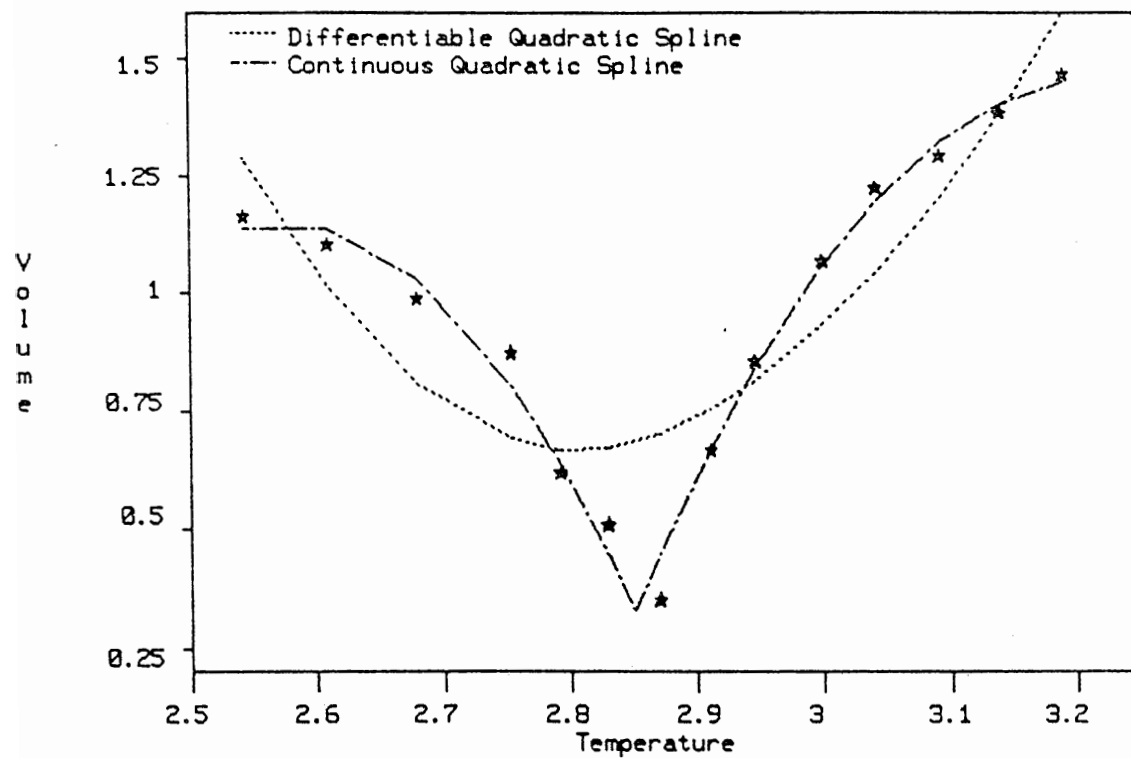
As an example of the use of model (1.6) with (2.11) consider the data on specific retention volume of methylene chloride in polyethelene terephthalate (c.f. Gallant (1977)) which is plotted, as a function of temperature, in Figure 4. A fit which uses $s_1(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_{21}(t-2.85)_+^2$ is clearly inadequate. Alternatively, the spline function $s_2(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_{11}(t-2.85)_+ + \beta_{21}(t-2.85)_+^2$ provides a good representation of the data that is much more consistent with what is indicated by the plot, namely, possible continuity but not differentiability of η, the true response function, at the knot. It should be noted that a test of $H_0 : \beta_{11} = 0$ results in rejection at the .0001 level and, hence, the continuity reduction at ξ_1 is also statistically warranted.

The use of robust regression techniques for spline models has been considered by Lenth (1977). Instead of estimating the coefficients in (2.11) by minimizing the sum of squared residuals, estimators are obtained by minimizing a function of the form

$$\sum_{j=1}^{n} \rho(y(t_j) - s(t_j)) ,$$

where ρ is some appropriately chosen loss measure. As illustrated by Lenth (1977), the combination of robust regression methodology with the inherent robustness already available from a spline model provides an extremely flexible curve fitting procedure.

FIGURE 4. Methylene Chloride Data

The concept of spline regression can also be extended to the case of more than one independent variable. Unfortunately, such extensions suffer from the lack of a totally satisfactory basis for splines in higher dimensions. Spline regression models involving several independent variables are discussed, for example, in Fuller (1969) and Whitten (1972).

The problem of optimal design construction for spline regression models appears quite difficult in general. Various types of optimal designs for certain special spline models have been derived by Studden and Van Arman (1969), Murty (1971), Draper, Guttman and Lipow (1974a,b, 1977) and Park (1978).

3. SMOOTHING SPLINES

A smoothing spline corresponding to model (1.3) is a spline of order 2m which has a knot at each of the observation points. Consequently, we assume throughout this section that

$$a < t_1 < \cdots < t_n < b .$$

An explicit expression for this type of spline will be given in Section 3.2. We will first, however, explore the statistical motivations for smoothing splines. These are, perhaps, best provided by simply detailing several of the regression related problems which they solve.

3.1 Optimality Properties of Smoothing Splines

Consider again model (1.3) where r represents the departure from a polynomial regression model. A measure of the size of this remainder is provided by

$$||r^{(m)}||^2 = \int_a^b (r^{(m)}(t))^2 \, dt = ||\eta^{(m)}||^2 , \qquad (3.1)$$

which is also a measure of curvature smoothness for η (see Wegman (1982) for a discussion of "derivative and curvature smoothness"). Since η is believed to be a smooth function, (3.1) is not expected to be large and, consequently, functions with large mth derivatives, in the sense of (3.1), should not be considered as estimators of the response function. To incorporate these ideas into an estimation scheme we might estimate η by using the function which minimizes a penalized least squares criterion such as

$$\frac{1}{n} \sum_{j=1}^{n} (y(t_j) - f(t_j))^2 + \lambda ||f^{(m)}||^2 \qquad (3.2)$$

over all f's in an appropriate function class, for some given $\lambda \epsilon (0,\infty)$. The criterion (3.2) represents a compromise between fidelity to the data and smoothness which, when minimized over

$$W_2^m[a,b] = \{f : f^{(j)} \text{ is absolutely continuous} \qquad (3.3)$$

$$j = 0, \ldots, m-1 \text{ and } ||f^{(m)}|| < \infty \},$$

results in the <u>smoothing spline</u> estimator for η, which will be denoted by $\hat{\eta}_{n,\lambda}$. It is of historical interest to note that smoothing splines were apparently first derived by Schoenberg (1964) who, motivated by the work of Whittaker (1923), proposed the use of (3.2) as a data smoothing criterion.

As an alternative to the previous development suppose that, either through assumptions or prior knowledge regarding the smoothness of η, it is possible to specify a bound for $||r^{(m)}||$ such as

$$||r^{(m)}|| \leq \alpha. \qquad (3.4)$$

Thus model (1.3) can now be thought of as an approximate linear model in the sense of Sacks and Ylvisaker (1978). Lacking better knowledge of η than (3.4), it would seem reasonable to predict η at $t \epsilon [a,b]$ by the linear estimator $\sum_{j=1}^{n} c_j^* y(t_j)$ which solves

$$\max_{||r^{(m)}|| \leq \alpha} E[(y(t) - \sum_{j=1}^{n} c_j^* y(t_j))^2]$$

$$= \min_{c_1,\ldots,c_n} \max_{||r^{(m)}|| \leq \alpha} E[(y(t) - \sum_{j=1}^{n} c_j y(t_j))^2].$$

$$(3.5)$$

If $\eta \in W_2^m[a,b]$, then the solution to (3.5), when viewed as a function of t, is identical to the minimizer of (3.2) with $\lambda = \sigma^2/n\alpha^2$. This result is due to Speckman (1982c).

Another method of deriving smoothing splines can be obtained by drawing a parallel with models for discrete time stochastic processes. A common model for discretely indexed time series is $y(t) = f(t) + \varepsilon(t)$, $t=0,\pm1,\ldots$, where ε is a white noise process, and f consists of deterministic and, possibly, stochastic components but is assumed uncorrelated with the ε process. In practice, f is usually adequately represented by removing a polynomial trend and treating the remainder as a moving average process of the form $\sum_{j=0}^{q} b_j \omega(t-j)$, where $\{\omega(t) ; t=0,\pm1,\ldots\}$ is also a white noise process. Thus, the approach has been to write

$$y(t) = \sum_{j=0}^{m-1} \alpha_j t^j + r(t) + \varepsilon(t)$$

and approximate $r(t)$ by a stochastic process. This idea has the obvious parallel of approximating $r(t)$ in (1.4) by a continuous time process such as $Z(t) = \int_a^b b(t-u)dW(u)$ where $\{W(u) ; u \in [a,b]\}$ is a zero mean process having the same covariance kernel as that of the Wiener process and $b(\cdot)$ is some specified function. In particular, the choice $b(x) = x_+^{m-1}/(m-1)!$ results in a Z process with covariance kernel

$$Q(s,t) = \int_a^b \frac{(t-u)_+^{m-1}(s-u)_+^{m-1}}{(m-1)!^2} du. \qquad (3.6)$$

Consequently, by analogy with results from time series, we might consider using an approximate model such as

$$y(t_i) = \sum_{j=0}^{m-1} \alpha_j t_i^j + \sigma_s Z(t_i) + \varepsilon(t_i), \quad i=1,\ldots,n, \qquad (3.7)$$

where σ_S is some positive scale parameter and the Z's are
assumed to have zero means, covariances given by (3.6) and
be uncorrelated with the ε's. For any given t the best linear
unbiased predictor (BLUP) of y(t) for this model was derived
by Kimeldorf and Wahba (1970b) but can be obtained by more
classical means through reference to Goldberger (1962).
Kimeldorf and Wahba (1970b) recognized that, when viewed as a
function of t, the BLUP for (3.7) is the same function which
minimizes (3.2) with $\lambda = \sigma^2/n\sigma_S^2$ and, hence, is a smoothing spline.
The idea of viewing smoothing splines as stemming from polynomial
approximation of η with the remainder modeled as a stochastic
process is due to Wecker and Ansley (1981, 1983).

In the event that $\varepsilon(t_i) = 0$, i=1,...,n, in (3.7) and,
hence, only the process $\sum_{j=0}^{m-1}\alpha_j t^j + \sigma_S Z(t)$ is observed, it
follows from Kimeldorf and Wahba (1970b) that the BLUP is the
smoothest spline, in the sense of (3.1), which interpolates the
data. Extensions of this result to more general processes such
as those having mean value functions that are not polynomials
are given in Peele and Kimeldorf (1977, 1979).

Instead of the three previous viewpoints we might adopt
the Bayesian philosophy of Blight and Ott (1975). They note
that once η has been represented as far as seems feasible by
a polynomial in (1.3), the experimenter then regards the $r(t_i)$'s
as errors with zero predicted values. Thus, although r is
deterministic, this belief can be represented by assuming that
the $r(t_i)$'s have a zero mean Gaussian prior distribution. The
α_j's can also be assumed to have a normal prior distribution
with mean $\underline{\mu}$ and variance-covariance matrix γI, which can be
justified by recalling that the Bayesian estimator for the α_j's
in ordinary polynomial regression reduces to the least squares
estimator when $\gamma \to \infty$. This leads us to assume that η has the
same prior distribution as the stochastic process

$$X(t) = \sum_{j=0}^{m-1} \alpha_j t^j + \sigma_S Z(t), \quad t \in [a,b] \quad , \tag{3.8}$$

where $\underline{\alpha}$ is $N_m(\underline{\mu}, \gamma I)$ and Z is a zero mean Gaussian process that is independent of the α_j's. Wahba (1978) assumes that $\underline{\mu} = \underline{0}$ and Z has covariance kernel (3.6). She then shows that as $\gamma \to \infty$ the BLUP of $\eta(t)$ for model (3.8) reduces to the solution of the previous problem. Thus spline smoothing is also equivalent to a Bayesian prediction problem for a polynomial regression model with a diffuse prior on the polynomial coefficients. For a discussion of the relationship between the Blight and Ott model (3.8) and similar models discussed in Young (1977) and O'Hagan (1978) see Steinberg (1983).

Before presenting an explicit form for the smoothing spline $\hat{\eta}_{n,\lambda}$, it should be noted that although both stochastic and deterministic models lead to the use of $\hat{\eta}_{n,\lambda}$ as an estimator of η these two types of models are philosophically quite different. For example, it can be shown (see Wahba (1981a)) that the sample paths for the Bayesian model are not in $W_2^m[a,b]$.

3.2 The Hat Matrix

For a given λ, the smoothing spline is a linear estimator. Consequently, the vector of fitted values, $\hat{\underline{\eta}}_{n,\lambda} = (\hat{\eta}_{n,\lambda}(t_1), \ldots, \hat{\eta}_{n,\lambda}(t_n))'$ can be written as

$$\hat{\underline{\eta}}_{n,\lambda} = H(\lambda)\underline{y} . \tag{3.9}$$

By analogy with regression analysis we call $H(\lambda)$ the hat matrix since it transforms $\underline{y}$ to $\hat{\underline{y}} = \hat{\underline{\eta}}_{n,\lambda}$. Various representations for $H(\lambda)$ can be found in work by Reinsch (1967, 1971), Anselone and Laurent (1968), Kimeldorf and Wahba (1970b, 1971), Demmler and Reinsch (1975) and Wahba (1978, 1980). Let

$$T = \{t_i^j\}_{\substack{i=1,n \\ j=0,m-1}} \tag{3.10}$$

and

$$Q_n = \{Q(t_i, t_j)\}_{i,j=1,n} \tag{3.11}$$

Then, Wahba (1978) has shown that given any $n \times (n-m)$ matrix U
of rank $n-m$ satisfying

$$U'T = 0 \quad , \tag{3.12}$$

where 0 is an $(n-m) \times m$ matrix of all zeroes, we have

$$I - H(\lambda) = n\lambda U(U'Q_n U + n\lambda U'U)^{-1}U' \quad . \tag{3.13}$$

Alternative forms for $H(\lambda)$, i.e., alternative bases for smoothing
splines, can then be regarded as corresponding to different
choices for U. A natural choice is to take

$$U'U = I \tag{3.14}$$

since, in this case,

$$I - H(\lambda) = n\lambda U \Gamma D(\lambda) \Gamma' U' \quad , \tag{3.15}$$

where Γ is the matrix of eigenvectors for $U'Q_n U$ and $D(\lambda)$ is a diag-
onal matrix involving the eigenvalues $d_1, \ldots, d_{n-m}$ of $U'Q_n U$ that
is given explicitly by

$$D(\lambda) = \text{diag}\left(\frac{1}{n\lambda + d_1}, \ldots, \frac{1}{n\lambda + d_{n-m}} \right) \quad . \tag{3.16}$$

The matrix $H(\lambda)$ is closely related to the hat matrix for
polynomial regression

$$H = T(T'T)^{-1}T' \quad . \tag{3.17}$$

It follows, for example, from Hoaglin and Welsch (1978) that
the elements of H satisfy i) $0 \leq h_{ii} \leq 1$, ii) $-1 \leq h_{ij} \leq 1$
and iii) $h_{ii} = 1$ iff $h_{ij} = 0$ for all $j \neq i$. Using representation
(3.15) it is possible to show that the elements of $H(\lambda) = \{h_{ij}(\lambda)\}$
have similar properties. More precisely, it can be shown that
(see Eubank (1983))

$$0 \leq h_{ii}(\lambda) \leq 1 \quad , \tag{3.18}$$

$$-1 \leq h_{ij}(\lambda) \leq 1, \tag{3.19}$$

$$h_{ii}(\lambda) = 1 \quad \text{iff} \quad h_{ij}(\lambda) = 0 \quad \text{for all } j \neq i \tag{3.20}$$

and, furthermore, that

$$h_{ii}(\lambda) \downarrow h_{ii} \;, \quad h_{ij}(\lambda) \to h_{ij} \quad \text{as } \lambda \to \infty \tag{3.21}$$

and

$$h_{ii}(\lambda) \uparrow 1 \;, \quad h_{ij}(\lambda) \to 0 \quad \text{as } \lambda \to 0. \tag{3.22}$$

By analogy with regression theory the values of $h_{ii}(\lambda)$ are called <u>leverage</u> <u>values</u> since they tell us the influence that $y(t_i)$ has on its own prediction. Due to (3.18)-(3.22) the $h_{ii}(\lambda)$'s can be used as a diagnostic tool in the same manner that leverage values are utilized in ordinary regression analysis (see Hoaglin and Welsch (1978)) to indicate sensitive points among the t_j's where $y(t_j)$ may be overly influential in determining the fit.

Other properties of $\hat{\eta}_{n,\lambda}$ follow from (3.15) and (3.21)-(3.22). For instance, $\hat{\eta}_{n,0}$ is now recognized as the smoothest spline which interpolates the data and $\hat{\eta}_{n,\infty} = \lim_{\lambda \to \infty} \hat{\eta}_{n,\lambda}$ satisfies

$$\hat{\underline{\eta}}_{n,\infty} = T(T'T)^{-1}T'\underline{y} \;. \tag{3.23}$$

This has the consequence that $\hat{\eta}_{n,\infty}$ is the least squares predictor obtained from model (1.3) by regressing $\underline{y}$ on the polynomials $1, t, \ldots, t^{m-1}$. The latter fact verifies that <u>smoothing</u> <u>splines</u> <u>are</u> indeed <u>a</u> <u>generalization</u> <u>of</u> <u>polynomial</u> <u>regression</u>.

3.3 Selection of λ

The quantity λ is usually called the <u>smoothing</u> <u>parameter</u> and can be regarded as the "tuning knob" which controls the tradeoff between fidelity to the data and smoothness. A visually satisfactory value of λ can usually be selected through trial and error. However, no optimality properties can be attributed to values selected in this fashion. Ideally, we might wish the selected λ to minimize the true mean square error (MSE),

$$\text{MSE}(\lambda, n) = n^{-1} \sum_{j=1}^{n} (\eta(t_j) - \hat{\eta}_{n,\lambda}(t_j))^2 \;. \tag{3.24}$$

This random variable is, of course, unobservable and the obvious alternative is to use its sample estimate

$$\hat{MSE}(\lambda,n) = n^{-1}\sum_{j=1}^{n} e_{n,\lambda}(t_j)^2 \tag{3.25}$$

where $e_{n,\lambda}(t_j)$ is the jth residual defined by

$$e_{n,\lambda}(t_j) = y(t_j) - \hat{\eta}_{n,\lambda}(t_j) . \tag{3.26}$$

Unfortunately, (3.25) is always minimized at $\lambda=0$, since $\hat{\eta}_{n,0}$ interpolates the data. This fact led Wahba and Wold (1975a) to suggest using the smoothing parameter value which minimized the cross-validation (CV) criterion

$$CV(\lambda,n) = n^{-1} \sum_{j=1}^{n} (y(t_j) - \hat{\eta}_{n,\lambda}^{[j]}(t_j))^2 \tag{3.27}$$

where $\hat{\eta}_{n,\lambda}^{[j]}$ is the smoothing spline fit to $y(t_1),\ldots,y(t_{j-1})$, $y(t_{j+1}),\ldots,y(t_n)$, $j=1,\ldots,n$. The use of (3.27) may be justified, intuitively, by the belief that a good value of λ should be one for which $\hat{\eta}_{n,\lambda}^{[j]}(t_j)$ is a good predictor of the missing data value $y(t_j)$.

Motivated again by standard regression terminology we define the _deleted residual_

$$e_{n,\lambda}^{[j]} = y(t_j) - \hat{\eta}_{n,\lambda}^{[j]}(t_j)$$

and note that Craven and Wahba (1979) have established the remarkable identity

$$e_{n,\lambda}^{[j]} = e_{n,\lambda}(t_j)/(1-h_{jj}(\lambda)) \tag{3.29}$$

which parallels results from linear regression (see Hoaglin and Welsch (1978)). Thus

$$CV(\lambda,n) = \sum_{j=1}^{n} e_{n,\lambda}(t_j)^2/(1-h_{jj}(\lambda))^2 \tag{3.30}$$

which can be utilized, along with (3.21)-(3.22), to understand how CV works.

It follows easily from Schoenberg (1964) that $\hat{MSE}(\lambda,n)$ is a monotone increasing function of λ and, hence, the residuals

tend to decrease in magnitude, on the average, as λ decreases. Since $(1-h_{jj}(\lambda))^{-1}$ increases monotonically as λ decreases, we now see that $CV(\lambda,n)$ is merely a weighted version of $\widehat{MSE}(\lambda,n)$ which utilizes the weights $(1-h_{jj}(\lambda))^{-2}$ to counteract the tendency to choose λ as zero. On the other hand, the CV criterion also guards against the choice $\lambda=\infty$ since the decrease in $(1-h_{jj}(\lambda))^{-2}$ obtained by increasing λ tends to be counteracted by an increase in the average size of the residuals. Hopefully, the value selected will then reflect the correct balance, somewhere between total fidelity to the data, obtained when $\lambda=0$, and the "smoothest possible" fit, realized at $\lambda=\infty$.

Equation (3.30) suggests that other criteria which utilize weights having similar properties to the $(1-h_{jj}(\lambda))^{-2}$'s, such as their average or median, might be expected to work as well. Craven and Wahba (1979) proposed using the former choice and termed the resulting criterion generalized cross-validation (GCV). Their criterion is given explicitly by

$$GCV(\lambda,n) = n^{-1}[\tfrac{1}{n} \text{ trace } (I-H(\lambda))]^{-2} \sum_{j=1}^{n} e_{n,\lambda}(t_j)^2. \qquad (3.31)$$

They then showed that (3.31) provides a rotation invariant version of (3.30) and established the important GCV Theorem. This theorem has the consequence that for $\eta \varepsilon W_2^m[a,b]$,

$$|E[MSE(\lambda,n)] + \sigma^2 - E[GCV(\lambda,n)]|/E[MSE(\lambda,n)] \le g(\lambda) \qquad (3.32)$$

where $g(\lambda)$ is a function involving n and the trace of $H(\lambda)$ and $H(\lambda)^2$. Equation (3.32) has the implication that, if $g(\lambda)$ is small, then $GCV(\lambda,n)$ is an estimator of $E[MSE(\lambda,n)]$ with near constant bias. Consequently, we would expect $GCV(\lambda,n)$ to track $MSE(\lambda,n)$ and both to be minimized at approximately the same value of λ. In the case of equally spaced t_i's, Craven and Wahba (1979) verified that $g(\lambda)$ is, in fact, small for sufficiently large n and appropriately chosen values for λ which allowed them to conclude that if $\{\tilde{\lambda}(n)\}$ is a sequence of minimizers of $E[GCV(\lambda,n)]$,

$$E[MSE(\tilde{\lambda}(n),n)]/\inf_{\lambda} E[MSE(\lambda,n)] \downarrow 1, \text{ as } n \to \infty. \qquad (3.33)$$

The case of unequally spaced sample points follows from work in Speckman (1981b). A data oriented version of (3.33) has been proved by Speckman (1982a) who shows that, uncer certain regularity conditions, a sequence of minimizers of $GCV(\lambda,n)$, $\{\hat{\lambda}(n)\}$ say, satisfies

$$MSE(\hat{\lambda}(n),n)/\inf_{\lambda} MSE(\lambda,n) \to_p 1 \quad \text{as } n \to \infty,$$

where $\to_p$ denotes convergence in probability.

If instead of a deterministic η we assume that η is stochastic as in model (3.8) the previous optimality results no longer apply. However, it follows from Wahba (1977a,b) that both $E_\eta E_\varepsilon[MSE(\lambda,n)]$ and $E_\eta E_\varepsilon[GCV(\lambda,n)]$ have the same minimizer, namely, $\lambda = \sigma^2/n\sigma_s^2$, where E_η and E_ε denote expectation with respect to the prior distribution of η and the distribution of the ε's. Thus, GCV can also be expected to work well in this case.

Efficient algorithms for the estimation of λ using GCV have been developed by Utreras (1979, 1980, 1981b) for equally spaced data. A program which provides a cubic smoothing spline fit to data with λ selected by GCV is contained in the IMSL package.

It follows from Wahba (1978) that for $\eta \in W_2^m[a,b]$ the asymptotically optimum λ is $0(n^{-2m/2m+1})$ and the corresponding asymptotically optimum MSE is bounded above by a term that is $0(n^{-2m/2m+1})$. It is shown in Wahba (1975) that, for equally spaced t_j's and a periodic $\eta \in W_2^{2m}[a,b]$, the optimal MSE enjoys the better rate $0(n^{-4m/4m+1})$. (See also Wahba and Wold (1975b) and Wahba (1977c) for other related results.) A similar set of conclusions have been obtained by Speckman (1981b) pertaining to the integrated mean square error (IMSE)

$$IMSE(\lambda,n) = \int_a^b (\eta(t) - \hat{\eta}_{n,\lambda}(t))^2 dt.$$

He shows that if $\eta \in W_2^m[a,b]$ then, by taking $\lambda = 0(n^{-2m/2m+1})$, it is always possible to achieve the rate $0(n^{-2m/2m+1})$ for the IMSE. This is consistent with Wahba's results for the MSE. However, a rate of $0(n^{-4m/4m+1})$ for the IMSE when $\eta \in W_2^{2m}[a,b]$ holds only if η satisfies the boundary conditions $\eta^{(j)}(a) = \eta^{(j)}(b) = 0$,

$j = m, \ldots, 2m-1$ as otherwise the behavior of the IMSE near a and b is dominated by boundary terms. (Similar results have been obtained by Rice and Rosenblatt (1981, 1983) and this behavior was observed empirically by Wold (1974).) This fact has an interesting consequence for the comparison of smoothing and variable knot splines which will be discussed in Section 4.

A variety of other applications for GCV can be found in Golub, Heath and Wahba (1979) and Wahba (1977a,b, 1980, 1982a,b). In particular, GCV can be used to select an appropriate value for m. This possibility is explored, for example, by Gamber (1979a,b), Wahba and Wendelberger (1980) and Speckman (1982a).

3.4 Inference and Diagnostics

For the Bayesian model it is shown in Wahba (1981a) that the posterior variance-covariance matrix for $\underline{\eta} = (\eta(t_1), \ldots, \eta(t_n))'$ satisfies

$$V(\underline{\eta} | \underline{y}) = \sigma^2 H(\lambda). \tag{3.34}$$

Thus, as suggested by Gamber (1979b) and Wahba (1981a), we might use

$$\hat{\eta}_{n,\lambda}(t_i) \pm Z_{\alpha/2} \sqrt{\sigma^2 h_{ii}(\lambda)} , \tag{3.35}$$

where $Z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentage point of the standard normal distribution, to provide an approximate $100(1-\alpha)\%$ confidence interval for $\eta(t_i)$. Of course σ^2 in (3.35) is usually unknown. To overcome this difficulty, Wahba (1981a) has proposed an estimator of σ^2, patterned after the usual variance estimate in linear models, which we define by

$$\hat{\sigma}^2_{n,\lambda} = \sum_{j=1}^{n} e_{n,\lambda}(t_j)^2 / \mathrm{tr}(I - H(\lambda)). \tag{3.36}$$

Her simulation results indicate that the confidence intervals obtained from (3.35) with σ^2 estimated by $\hat{\sigma}^2_{n,\lambda}$ can be quite satisfactory and she also gives an argument for why they should work for $\eta \in W_2^m[a,b]$ when λ is selected as the GCV estimate. Alternatives to (3.35) include interval estimates proposed

by Wecker and Ansley (1983) as well as those that can be obtained by jackknifing. This latter approach has been suggested by Wold (1974) and Nougues and Sielken (1980).

As an example consider the data in Figure 5 which was simulated using $\eta(t) = 4.26(e^{-3.25t} - 4e^{-6.5t} + 3e^{-9.75t})$ and normal errors with $\sigma = .1$. This function is a rescaled version of one utilized by Wahba and Wold (1975a) for their illustrations. Plotted along with the data are the true response function, the smoothing spline estimate of η with λ select through GCV, and Bayesian approximate 95% confidence intervals for $\eta(t_j)$, j=1,...,n.

Define the residual vector $\underline{e}_{n,\lambda} = (e_{n,\lambda}(t_1),...,e_{n,\lambda}(t_n))'$ and note that, for the model (3.7),

$$V(\underline{e}_{n,\lambda}) = V((I-H(\lambda))\underline{y}) = \sigma^2(I-H(\lambda)). \qquad (3.37)$$

This relationship parallels results for the variances and co-variances of residuals from linear models and suggests a variety of diagnostic measures patterned after their counterparts in regression analysis. Examples include the "studentized residuals" $\tau_{n,\hat{\lambda}}(t_i) = e_{n,\hat{\lambda}}(t_i)/\hat{\sigma}_{n,\hat{\lambda}}(1-h_{ii}(\hat{\lambda}))^{1/2}$, i=1,...,n, where $\hat{\lambda}$ is the GCV estimate of λ, "Cook's distance measures"

$$D_{n,\hat{\lambda}}(t_i) = \tau_{n,\hat{\lambda}}(t_i)^2 h_{ii}(\hat{\lambda})/[(1-h_{ii}(\hat{\lambda})) \ \text{tr}(H(\hat{\lambda}))]$$
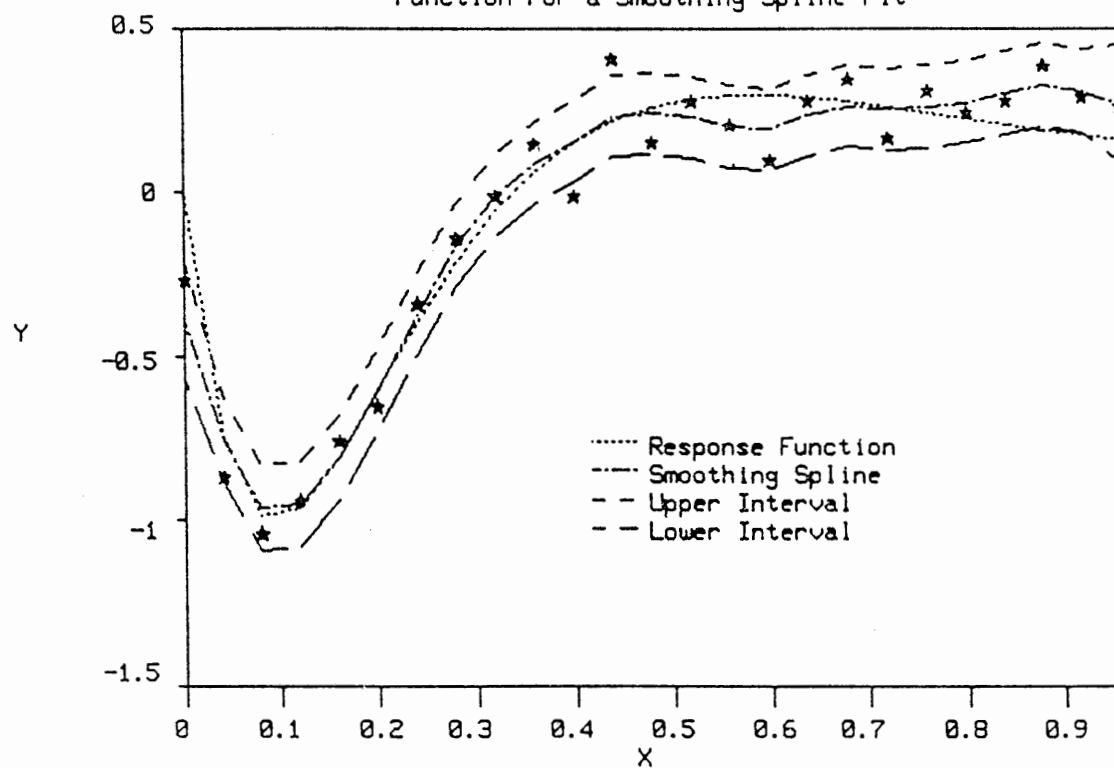
and the "studentized deleted residuals"

$$\tau_{n,\hat{\lambda}}^{[i]}(t_i) = e_{n,\hat{\lambda}}(t_i)/\hat{\sigma}_{n,\hat{\lambda}}^{[i]}(1-h_{ii}(\hat{\lambda}))^{1/2}$$

where $\hat{\sigma}_{n,\hat{\lambda}}^{[i]}$ is the estimator of σ corresponding to the observation set $y(t_1),...,y(t_{i-1}),y(t_{i+1}),...,y(t_n)$ (see Eubank (1983)). Wendelberger (1981) also notes the usefulness of probability plots for the residuals and Wahba (1980) suggests that, when σ^2 is known, $\text{tr}(I-H(\lambda))\hat{\sigma}_{n,\lambda}^2/\sigma^2$ can be compared to values of a χ^2 random variable with $\text{tr}(I-H(\lambda))$ "degrees of freedom" to assess goodness-of-fit. The development of diagnostic methods for smoothing

FIGURE 5. Bayesian Confidence Intervals and the Response
Function For a Smoothing Spline Fit

splines has, as yet, to receive much attention in the literature. Perhaps, some of the measures proposed here will provide a starting point for more rigorous study and future developments.

3.5 Generalizations and Related Work

As is the case with spline regression, the concept of robust smoothing splines can also be developed by considering functions which minimize a criterion such as

$$\sum_{j=1}^{n} \rho(y(t_j) - f(t_j)) + \lambda||f^{(m)}||^2$$

for some suitable function ρ which measures the size of a residual. Under quite general conditions the solution is still found to be a spline. This type of robust smoothing spline has been investigated by Anderssen, Bloomfield and McNeil (1974), Huber (1979), Utreras (1981a) and Cox (1983).

If we modify model (1.3) to obtain the more general model

$$y(t_i) = \sum_{j=0}^{m-1} \alpha_j \phi_j(t_i) + r(t_i) + \varepsilon(t_i) \qquad (3.38)$$

where the ϕ_j's, j=0,...,m-1, are now m arbitrary regression functions, a generalized version of our basic smoothing spline can be derived as the minimizer of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (y(t_i) - \psi_i(f))w_{ij}(y(t_j) - \psi_j(f)) + \lambda|||Lf|||^2, \qquad (3.39)$$

where the w_{ij} are positive weights, the ψ_j's are given linear functionals, L is a linear differential operator and $|||\cdot|||$ is some appropriate norm. The smoothing spline studied in previous sections correponds to $\phi_j(t) = t^j$, $w_{ii}=1$ and $w_{ij}=0$ (i≠j), $\psi_j(f)=f(t_j)$, $L = d^m/dt^m$ and $|||\cdot||| = ||\cdot||$. Thus, (3.38) - (3.39) represent a variety of modifications to the problem considered previously. Results corresponding to smoothing splines which stem from extensions such as these and their connection with BLU, Bayesian and minimax prediction can be found in Kimeldorf and Wahba (1970a,b, 1971), Wahba (1978), Weinert (1978),

Weinert and Kailath (1974), Weinert and Sidhu (1978), Weinert, Desai and Sidhu (1979), Weinert, Byrd and Sidhu (1980), and Speckman (1979, 1981a). Constraints can also be added to the estimation criterion such as those considered by Wright and Wegman (1980) and Wegman (1982).

There are also multivariate extensions of the smoothing spline concept. For example, in the case of two independent variables, t and x, and observations (y_i, t_i, x_i), $i=1, \ldots, n$, an analog of model (3.1) is

$$y_i = \sum_{j+k \leq m-1} \alpha_{j,k} t_i^j x_i^k + r(t_i, x_i) + \varepsilon_i$$

for which a reasonable smoothing criterion is

$$n^{-1} \sum_{j=1}^{n} (y_j - f(t_j, x_j))^2 + \lambda \sum_{j=0}^{m} \binom{m}{j} \int_{\mathbb{R}^2} \left\{ \frac{\partial^m}{\partial t^j \partial x^{m-j}} f(t,x) \right\}^2 dx dt.$$

$$(3.40)$$

When formulated in the proper function space, the minimizer of (3.40) is a bivariate spline known as the thin plate spline. Various modification and generalizations of this type of spline and its extension to more than two dimensions are investigated in Dyn, Wahba and Wong (1979), Wahba (1979, 1981b,c, 1982b), Wahba and Wendelberger (1980), Wong (1981) and Wendelberger (1981, 1982). It is of interest to note that the predicted values for such multivariate models can still be expressed as $\hat{\eta} = H(\lambda) \underline{y}$ where $H(\lambda)$ has a form similar to (3.15). Thus, analogs of equations (3.10)-(3.22) and (3.29) hold in the multivariate setting which can be utilized to motivate the use of GCV for the selection of λ and suggest various inferential procedures and diagnostics. The use of leverage values may be particularly useful in this case.

Finally, the question arises of how the t_i's should be selected in designed experiments. A general formulation of this problem has been given by Wahba in her discussion to O'Hagan (1978). However, this problem remains, as yet,

unsolved. Some problems closely related to the one she posed are considered by Sacks and Ylvisaker (1970), Wahba (1971) and Eubank, Smith and Smith (1982). In the case of interpolating splines ($\lambda = 0$) some design questions have been answered by Speckman (1982b).

4. SPLINE REGRESSION: FREE KNOTS

In this section we again consider the generalized spline model (1.6) wherein s is now defined by (2.11). However, in contrast to the development in Section 2, values for the knots are no longer assumed to be known or specified and, hence, the ξ_j's are now included as parameters to be estimated from the data. Let $d = m + \sum_{j=1}^{k}(m-\nu_j)$ and denote the d x 1 vector of linear parameters in (2.11) by

$$\underline{\theta} = (\alpha_0, \ldots, \alpha_{m-1}, \beta_{\nu_1,1}, \ldots, \beta_{m-1,1}, \ldots, \beta_{\nu_k,k}, \ldots, \beta_{m-1,k})'. \quad (4.1)$$

Also define the collection of "all possible knot sets" by

$$D_k = \{(\xi_0, \ldots, \xi_{k+1}): a = \xi_0 < \xi_1 < \cdots < \xi_k < \xi_{k+1} = b\} \quad (4.2)$$

and, for a given $\underline{\theta} \in \mathbb{R}^d$ and $\underline{\xi} \in D_k$, denote the ith residual by

$$e_i(\underline{\theta},\underline{\xi}) = y(t_i) - \sum_{j=0}^{m-1} \alpha_j t_i^j - \sum_{r=1}^{k} \sum_{j=\nu_r}^{m-1} \beta_{j,r}(t_i-\xi_r)_+^j . \quad (4.3)$$

Then, for some specified integer k and vector of continuity constraints $\underline{\nu}$, the problem we consider, initially, is the selection of $\underline{\theta}^* \in \mathbb{R}^d$ and $\underline{\xi}^* \in D_k$ which minimize

$$SSE(\underline{\theta},\underline{\xi}) = \sum_{j=1}^{n} e_j(\underline{\theta},\underline{\xi})^2 . \quad (4.4)$$

The simultaneous estimation of both $\underline{\theta}$ and $\underline{\xi}$ using (4.4) is called free or variable knot spline regression and, since s is not linear in the ξ_j's, is recognized as a problem of non-linear least squares.

Perhaps the simplest example of a variable knot spline problem, which is of practical interest, corresponds to the

case of a discontinuous linear spline with one knot. Such
models can be written as

$$y(t_i) = \begin{cases} \delta_{00} + \delta_{10}t_i + \varepsilon(t_i) \ , \ i=1,\ldots,q \ , \\ \\ \delta_{01} + \delta_{11}t_i + \varepsilon(t_i) \ , \ i=q+1,\ldots,n \ , \end{cases} \tag{4.5}$$

with q being the number of observations obtained from the first
linear segment. Selection of q is equivalent to selecting ξ_1,
in this case, since the precise location of a knot between any
two observations is immaterial in discontinuous models. The
problem of estimating the parameters in (4.5) and/or testing
the hypothesis that a change in regression regimes has occurred
is classically known as the problem of switching regressions.
It was first considered by Quandt (1958, 1960) and has, since,
received considerable attention in the literature. Many of the
important references and an interesting new approach to this
problem can be found in Worsley (1983).

Sprent (1961) was perhaps the first to consider a con-
tinuous version of (4.5), i.e., to assume that the two line
segments intersect at some point, ξ_1, between t_q and t_{q+1}.
Under the assumption that q was known and the errors were normal,
he developed a likelihood ratio test for the hypothesis that
ξ_1 has some specified value. This approach was later generalized
by Robinson (1964), to include higher order polynomial segments.
Hudson (1966) relaxed the assumption that q be known and de-
veloped an algorithm for estimation of both the coefficients in
each segment and the knot ξ_1. His results also extend to higher
order polynomial segments and $k+1 > 2$ segments. Hudson's
algorithm, in the case of k=1, was improved by Hinkley (1969)
(see also Hinkley (1971)) who also derived the asymptotic dis-
tribution of the estimators for this case. Other algorithms
for parameter estimation in linear splines with $k \geq 1$ knots
have been developed by Bellman and Roth (1969), and Ertel and
Fowlkes (1976). The latter authors even allow k to be a free

variable but consider only knots of the form $\xi_i = (t_i + t_{i+1})/2$.

A Bayesian approach to parameter estimation for linear splines with a single knot has been investigated by Smith and Cook (1980). Bacon and Watts (1971) also utilize a Bayesian framework but assume that some specified transition function can be treated as governing the change from one segment to another and, consequently, they estimate the value of a transition parameter instead of a knot.

Estimation in spline models which allow for general values of m and k, but no continuity constraints (i.e., $\nu_j=0$, $j=1,\ldots,k$ in (4.3)), has been considered by Cox (1971) and Guthery (1974) who both employ a dynamic programming approach. McGee and Carleton (1970) have developed an approximate estimation procedure for such models through the use of cluster analysis which also allows for the estimation of k.

Gallant (1974) and Gallant and Fuller (1973) in their study of grafted polynomials were apparently among the first to utilize nonlinear regression methodology for parameter estimation in spline models. It follows from their work that nonlinear regression procedures such as the modified Gauss-Newton algorithm (Hartley (1961)) are appropriate for continuously differentiable spline models.

The use of nonlinear least squares to estimate the parameters $\underline{\theta}$ and $\underline{\xi}$ of the spline model is facilitated by observing that these parameters form two disjoint sets. Thus splines are a special case of partially linear models whose linear and nonlinear parameters separate in the sense that, given a value for $\underline{\xi}$, $\underline{\theta}$ may be estimated through linear regression. To see the benefits which stem from this fact first define

$$X(\underline{\xi}) = [T \mid Q_\nu(\underline{\xi})] \qquad (4.6)$$

where T is given in (3.10) and $Q_\nu(\underline{\xi})$ is a $n \times (d-m)$ matrix with typical element $(t_i - \xi_j)_+^r$, $i=1,\ldots,n$ $j=1,\ldots,\kappa$ and $r=\nu_j,\ldots,m-1$.

Then, for any $\underline{\xi} \in D_k$, it is well known that the solution, $\hat{\underline{\theta}}(\underline{\xi})$ say, to

$$X(\underline{\xi})'X(\underline{\xi})\underline{\theta} = X(\underline{\xi})'\underline{y} \qquad (4.7)$$

minimizes $SSE(\underline{\theta},\underline{\xi})$ with respect to $\underline{\theta}$. This suggests replacing $\underline{\theta}$ by $\hat{\underline{\theta}}(\underline{\xi})$ in $e_i(\underline{\theta},\underline{\xi})$ and minimizing

$$SSE(\underline{\xi}) = \sum_{j=1}^{n} e_j(\hat{\underline{\theta}}(\underline{\xi}),\underline{\xi})^2 \qquad (4.8)$$

instead of (4.4). The function in (4.8) depends only on the knots and its minimization as a function of $\underline{\xi}$ is equivalent to iterated minimization of (4.4), i.e., $\min_{\underline{\xi} \in D_k} SSE(\underline{\xi}) = \min_{\underline{\xi} \in D_k} \min_{\underline{\theta} \in \mathbb{R}^d} SSE(\underline{\theta},\underline{\xi})$. The benefits from using this iterated approach for fitting splines with free knots are far from trivial. Not only does it provide an effective reduction in the number of parameters under consideration but, more importantly, reasonable starting values for the ξ_j's, which are required to initialize most nonlinear optimization algorithms, can usually be obtained from visual inspection of the data. An initial value for the $\underline{\theta}$ vector is not so easily obtained, however. It is, therefore, comforting to note that, under certain restrictions, Golub and Pereyra (1973) have shown that minimizing (4.8) is equivalent to minimizing (4.4). For their results to apply we must assume that s is continuously differentiable and that the rank of $X(\underline{\xi})$ is constant over the knot variations we employ. Jupp (1978) provides reasoning for why the latter assumption should usually be satisfied.

As noted by Jupp (1978), algorithms for minimizing $SSE(\underline{\xi})$ exhibit a lethargy property in that they tend to become trapped near the boundary of D_k and converge (numerically) quite far from the boundary solution. Solutions on the boundary of D_k correspond to a reduction in the stated continuity constraints and, as a result, we will generally prefer an interior minimum. These facts led Jupp (1978) to propose reparameterizing the problem in terms of the new variables

$$\omega_i = \ln\left(\frac{\xi_{i+1} - \xi_i}{\xi_i - \xi_{i-1}}\right) \quad, \quad i=1,\ldots,k \quad . \tag{4.9}$$

The transformation (4.9) maps D_k smoothly onto $\mathring{R}^k$ and takes the boundaries to $\pm\infty$. Jupp (1972, 1978) has used (4.9) effectively to fit a cubic (m=4) spline with five knots and $\nu_j = 3$, j=1,...,5 to the titanium heat data. For various starting values, his approach led to the interior minimum more frequently than an algorithm developed by deBoor and Rice (1968a,b), available on the IMSL package, to which it was compared.

Jupp's spline fit to the titanium heat data provides an illustration of the gains to be realized by optimizing knot locations. To demonstrate this point both the cubic spline with five optimal knots and a cubic spline with five knots uniformly spaced over the interior of [595,1075] have been plotted with the data in Figure 6.

Jupp (1978) established a relationship between the Jacobian matrices for SSE($\underline{\xi}$) and SSE($\underline{\omega}$), the SSE function using the transformed variables (4.9). Specifically he showed that the Jacobian matrices, $J(\underline{\xi})$ and $J(\underline{\omega})$, for these two functions satisfy
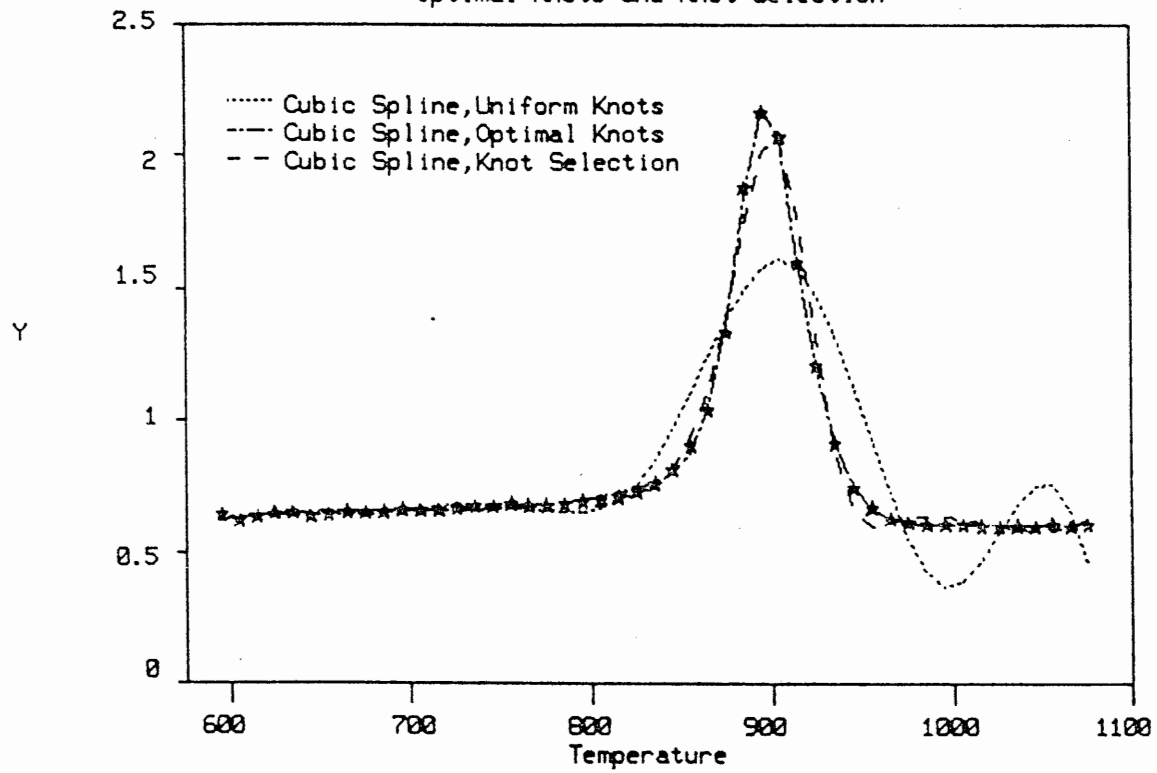
$$J(\underline{\omega}) = J(\underline{\xi}) \; G(\underline{\xi})^{-1} \tag{4.10}$$

where $G(\underline{\xi})$ is a $k \times k$ tridiagonal matrix having nonzero elements $g_{j,j-1} = (\xi_j - \xi_{j-1})^{-1}$, $g_{j,j} = -(\xi_{j+1}-\xi_j)^{-1} - (\xi_j-\xi_{j-1})^{-1}$ and $g_{j,j+1} = (\xi_{j+1} - \xi_j)^{-1}$. By using a local linearization it was then shown that the vectors of corrections (or changes), $\underline{\delta\omega}$ and $\underline{\delta\xi}$, which are made to the ω_j's and ξ_j's at each iteration of the optimization process are related by

$$\underline{\delta\xi} \doteq G(\underline{\xi})^{-1} \; \underline{\delta\omega} . \tag{4.11}$$

This has the consequence that a local approximation to the solution path to optimal knots corresponding to transformation (4.9) can be obtained using only the variables of interest, i.e., the ξ_j's, for many nonlinear least squares algorithms.

FIGURE 6. Titanium Heat Data Fitted Using Uniform Knots, Optimal Knots and Knot Selection

For example, if we apply the Levenberg-Marquardt algorithm (Marquardt (1963)) to $SSE(\underline{\omega})$ then at each iteration we make a change of the form $\underline{\delta\omega} = (J(\underline{\omega})'J(\underline{\omega}) + \lambda I)^{-1}J(\underline{\omega})'\underline{e}$ where $\underline{e}$ is the vector of residuals and $\lambda > 0$. Using (4.10) and (4.11), the corresponding approximate change in the knots is

$$\underline{\delta\xi} = (J(\underline{\xi})'J(\underline{\xi}) + \lambda G(\underline{\xi})^2)^{-1}J(\underline{\xi})'\underline{e} \qquad (4.12)$$

which does not involve $\underline{\omega}$.

It is of interest to compare (4.12) to the correction term, $(J(\underline{\xi})'J(\underline{\xi}) + \lambda I)^{-1}J(\underline{\xi})'\underline{e}$, one obtains by applying the Levenberg-Marquardt algorithm directly to the ξ_j's. It is well known (Marquardt (1963)) that this latter correction solves the problem

$$\min_{\underline{d}\in\mathbb{R}^k} \{||\underline{e} - J(\underline{\xi})\underline{d}||_E^2 + \lambda||\underline{d}||_E^2\} ,$$

with $||\cdot||_E$ denoting the usual Euclidean norm, whereas (4.12) solves an analogous problem except with $||\underline{d}||_E$ replaced by $||G(\underline{\xi})\underline{d}||$. Examination of $G(\underline{\xi})\underline{d}$ reveals that this is a vector of second differences and, hence, (4.12) represents a smoothing of the solution path for the Levenberg-Marquardt procedure. This, of course, suggests the possibility of replacing $G(\underline{\xi})$ in (4.12) with alternatives which provide other types of smoothing. Such possibilities have apparently not, as yet, been investigated.

Under the assumption of normal errors we may also set confidence intervals and test hypotheses about the parameters in a spline model. It is, in fact, possible to obtain exact tests and confidence regions in certain cases using a method essentially due to Halperin (1962) and Hartley (1964). Their approach is applicable to any nonlinear model which, under a null hypothesis of interest, is linear in the unspecified parameters. For example, in a spline model if we wish to test $H_0 : \underline{\xi} = \underline{\xi}_0$ for some specified vector $\underline{\xi}_0$, the regression model under H_0 is, in matrix notation,

$$\underline{y} = X(\underline{\xi}_0)\underline{\theta} + \underline{\varepsilon} \qquad\qquad (4.13)$$

which is linear in the unspecified parameter vector $\underline{\theta}$. Consequently, by fitting the model

$$\underline{y} = X(\underline{\xi}_0)\underline{\theta} + Z(\underline{\xi}_0)\underline{\delta} + \underline{\varepsilon} \quad , \qquad\qquad (4.14)$$

where $Z(\underline{\xi})$ is an arbitrary matrix that is possibly dependent on $\underline{\xi}$ and chosen subject to rank considerations, a test of H_0 can be obtained through the standard F-test for $H_0^* : \underline{\delta} = 0$. Gallant (1974, 1977) discusses how the elements of $Z(\underline{\xi})$ should be selected to provide good power for this test and notes that, by finding all the knot sets for which H_0^* is not rejected, a confidence region for $\underline{\xi}$ can be obtained. El-Shaarawi and Shah (1980) make the specific choice

$$Z(\underline{\xi}) = - \left\{ \frac{\partial e_i(\underline{\theta},\underline{\xi})}{\partial \xi_j} \right\}_{\substack{i=1,n \\ j=1,k}}$$

and extend the work of Halperin (1962) to obtain statistics which, for spline models, can be utilized to test hypotheses of the form $H_0 : \underline{\xi} = \underline{\xi}_0$ or $H_0 : \underline{\xi} = \underline{\xi}_0, \underline{\theta} = \underline{\theta}_0$. By inverting these test statistics one can also, in theory, obtain confidence regions for the knots or for the knots and coefficient vector.

The asymptotic distribution theory for least squares parameter estimates from spline models has been studied by Gallant (1974) and Feder (1975a,b). In particular, it follows from Gallant (1974) that, for a continuously differentiable spline model, if we let $J(\underline{\theta},\underline{\xi})$ and $(\hat{\underline{\theta}},\hat{\underline{\xi}})$ denote, respectively, the Jacobian matrix and minimizer of the unreduced functional (4.4) then, for large n, $(\hat{\underline{\theta}},\hat{\underline{\xi}})$ may be treated as having a $d+k$ dimensional normal distribution with mean $(\underline{\theta},\underline{\xi})$ and a variance-covariance matrix which is consistently estimated by

$$C = \frac{SSE(\hat{\underline{\theta}},\hat{\underline{\xi}})}{n} (J(\hat{\underline{\theta}},\hat{\underline{\xi}})'J(\hat{\underline{\theta}},\hat{\underline{\xi}}))^{-1} .$$

This fact motivates the use of intervals such as $\hat{\xi}_i \pm Z_{\alpha/2}\sqrt{c_{m+i,m+i}}$ where c_{ij} is the ijth element of C, to provide approximate $100(1-\alpha)\%$ intervals for the knots. The linear parameters can, of course, be dealt with similarly. It is also possible to use SSE ratios to obtain approximate tests for certain hypotheses about the spline model. Unfortunately, tests which pertain to the absence of a knot or a continuity constraint violate certain regularity conditions required by Gallant (1974) in deriving their asymptotic distribution theory. To overcome such difficulties Gallant (1977) developed an asymptotic analog of the exact tests we considered previously that can be utilized for these types of hypotheses.

It is sometimes of interest to estimate both the number and placement of knots. To obtain an estimate of k one could, for example, use statistics such as (2.9) and (2.10) or, perhaps, a C_p type statistic such as utilized by Ertel and Fowlkes (1976). The use of their statistic, however, requires an estimate of σ^2 which they obtain by overfitting. As GCV can be regarded as a competitor to C_p which does not require estimation of σ (see Golub, Heath and Wahba (1979)) one could also estimate k by the minimizer of

$$GCV(k) = n^{-1}SSE(\hat{\underline{\xi}}_k)/[n^{-1}tr(I-H(\hat{\underline{\xi}}_k))]^2$$

where $\hat{\underline{\xi}}_k$ is the vector of knot estimates for the k knot model and $H(\hat{\underline{\xi}}_k) = X(\hat{\underline{\xi}}_k)(X(\hat{\underline{\xi}}_k)'X(\hat{\underline{\xi}}_k))^{-1}X(\hat{\underline{\xi}}_k)'$. In view of the asymptotic relationship between GCV and Akaike's information criterion (AIC) (see Golub, Heath and Wahba (1979)) this approach is closely related to an AIC based procedure used by Brannigan (1981) to select k, as well as $\underline{\xi}$, which appears quite successful. As another alternative, cluster analysis methodology, such as employed by McGee and Carleton (1970), could be used. Other algorithms which allow for variable k have been proposed by Powell (1970) and Smith and Smith (1979).

An asymptotic solution to the problem of optimally selecting the number and positioning of the knots, as well as the optimal placement of observation points, has been provided by Agarwal (1978) and Agarwal and Studden (1978a, 1978b, 1980). They also develop an adaptive algorithm to accomplish optimization with respect to these three variables which can be used when the process under study can be sampled repeatedly. It also follows from their work that the asymptotically optimal IMSE for variable knot splines of order m is $0(n^{-2m/2m+1})$. By comparing this to similar results for smoothing splines in the previous section, we observe that to obtain this rate with smoothing splines one must either use a spline of order 2m or handle the boundary terms separately (see Speckman (1981b)). Consequently, free knot splines can be considered as, asymptotically, more parsimonious than smoothing splines.

Essentially any algorithm which allows the knots to be free variables will require considerable computation time. An interesting alternative to such procedures has been proposed by Smith (1983) (see also Marsh (1983)). She chooses a pool of k' > k knots, distributed appropriately over [a,b] and then uses variable selection techniques to delete knots from consideration whose corresponding coefficients do not statistically differ from zero. The resulting fitted model is usually obtained with much less labor and, provided the knot pool is chosen well, often provides a fit that is competitive with those obtained via nonlinear regression. An illustration of typical results obtained from her method is provided by Figure 6 where a cubic spline has been fitted to the titanium heat data using a knot pool consisting of 10 equally spaced values over [785,1055] and step-down selection. It is interesting to note that after nonsignificant terms are deleted the final fit has five knots.

ACKNOWLEDGMENTS

REFERENCES

Agarwal, Girdhar, G. (1978). Asymptotic design and estimation in polynomial spline regression. Department of Statistics Mimeograph Series No. 78-11, Purdue University.

Agarwal, Girdhar G. and Studden, W. J. (1978a). An algorithm for selection of design and knots in the response curve estimation by spline functions. Department of Statistics Mimeograph Series No. 78-15, Purdue University.

Agarwal, Girdhar G. and Studden, W. J. (1978b). Asymptotic design and estimation using linear splines. *Comm. Statist.-Simula. Computa.* B7(4), 309-319.

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

Anderssen, R. S., Bloomfield, P. and McNeil, D. R. (1974). Spline functions in data analysis. Tech. Rep. 69, Series 2, Department of Statistics, Princeton University.

Anselone, P. M. and Laurent, P. J. (1968). A general method for the construction of interpolating or smoothing spline-functions. *Numer. Math.* 12, 66-82.

Bacon, David W. and Watts, Donald G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* 58, 525-534.

Barth, James, Kraft, Arthur and Kraft, John (1976). Estimation of the liquidity trap using spline functions. *Rev. Econ. Statist.* 58, 218-222.

Bellman, Richard and Roth, Robert (1969). Curve fitting by
    segmented straight lines. *J. Amer. Statist. Assoc. 64*,
    1079-1084.

Blight, B.J.N. and Ott, L. (1975). A Bayesian approach to
    model inadequacy for polynomial regression. *Biometrika
    62*, 79-88.

de Boor, Carl (1978). A *Practical Guide to Splines*. Springer-
    Verlag: New York.

de Boor, Carl and Rice, John R. (1968a). Least squares cubic
    spline approximation I - fixed knots. Computer Science
    Department Tech. Report No. 20, Purdue University.

de Boor, Carl and Rice, John R. (1968b). Least squares cubic
    spline approximation II - variable knots. Computer Science
    Dept. Tech. Report No. 21, Purdue University.

Brannigan, M. (1981). An adaptive piecewise polynomial curve
    fitting procedure for data analysis. *Comm. Statist.-Theor.
    Meth. A10(18)*,1823-1848.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques
    for testing the constancy of regression relationships
    over time. *J. Roy. Statist. Soc. Ser. B 37*, 149-192.

Buse, A. and Lim, L. (1977). Cubic splines as a special case
    of restricted least squares. *J. Amer. Statist. Assoc.
    72*, 64-68.

Chow, Gregory C. (1960). Tests of equality between sets of
    coefficients in two linear regressions. *Econometrica 28*,
    591-605.

Cox, M. G. (1971). Curve fitting with piecewise polynomials.
    J. Inst. Maths. Applics. 8, 36-52.

Cox, Dennis D. (1981). Asymptotics for M-type smoothing splines.
    Ann. Statist. 11, 530-551.

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

Demmler, A. and Reinsch, C. (1975). Oscillation matrices with
    spline smoothing. *Numer. Math. 24*, 375-382.

Draper, Norman R., Guttman, Irwin and Lipow, Peter (1974a).
All-bias designs for spline functions, Part I. MRC
Technical Summary Report No. 1434, University of
Wisconsin-Madison.

Draper, Norman R., Guttman, Irwin and Lipow, Peter (1974b).
All-bias designs for spline functions, Part II. MRC
Technical Summary Report No. 1441. University of
Wisconsin-Madison.

Draper, N. R., Guttman, Irwin and Lipow, Peter (1977). All-bias
designs for spline functions joined at the axes. *J. Amer.
Statist. Assoc.* 72, 424-429.

Dyn, N., Wahba G., and Wong, W. H. (1979). Comments to "Smooth
pycnophylactic interpolation for geographical regions"
by W. Tobler. *J. Amer. Statist. Assoc.* 74, 530-535.

El-Shaarawi, A. and Shah, K. R. (1980). Interval estimation in
nonlinear models. *Sankhyā Ser. B* 42, 229-232.

Ertel, James E. and Fowlkes, Edward B. (1976). Some algorithms
for linear spline and piecewise multiple linear regression.
J. Amer. Statist. Assoc. 71, 640-648.

Eubank, R. L. (1983). The hat matrix for smoothing splines.
Statistics and Probability Letters, to appear.

Eubank, R. L., Smith, P. L. and Smith, P. W. (1982). A note on
optimal and asymptotically optimal designs for certain time
series models. *Ann. Statist.* 10, 1295-1301.

Feder, Paul I. (1975a). On asymptotic distribution theory in
segmented regression problems - identified case. *Ann.
Statist.* 3, 49-83.

Feder, Paul I. (1975b). The log likelihood ratio in segmented
regression. *Ann. Statist.* 3, 84-97

Fuller, W. A. (1969). Grafted polynomials as approximating
functions. *Australian J. of Agr. Econ.* 13, 35-46.

Fuller, W. A. (1976). *Introduction to Statistical Time Series.*
John Wiley: New York.

Gallant, A. Ronald (1974). The theory of nonlinear regression as
it relates to segmented polynomial regressions with estimated
join points. Institute of Statistics Mimeograph Series No.
925, North Carolina State Univ.

Gallant, A. Ronald (1975). Nonlinear regression. *Amer. Statist.*
29, 73-81.

Gallant, A. Ronald (1977). Testing a nonlinear regression specification: a nonregular case. *J. Amer. Statist. Assoc.* 72, 523-530.

Gallant, A. R. and Fuller, Wayne A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *J. Amer. Statist. Assoc.* 68, 144-147.

Gamber, Heather A. (1979a). Choice of an optimal shape parameter when smoothing noisy data. *Comm. Statist.-Theor. Meth.* A8(14), 1425-1435.

Gamber, Heather A. (1979b). Confidence regions for periodic functions. *Comm. Statist.-Theor. Meth.* A8(14), 1437-1446.

Gbur, E. E. and Dahm, P. F. (1982). Estimation of segmented regression models in the presence of measurement error. Manuscript, Texas A&M University.

Gbur, E. E., Thomas, G. L. and Miller, F. R. (1979). The use of segmented regression models in the determination of the base temperature in heat accumulation models. *Agronomy J.* 71, 949-953.

Gerig, Thomas M. and Gallant, A. Ronald (1975). Computing methods for linear models subject to linear parametric constraints. *J. Statist. Comput. Simul.* 3, 283-296.

Goldberger, Arthur S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.* 57, 369-375.

Golub, Gene H., Heath, Michael and Wahba, Grace (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21, 215-223.

Golub, G. H. and Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.* 10, 413-432.

Gunst, Richard F. and Mason, Robert L. (1980). *Regression Analysis and Its Applications: A Data-Oriented Approach.* Marcel Dekker: New York.

Guthery, Scott B. (1974). Partition regression. *J. Amer. Statist. Assoc.* 69, 945-947.

Hartley, H. O. (1961). The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. *Technometrics* 3, 269-280.

Hartley, H. O. (1964). Exact confidence regions for the parameters in non-linear regression laws. *Biometrika* 51, 347-353.

Halperin, Max (1962). Confidence interval estimation in non-linear regression. *J. Roy. Statist. Soc. Ser. B*, 25, 330-333.

Hinkley, D. V. (1969). Inference about intersection in two-phase regression. *Biometrika* 56, 495-504.

Hinkley, David V. (1971). Inference in two-phase regression. *J. Amer. Statist. Assoc.* 66, 736-743.

Hoaglin, D. C. and Welsch, R. F. (1978). The hat matrix in regression and ANOVA. *Amer. Statist* 32, 17-22. Corrigenda 32, 146.

Holt, J. N. and Jupp, D.L.B. (1978). Free-knot spline inversion of a Fredholm integral equation from astrophysics. *J. Inst. Math. Applics.* 21, 429-443.

Huber, Peter J. (1979). Robust smoothing. In *Robustness in Statistics* (Launer and Wilkinson, eds.), 33-47. Academic Press: New York.

Hudson, Derek, J. (1966). Fitting segmented curves whose join points have to be estimated. *J. Amer. Statist. Assoc.* 61, 1097-1129.

Jupp, D.L.B. (1972). Curve fitting by splines as an application of unconstrained optimization. In *Optimization* (R. S. Anderssen, L. S. Jennings and D. M. Ryan, eds.) 49-59. Univ. of Queensland Press: Brisbane.

Jupp, David L.B. (1978). Approximation to data by splines with free knots. *SIAM J. Numer. Anal.* 15, 328-343.

Jupp, D.L.B. and Vozoff, K. (1975). Stable iterative methods for the inversion of geophysical data. *Geophys. J. Roy. Astronom. Soc.* 42, 957-976.

Kimeldorf, George S. and Wahba, Grace (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* 41, 495-502.

Kimeldorf, George S. and Wahba, Grace (1970b). Spline functions and stochastic processes. *Sankhya Ser.* A 32, 173-180.

Kimeldorf, George and Wahba, Grace (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl. 33,* 82-95.

Lechner, James A., Reeve, Charles P. and Spiegelman, Clifford H. (1982). An implementation of the Scheffé approach to calibration using spline functions, illustrated by a pressure-volume calibration. *Technometrics 24,* 229-234.

Lenth, Russell V. (1977). Robust splines. *Comm. Statist.-Theor. Meth.* A6(9), 847-854.

Marquardt, Donald W. (1963). An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math. 11,* 431-441.

Marsh, L. C. (1983). On estimating spline regressions. Proc. 8th Annual SAS Users Group International Conference.

McGee, Victor E. and Carleton, Willard T. (1970). Piecewise regression. *J. Amer. Statist. Assoc. 65,* 1109-1124.

Murty, V. N. (1971). Optimal designs with a polynomial spline regression with a single multiple knot at the center. *Ann. Math. Statist. 42,* 952-960.

Nougues, A. and Sielken, R. L. Jr. (1980). A comparative study of spline regression. Institute of Statistics Tech. Report No. 67, Texas A&M Univ.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *J. Roy. Statist. Soc. Ser.* B 40, 1-42.

Park, Sung H. (1978). Experimental designs for fitting segmented polynomial regression models. *Technometrics 20,* 151-154.

Peele, Lawrence and Kimeldorf, George (1977). Prediction functions and mean estimation functions for a time series. *Ann. Statist.* 5, 709-721.

Peele, Lawrence and Kimeldorf, George (1979). Time series prediction functions based on imprecise observations. *Ann. Statist.* 7, 801-811.

Poirier, Dale J. (1973). Piecewise regression using cubic splines. *J. Amer. Statist. Assoc. 68,* 515-524.

Poirier, Dale J. (1975). On the use of Cobb-Douglas splines.
    Internat. Econom. Rev. 16, 733-744.

Powell M.J.D. (1970). Curve fitting by splines in one variable.
    In *Numerical Approximation to Functions and Data* (J.G. Hayes,
    ed.), 65-83. Athlone Press: London.

Quandt, Richard E. (1958). The estimation of the parameters of a
    linear regression system obeying two separate regimes.
    J. Amer. Statist. Assoc. 53, 873-880.

Quandt, Richard E. (1960). Tests of the hypothesis that a linear
    regression system obeys two separate regimes. *J. Amer.
    Statist. Assoc. 55*, 324-330.

Reinsch, Christian H. (1967). Smoothing by spline functions.
    Numer. Math. 10, 177-183.

Reinsch, Christian H. (1971). Smoothing by spline functions, II.
    Numer. Math. 16, 451-454.

Rice, John and Rosenblatt, Murray (1981). Integrated mean square
    error of a smoothing spline. *J. Approx. Theor. 33*, 353-369.

Rice, John and Rosenblatt, Murray (1983). Smoothing splines:
    regression, derivatives and deconvolution. *Ann. Statist.
    11*, 141-156.

Robinson, D. E. (1964). Estimates for the points of intersection
    of two polynomial regressions. *J. Amer. Statist. Assoc. 59*,
    214-224.

Royden, H. L. (1968). *Real Analysis*. MacMillan Publishing Co.:
    New York.

Sacks, J. and Ylvisaker, D. (1970). Designs for regression pro-
    blems with correlated errors, III. *Ann. Math. Statist.
    41*, 2057-2074.

Sacks, J. and Ylvisaker, D. (1978). Linear estimation for
    approximately linear models. *Ann. Statist. 6*, 1122-1137.

Schoenberg, I. J. (1964). Spline functions and the problem of
    graduation. *Proc. Nat. Acad. Sci. USA 52*, 947-950.

Smith, A.F.M. and Cook, D. G. (1980). Straight lines with a
    change point: a Bayesian analysis of some renal trans-
    plant data. *Appl. Statist. 29*, 180-189.

Smith, Patricia L. (1979). Splines as a useful and convenient
    statistical tool. *Amer. Statist. 33*, 57-62.

Smith, Patricia L. (1982). Hypothesis testing in B-spline regression. *Comm. Statist.-Simula. Computa.* B11(2), 143-157.

Smith, Patricia L. (1983). Curve fitting and modeling with splines using statistical variable selection techniques. Preprint.

Smith, Patricia L. and Klein, Vladislav (1982). The selection of knots in polynomial splines using stepwise regression. Preprint.

Smith, Patricia L. and Smith, Philip W. (1979). Empirical stopping rules for data fitting by splines. *Proc. 1979 Army Numerical Analysis and Computers Conference*, 449-456.

Speckman, Paul (1979). On minimax estimation of linear operators in Hilbert spaces from noisy data. Preprint.

Speckman, Paul (1981a). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.*, to appear.

Speckman, Paul (1981b). The asymptotic integrated error for smoothing noisy data by splines. *Numer. Math.*, to appear.

Speckman, Paul (1982a). Efficient nonparametric regression with cross-validated smoothing splines. Preprint.

Speckman, Paul (1982b). L_p approximation of autoregressive Gaussian processes. *J. Multivariate Anal.*, to appear.

Speckman, Paul (1982c). Minimax estimates of linear functionals in a Hilbert space. *Ann. Statist.*, to appear.

Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics* 17, 634-645.

Steinberg, David M. (1983). Bayesian models for response surfaces of uncertain functional form. MRC Technical Summary Report No. 2474, Univ. of Wisconsin-Madison.

Stone, Charles J. (1977). Consistent nonparametric regression. *Ann. Statist.* 5, 595-645.

Studden, W. and Van Arman, D. (1969). Admissible designs for polynomial spline regression. *Ann. Math. Statist.* 40, 1557-1569.

Utreras, F. (1979). Cross-validation techniques for smoothing spline functions in one or two dimensions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.), 196-232. Lecture notes in Mathematics No. 757. Springer: New York.

Utreras, F. (1980). A package for smoothing noisy data with splines. Department of Mathematics Tech. Report No. MA-80-B-209, Univ. of Chile.

Utreras, F. (1981a). On computing robust splines and applications. *Siam J. Sci. Stat. Comput. 2*, 153-163.

Utreras, F. (1981b). Optimal smoothing of noisy data using spline functions. *Siam J. Sci. Stat. Comput. 2*, 349-362.

Wahba, Grace (1971). On the regression design problem of Sacks and Ylvisaker. *Ann. Math. Statist. 42*, 1035-1053.

Wahba, Grace (1975). Smoothing noisy data with spline functions. *Numer. Math. 24*, 383-393.

Wahba, Grace (1977a). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (P. R. Krishnaiah, ed.) 507-523. North Holland: Amsterdam.

Wahba, Grace (1977b). Optimal smoothing of density estimates. In *Classification and Clustering* (J. Van Ryzin, ed.), 423-458. Academic Press: New York.

Wahba, Grace (1977c). Practical approximate solutions to linear operator equations when the data are noisy. *Siam J. Numer. Anal. 14*, 651-667.

Wahba, Grace (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, Grace (1979). Convergence rates of "Thin Plate" smoothing splines when the data are noisy. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.) 233-245. Lecture Notes in Mathematics No. 757. Springer: New York.

Wahba, Grace (1980). Ill posed problems: numerical and statistical methods for mildy, moderately and severely ill posed problems with noisy data. In *Proceedings of the International Conference on Ill Posed Problems*. (M.Z. Nashed, ed.) to appear.

Wahba, Grace (1981a). Bayesian confidence intervals for the cross validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, to appear.

Wahba, Grace (1981b). Numerical experiments with the thin plate histospline. *Comm. Statist.-Theor. Meth.* A10(24), 2475-2514.

Wahba, Grace (1981c). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* 2, 5-16. Erratum, 3, 385-386.

Wahba, Grace (1982a). Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine. In *Statistical Decision Theory and Related Topics III, Vol. 2* (S. S. Gupta and J. O. Berger, eds.), 383-418. Academic Press: New York.

Wahba, Grace (1982b). Cross validated spline methods for direct and indirect sensing experiments. In *Proceeding of the Signal Processing in the Ocean Environment Workshop*, (E. J. Wegman and J. Smith, eds.) Marcel Dekker: New York.

Wahba, Grace and Wendelberger, James (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review 108*, 1122-1143.

Wahba, G. and Wold, S. (1975a). A completely automatic French curve: fitting spline functions by cross validation. *Comm. Statist.-Theor. Meth.* A4(1), 1-17.

Wahba, G. and Wold, S. (1975b). Periodic splines for spectral density estimation: the use of cross validation for determining the degree of smoothing. *Comm. Statist.-Theor. Meth.* A4(2), 125-141.

Wecker, W. E. and Ansley, C. F. (1981). Extensions and examples of the signal extraction approach to regression. To appear in *Applied Time Series Analysis of Economic Data* (A. Zellner, ed.).

Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Amer. Statist. Assoc.* 78, 81-89.

Wegman, Edward J. (1982). Optimal nonparametric function estimation. Preprint.

Wegman, Edward J. and Wright, Ian W. (1983). Splines in statistics. *J. Amer. Statist. Assoc.* 78, 351-365.

Weinert, Howard L. (1978). Statistical methods in optimal curve fitting. *Comm. Statist.-Simula. Computa.* B7(4), 417-435.

Weinert, Howard L. and Kailath, Thomas (1974). Stochastic interpretations and recursive algorithms for spline functions. *Ann. Statist.* 2, 787-794.

Weinert, Howard L. and Sidhu, Gursharan, S. (1978). A stochastic framework for recursive computation of spline functions: part I, interpolating splines. *IEEE Trans. Inform. Theory* IT-24, 45-50.

Weinert, H. L., Desai, U. B. and Sidhu, G. S. (1979). ARMA splines, system inverses, and least-squares estimates. *SIAM J. Control Optim.* 17, 525-536.

Weinert, H. L., Byrd, R. H. and Sidhu, G. S. (1980). A stochastic framework for recursive computation of spline functions: part II, smoothing splines. *J. Optim. Theory Appl.* 30, 255-268.

Wendelberger, James G. (1981). The computation of Laplacian smoothing splines with examples. Tech. Report No. 648, Dept. of Statistics, University of Wisconsin-Madison.

Wendelberger, James G. (1982). Estimation of divergence and vorticity using multidimensional smoothing splines. In *Proceedings of the Workshop on Density Estimation and Smoothing* (E. Parzen, ed.). To appear.

Whittaker, E. T. (1923). On a new method of graduation. *Proc. Edinburg Math. Soc.* 41, 63-75.

Whitten, W. J. (1971). The use of multi-dimensional cubic spline functions for regression and smoothing. *Austral. Comput. J.* 3, 81-88.

Whitten, W. J. (1972). The use of periodic spline functions for regression and smoothing. *Austral. Comput. J.* 4, 31-34.

Wold, S. (1971). Analysis of kinetic data by means of spline functions. *Chem. Scripta* 1, 97-102.

Wold, S. (1974). Spline functions in data analysis. *Technometrics* 16, 1-11.

Wong, Wing Hung (1981). On constrained splines and their ap-
     proximations. Tech. Report No. 127, Dept. Statist.,
     University of Chicago .

Worsley, K. J. (1983). Testing for a two-phase multiple
     regression. *Technometrics* 25, 35-42.

Wright, Ian W. and Wegman, Edward J. (1980). Isotonic, convex
     and related splines. *Ann. Statist.* 8, 1023-1035.

Young, A. S. (1977). A Bayesian approach to prediction using
     polynomials. *Biometrika* 64, 309-317.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SMU/DS/TR-180	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* APPROXIMATE REGRESSION MODELS AND SPLINES		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER SMU/DS/TR-180
7. AUTHOR(s) R. L. Eubank		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0207
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-479
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE August 1983
		13. NUMBER OF PAGES 56
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)*
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any putpose of The United States Government.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purposes of the United States Government.

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Bayesian estimation; best linear unbiased predictor; cross-validation; fixed knots; nonlinear regression; polynomial regression; smoothing splines; variable knots

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The literature pertaining to splines in regression analysis is reviewed. Spline regression is motivated as a simple extension of the basic polynomial regression model. Using this framework, the concepts of fixed and variable knot spline regression are developed and corresponding inferential procedures are considered. Smoothing splines are also seen to be an extension of poly-nomial regression and various optimality properties, as well as inferential and diagnostic methods, for these types of splines are discussed.

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