An Algorithm for Selection of Design and Knots in the Response Curve Estimation by Spline Functions*

by

Girdhar G. Agarwal

and

W. J. Studden

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-15

July 1978

*This research was supproted in part by NSF Grant No. MCS76-08235 A01.

I. INTRODUCTION

Suppose a functional relationship

$$\eta = g(x)$$

exists between a response η and an independent (or controllable) variable x, where x lies in the interval $[0,1]$. The problem to be considered is to estimate g using n measurements of η. At each x_i, $i=1,\ldots,r$, $n_i = n\mu_i$ measurements are taken. The probability measure assigning mass μ_i to the point x_i $(\Sigma\mu_i = 1)$ is referred to as the design. In observing the response η we assume that an additive experimental error, denoted by ε, exists so that, for each observation y_{ij}, $j=1,\ldots,n_i$, $i=1,\ldots,r$; we can write

$$y_{ij} = \eta(x_i) + \varepsilon_{ij} = g(x_i) + \varepsilon_{ij}.$$

We assume that ε_{ij} are uncorrelated and identically distributed with mean zero and an unknown common variance σ^2 independent of x.

If it is known that the true functional relationship $\eta = g(x)$ has a certain form depending on a few parameters, then the problem is usually to estimate these parameters. If the form of the true functional relationship is unknown, the problem is to approximate the function $g(x)$ by some graduating function. In this paper we are interested in the latter problem. In the absence of the knowledge of the true functional relationship, it has been a common practice to use a polynomial as an approximating function. But when the degree of polynomials is high, a number of unpleasant features begin to appear, one of which is the high oscillatory behavior of the approximating polynomial. Spline functions (for definition etc. see Greville 1969) are considerably less oscillatory. As an example see Jupp (1978) where he has fitted a polynomial of degree 9 as well as a cubic spline to the data of world sugar prices over a 31 year period (Guest 1961, p.194). The improvement in the fit to the data achieved by cubic splines is somewhat obvious since it shows less oscillation compared to polynomial fit. Furthermore, the behavior of a polynomial in an arbitrarily small region defines, through the concept of analytic continuity, its behavior everywhere. On the other hand, the spline functions possess the property of having local behavior that is less dependent on their behavior elsewhere. Because of these properties spline functions are more and more being used in the exploration of response curves for physical processes. Low order splines are commonly used in geophysics in the form of layered earth models (for example, see Vozoff and Jupp 1975 and its list of references or Jupp and Stewart 1974). In astrophysics, Holt (1974) has used piecewise linear splines to model the radiation profile of the sun's atmosphere. Wold (1971,1974) has used the spline functions in the analysis of response curves in pharmacokinetics.

Here the function $g(x)$ will be approximated by a spline function $s(x)$ of degree d. The function $s(x)$ has the representation

$$s(x) = \sum_{i=0}^{d} \theta_i x^i + \sum_{i=1}^{k} \theta_{d+i} (x-\xi_i)_+^d \qquad (1.1)$$

where $(x - \xi_i)_+ = (x - \xi_i)$ if $x > \xi_i$ and zero otherwise. The points ξ_i are called knots and we assume that $\xi_0 = 0 < \xi_1 < \ldots\ldots < \xi_k < \xi_{k+1} = 1$. The function s(x) has generally d-1 continuous derivatives at each ξ_i. Lower order differentiability can be assumed by introducing terms of the form $(x - \xi_i)_+^j$, j=d-1,d-2,...,0.

Let $\bar{y}_i$ denote the average of n_i observations taken at x_i. Estimates, which are linear in $\bar{y}' = (\bar{y}_1,\ldots,\bar{y}_r)$, will be used to estimate the vector of parameters $\theta' = (\theta_0,\theta_1,\ldots,\theta_{k+d})$. As our criterion for the goodness of estimate we shall use an integrated mean square error (IMSE); the integration being taken with respect to Lebesgue measure on [0,1]. Our estimate of the response is

$$\hat{g}(x) = \hat{\theta}' f(x),$$

where $f'(x) = (1,x,\ldots,x^d,(x-\xi_1)_+^d,\ldots,(x-\xi_k)_+^d)$; and the IMSE is

$$J = \int_0^1 E(\hat{\theta}' f(x) - g(x))^2 dx$$

It is easy to show that

$$\begin{aligned} J &= V + B \\ &= \text{Tr}\, M_0\, \text{Var}(\hat{\theta}) + \int_0^1 (g(x) - f'(x)E(\hat{\theta}))^2 dx \qquad (1.2) \end{aligned}$$

where M_0 is the (k+d+1) × (k+d+1) matrix $\int f(x) f'(x)dx$. Note that V and B denote the integrated variance and integrated squared bias respectively. Using the different choices of estimator, we study in Section 2 the asymptotic behavior of IMSE for large n and k. The asymptotic expression for IMSE depends on three variables: (a) the design μ or allocation of observations; (b) the distribution of the knots $\xi_1 < \xi_2, \ldots, < \xi_k$ and (c) the number of knots. We propose to adaptively estimate the response function g(x) by $\hat{\theta}'f(x)$, attempting to minimize the IMSE with respect to the variables (a), (b) & (c). The minimization of the asymptotic expression for the IMSE is presented in Section 3.

To take full advantage of the benefits of the spline approach, the choice of the number and position of the knots is an important and difficult problem. In an attempt to resolve this problem as well as the problem of choice of design, we present in Section 4 an algorithm for positioning the knots and the allocation of the observation points. The algorithm which is based on the results of the minimization of the asymptotic value of IMSE, is defined in such a way that explicit knowledge of the response function g(x) is not required.

Finally in Section 5 an example is presented which shows the behavior of the algorithm.

The main idea for the approach used here is from (Dodson 1972; Rice 1969; and Burchard 1974) where non-statistical approaches were used. Further discussion of the results can be found in (Agarwal 1978; and Agarwal and Studden 1978). Proofs of theorems in Sections 2 and 3 for general splines will be provided elsewhere. The emphasis in the present paper is on the details of the algorithm in Section 4. Only the case d=1 is considered here. Computational and programming aspects of the algorithm will be written in a separate report.

2. ASYMPTOTIC VALUE FOR IMSE

In considering the asymptotic behavior for the IMSE, we will be concerned with the sequences $T_k = \{\xi_0, \xi_1, \ldots, \xi_k, \xi_{k+1}\}$ of knots defined by

$$\int_0^{\xi_i} p(x)dx = i/(k+1), \quad i=0,1,\ldots,k+1 \qquad (2.1)$$

where $p(x)$ is a positive continuous density on $[0,1]$. Note that $\xi_0=0$, $\xi_{k+1}=1$. Sacks and Ylvisaker (1970) call the sequence $\{T_k, k \geq 1\}$ so defined as a Regular Sequence generated by p (RS(p)). We will also assume that the design μ shows some regularity as k on n become large.

2.1 Variance Minimizing Estimate

The classical type regression theory assumes that the approximating spline function $s(x)$ provides a perfect representation of the response $g(x)$ and the fitted value $\hat{y}(x)$ is supposed to be an unbiased estimate of $g(x)$, the true value. Under this assumption B=0 in (1.2) and J, the IMSE consists of only the integrated variance term. If this is the case then J is minimized by the usual least square estimate (LSE),

$$\hat{\underset{\sim}{\theta}}_{LSE} = \underset{\sim}{M}^{-1}(\mu)\int \underset{\sim}{f}(x)\bar{y}_x d\mu(x). \qquad (2.2)$$

Here μ represents the design measure placing mass μ_i on x_i, $i=1,2,..r$; $\underset{\sim}{M}(\mu) = \int \underset{\sim}{f}(x)\underset{\sim}{f}'(x)d\mu(x)$ is the $(k+d+1) \times (k+d+1)$ information matrix; and $\bar{y}_x$ is the average of the observations at x.

Our first theorem concerns the asymptotic value for IMSE when $g(x)$ is being approximated by spline functions of degree d=0.

Theorem 2.1: Let the response function g(x) be continuously differentiable on [0,1] and the approximating spline function have degree zero. If the design measure μ has a continuous strictly positive density h(x), the LSE is used and $\{T_k\}$ is RS(p), then

$$J \approx \frac{k\sigma^2}{n} \int_0^1 \frac{p(x)}{h(x)}\, dx + \frac{1}{12k^2} \int_0^1 \frac{(g'(x))^2}{(p(x))^2} \quad . \qquad (2.3)$$

The first term on the right corresponds to asymptotic value for integrated variance and the second term corresponds to asymptotic value for integrated squared bias. Note the asymptotics found here are with respect to the number of knots k going to infinity. The number of observations n should be at least (k+1) and will usually be increasing much faster than k.

In the next theorem we find the asymptotic expression for the IMSE when function g(x) is being approximated by spline function of degree one.

Theorem 2.2: Let the response function g(x) be twice continuously differentiable on [0,1] and the approximating spline function be of degree one. If the design measure μ has a continuous strictly positive density h(x), the LSE is used and $\{T_k\}$ is RS(p), then

$$J \approx \frac{k\sigma^2}{n} \int_0^1 \frac{p(x)}{h(x)}\, dx + \frac{1}{720k^4} \int_0^1 \frac{(g''(x))^2}{(p(x))^4}\, dx \qquad (2.4)$$

For a proof of these two theorems see Agarwal (1978).

2.2 Bias Minimizing Estimator

Various authors, for example, Box and Draper (1959) and Karson Manson and Hader (1969) have proposed attaching more importance to the bias part B.

The integrated squared bias B is minimized if

$$E(\hat{\underset{\sim}{\theta}}) = \underset{\sim}{M}_0^{-1}\int \underset{\sim}{f}(x)g(x)dx.$$

Our "bias minimizing" estimator (BME) will then be of the form

$$\hat{\underset{\sim}{\theta}} = \underset{\sim}{M}_0^{-1}\int \underset{\sim}{f}(x)\hat{g}(x)dx.$$

where $\hat{g}(x)$ is some estimate of $g(x)$. In Section 4 and 5 (where the algorithm and example are discussed) we take $\hat{g}(x)$ to be the first degree spline which interpolates $g(x)$ at x_i, $i=1,\ldots,r$. Recall that the x_i's are points where observations are taken. Let $\underset{\sim}{L}'(x) = (L_1(x),\ldots,L_r(x))$, where $L_i(x)$, $i=1,\ldots,r$ are the linear spline functions such that $L_i(x_j) = \delta_{ij}$, $i,j=1,\ldots,r$. We can represent the interpolating function $\hat{g}(x)$ in terms of the 'roof or triangle-shaped function' L_i's as

$$\hat{g}(x) = \sum_{i=1}^{r} g(x_i)L_i(x).$$

A requiring property about $\hat{g}$ is that if g is continuous, then $\hat{g}$ converges to g pointwise as $\max_{2\le i\le r}(x_i - x_{i-1})$ goes to zero (e.g. see Prenter 1975). Therefore our BME will look like

$$\hat{\underset{\sim}{\theta}}_{BME} = \underset{\sim}{M}_0^{-1}\int_0^1 \underset{\sim}{f}(x)\underset{\sim}{L}'(x)\underset{\sim}{\bar{y}}dx, \qquad (2.5)$$

where $\underset{\sim}{\bar{y}}' = (\bar{y}_1,\ldots,\bar{y}_r)$.

In the next theorem, we find the asymptotic expression for IMSE when $g(x)$ is being approximated by spline function of degree one and the estimator used is $\hat{\theta}_{BME}$. This theorem concerns choosing the design μ to have weights μ_i on x_i such that

$$\mu_i = \int L_i(x)h(x)dx, \quad i=1,\ldots,r \tag{2.6}$$

for some continuous and positive density $h(x)$.

Theorem 2.3: Let the response function $g(x)$ be twice continuously differentiable and the approximating spline functions have degree one. If the design is chosen using (2.6) and the estimator $\hat{\theta}_{BME}$, given in (2.5), is used and $\{T_k\}$ is RS(p), then

$$J \approx \frac{k\sigma^2}{n} \int_0^1 \frac{p(x)}{h(x)}\, dx + \frac{1}{720}\, \frac{1}{k^4} \int_0^1 \frac{(g''(x))^2}{(p(x))^4}\, dx. \tag{2.7}$$

In the above we have suggested one choice for BME, but we can suggest some other choices, too, which would involve estimating g in $M_0^{-1} \int_0^1 f(x) g(x) dx$.

For a proof of theorem 2.3 see Agarwal (1978).

Note that (2.7) and (2.4) are the same. In a more practical situation the design measure μ will be discrete on a finite number of points. With some regularity conditions the IMSE is still of the form (2.7), however the constants appearing in front of the two terms will be different. Generally a smoother design (resembling the uniform) will keep the bias term small and give slightly larger values for the variance term. An appropriate discrete design will give smaller values for the variance term but will increase the bias (Agarwal and Studden 1978).

3. MINIMIZATION OF IMSE

In the last section we indicated that the asymptotic value of the IMSE, when $g(x)$ is estimated by spline function of degree one, is

$$J \approx \frac{k\sigma^2}{n} \int_0^1 \frac{p(x)}{h(x)}\, dx + \frac{1}{720k^4} \int_0^1 \frac{(g''(x))^2}{(p(x))^4}\, dx. \qquad (3.1)$$

Here we shall minimize the asymptotic value of the IMSE with respect to the three "variables" (i) k, the number of knots, (ii) $p(x)$, the displacement of knots and (iii) $h(x)$, the allocation of observations. The case when $g(x)$ is being approximated by spline function of degree zero can be dealt in a similar manner.

In the expression (3.1) for J, only the first term contains the factor $h(x)$. Using the Schwarz's inequality and the fact that $h(x)$ is a density, we can show that the first term in (3.1) is minimized by

$$h(x) = \sqrt{p(x)} \Big/ \int_0^1 \sqrt{p(y)}\, dy \qquad (3.2)$$

Substituting this value of $h(x)$ in (3.1) yields

$$J = \frac{\sigma^2 k}{n} \left(\int_0^1 \sqrt{p(x)}\, dx\right)^2 + \frac{1}{720k^4} \int_0^1 \frac{(g''(x))^2}{(p(x))^4}\, dx \qquad (3.3)$$

Now the problem is reduced to minimize J with respect to p and k. It will be shown in the following theorem that the minimizing p and k are given by

$$p(x) = \{g''(x)\}^{4/9}/\int_0^1 \{g''(y)\}^{4/9}dy \qquad (3.4)$$

and

$$k = \int_0^1 (g'')^{4/9}\{(n/180\sigma^2)(1/\int(g'')^{2/9})\}^{1/5} \qquad (3.5)$$

Substituting from (3.2), (3.4) and (3.5) the values of h, p and k in (3.1) we see that the second factor on right (squared bias B) is equal to

$$.25(180)^{-1/5}(\sigma^2/n)^{4/5}(\int(g'')^{2/9})^{9/5} \qquad (3.6)$$

and the first factor (variance V) is four times B. Therefore the minimum value for IMSE is

$$J \simeq C\, n^{-4/5} \qquad (3.7)$$

where $C = 1.25(180)^{-1/5}(\sigma^2)^{4/5}(\int(g'')^{2/9})^{9/5}$.

Theorem 3.1: The functional J given in (3.1) is absolutely minimized by h, p and k given in (3.2), (3.4) and (3.5) respectively.

Proof: We have already shown the minimization of J with respect to h. Now differentiating the expression (3.3) with respect to k and equating it to zero, we get

$$\frac{\sigma^2}{n}(\int\sqrt{p})^2 - \frac{1}{180k^5}\int\frac{(g'')^2}{p^4} = 0$$

which yields

$$k = \{(n/\sigma^2)\int\frac{(g'')^2}{p^4} \cdot \frac{1}{(\int\sqrt{p})^2}\}^{1/5} \qquad (3.8)$$

We can verify that this k minimizes J for each p. Substituting this value of k in (3.3), we get

$$J = \rho\{\int \frac{g''^2}{p^4}\}^{1/5}\{\int\sqrt{p}\}^{8/5} \qquad (3.9)$$

where ρ is a constant, independent of p. Finally we have to minimize J given in (3.9) with respect to p(x). This can be done by using Holder's inequaty

$$\int\psi\phi \le (\int\psi^{\alpha})^{1/\alpha}(\int\phi^{\beta})^{1/\beta}. \qquad (3.10)$$

In (3.10), let us take $\psi^{\alpha} = (g'')^2/p^4$, $\phi^{\beta} = \sqrt{p}$. Now we want to choose α and β in such a way that $\Psi\phi$ is independent of p and $1/\alpha + 1/\beta = 1$. This can be done by choosing $\alpha = 9$, $\beta = 9/8$, $\psi = (g'')^{2/9}/p^{4/9}$ and $\phi = p^{4/9}$. Substituting these values in (3.10) gives

$$\int(g'')^{2/9} \le \{\int \frac{(g'')^2}{p^4}\}^{1/9}\{\int\sqrt{p}\}^{8/9}.$$

In the above, equality holds if and only if $p(x) = \alpha(g''(x))^{4/9}$, where α is a constant. This shows that J in (3.7) is minimized by $p = \alpha(g'')^{4/9}$, where $\alpha = 1/\int(g'')^{4/9}$, since p is a density. Now putting this value of p in (3.8) we get the desired result. q.e.d.

These minimization results indicate that knots should be placed where g'' is large. The relation h α $\sqrt{p}$ indicates that h should move away from p becoming more uniform. The relation (3.5) indicates that there should be many more observations than knots.

Remark: In theorem 3.1, we have assumed that all of the three variables, namely k, p and h are unknown. We might confront situations when one or two of these variables are known, e.g. we might be given the number of knots or the distribution of observations or both, and so on. Also in the theorem minimizing k should be less than or equal to n. When k given by (3.5) becomes greater than n, it is hard to find solutions where $k \leq n$ and we have to resort to approximate solutions. For a discussion of these approximate solutions and the partial minimization problem see Agarwal (1978).

4. ALGORITHM

Here we shall discuss how the theoretical results indicated in the previous sections can be exploited in adaptively estimating a more or less arbitrary response function g. To make matters simple, in this section, we shall again estimate g by simple linear spline functions.

We consider the following iterative procedure. We are given N_0 observations on g which are distributed among r_0 points $x_i^0,\ldots,x_{r_0}^0$ with n_i^0 observations at x_i^0, $i=1,..,r_0$, so that $\sum_{i=1}^{r_0} n_i^0 = n_0$.

Let y_{ij} denote the jth observation $(j=1,\ldots,n_i^0)$ at x_i^0. We begin with a knot-set

$$\Pi_0: \quad 0 = \xi_0^0 < \xi_1^0 < \ldots. < \xi_k^0 < \xi_{k+1}^0 = 1.$$

This partition Π_0 is an initial guess perhaps based on some information about the function g being estimated. In the absense of information about g, a uniform knot spacing may be used. Let us now spell out the steps involved in this iterative procedure and then explain the implementation of these steps.

I. Estimate $g(x)$, $g''(x)$.

II. Estimate σ^2 and IMSE.

III. Find $\hat{h}$, an estimate of best design h. Add, say, r_1, points of observations $x_1^1, \ldots, x_{r_1}^1$. Let n_1 more observations are taken. Distribute the n_1 observations among r_0+r_1 points $x_1^0, \ldots, x_{r_0}^0$, $x_1^1, \ldots, x_{r_1}^1$ so that the displacement of the combined set of n_0+n_1 observations resembles $\hat{h}$ as closely as possible.

IV. Estimate $g(x)$ and $g''(x)$.

V. Estimate σ^2.

VI. Find $\hat{k}$, an estimate of k and $\hat{p}$, an estimate of knot displacement p.

VII. Estimate $g(x)$, $g''(x)$.

VIII. Find estimate of IMSE.

Steps III through VIII are called a cycle. The cycle is repeated unless a termination is encountered. We shall talk about the termination criterion later. Now we discuss the implementation of the above algorithm.

Step I: To estimate g we use two kinds of estimators: (a) variance minimizing estimator ($\hat{\underset{\sim}{\theta}}_{LSE}$) and (b) bias minimizing estimator ($\hat{\underset{\sim}{\theta}}_{BME}$). The estimation of $g''(x)$ can be done in many different ways. We indicate three of them: (a) Find the LSE of $g(x)$ by a cubic spline and then take the second derivative of the fitted cubic spline to get an estimate of $g''(x)$. There are some standard routines for estimating g by a cubic spline, e.g. see de Boor and Rice (1968a and 1968b). (b) g'' can be estimated on intervals centered at the knots by a simple second difference with estimated value of g at adjacent knots. (c) Let $s_0(x) = \hat{g}(x)$ be the linear spline approximation to $g(x)$ which is obtained using one of the two estimators. We hope that the approximation s_0 contains enough information about g so that a reasonable approximation to g'' may be obtained from it. We cannot use s_0'' as an approximation to g'' for the obvious reason that s_0'' is zero except at the knots, where it does not exist. As suggested by Dodson (1972), we first find a broken line approximation of g'; its derivative may then be used as the desired approximation to g''.

For this purpose s_0', the derivative of s_0, is used. Note that s_0' is constant on the intervals (ξ^0_{i-1},ξ^0_i), $i=1,\ldots,k+1$. We put a breakpoint τ_i at the center of each interval $[\xi^0_{i-1},\xi^0_i]$, $i=1,..,k+1$; and define a broken line function $\tilde{s}$ on $\{\tau_i\}$ so as to interpolate s_0' at each τ_i. In the first and last intervals, $[0,\tau_1]$ and $[\tau_{k+1},1]$, the continuation of the broken line in the adjacent interior interval is used. Finally, with $\tilde{s}$ so defined, we use the step function $\hat{g}'' \overset{\text{def}}{=} \tilde{s}'$ as an estimate to g''.

<u>Step II</u>: We can base the estimate of σ^2 on the residual sum of squares

$$S_R = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

if the model obtained is correct but not otherwise. If the prior estimate of σ^2 is available we can see (or test by an F-test) whether or not the residual mean square is significantly greater than the prior estimate. If it is significantly greater we say that there is lack of fit and we would reconsider the model. If no prior estimate of σ^2 is available, but repeat measurements of y have been made at the same value of x, we can use the mean square for pure error defined by

$$MS_{P.E.} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (y_{ij}-\bar{y}_i)^2 / (\sum_{i=1}^{r} n_i - r)$$

as an estimate of σ^2.

The IMSE for the LSE is

$$J = \int_0^1 E(g(x) - \underset{\sim}{f}'(x)\underset{\sim}{M}^{-1}(\mu)\int \underset{\sim}{f}(x)\bar{y}_x d\mu(x))^2 dx$$

$$= V + B$$

where

$$V = \frac{\sigma^2}{n} \text{ Tr } \underset{\sim}{M}^{-1}(\mu)\underset{\sim}{M}_0, \tag{4.1}$$

and

$$B = \int_0^1 (g(x) - \underset{\sim}{f}'(x)\underset{\sim}{M}^{-1}(\mu)\int \underset{\sim}{f}(x) g(x) d\mu(x))^2 dx \tag{4.2}$$

The only unknown parameter in V is σ^2, so to find an estimate of V we replace σ^2 in (4.1) by $\hat{\sigma}^2$, the estimate of σ^2 obtained above. If g(x) is known we can evaluate the integral in (4.2) to get B. If the form of g(x) is unknown, which is usually the case, we can find an estimate of B as follows. First using the trapezoidal rule we replace the integral in (4.2) by a summation:

$$B \approx \sum_{i=2}^{r} \frac{1}{2}(x_i - x_{i-1})\{(g(x_i)-u(x_i))^2 + (g(x_{i-1})-u(x_{i-1}))^2\}$$

where

$$u(x_i) = \underset{\sim}{f}'(x_i)\underset{\sim}{M}^{-1}(\mu)\int\underset{\sim}{f}(x)g(\mathbf{x})d\mu(x), \ i=1,..,r.$$

Now we replace $g(x_i)$ by $\bar{y}_i$, the mean of the observations at x_i to get an estimate of B as

$$\hat{B} = \sum_{i=2}^{r} \frac{1}{2}(x_i - x_{i-1})\{(\bar{y}_i - \bar{u}(x_i))^2 + (\bar{y}_{i-1} - \bar{u}(x_{i-1}))^2\},$$

where

$$\bar{u}(x_i) = \underset{\sim}{f}'(x_i)\underset{\sim}{M}^{-1}(\mu)\int\underset{\sim}{f}(x)\bar{y}_x d\mu(x), \ i=1,..,r.$$

The above method gives a good estimate of B if we have many more observations than the number of knots.

We can similarly find the estimate of the IMSE for the other estimator $(\hat{\underset{\sim}{\theta}}_{BME})$.

<u>Step III</u>: We have obtained $\hat{g}''$ in step I. So from (3.2) and (3.4), we get

$$\hat{h}(x) = (\hat{g}''(x))^{2/9}/\int(\hat{g}''(y))^{2/9}dy.$$

(a) Find r_0+r_1 points t_i's according to the quantiles of $\hat{h}$, i.e. find t_i's such that

$$\int_0^{t_i} \hat{h}(x)dx = (i-1)/(r_0+r_1-1), \ i=1,..,r_0+r_1.$$

(b) Among the r_0+r_1 t_i's find the ones which are close to x_j^0, $j=1,..,r_0$, the remaining r_1 t_i's will be the points x_j^1, $j=1,...,r_1$ which are to be added at this stage. We arrange these r_0+r_1 points in increasing order and denote the ordered set as $x_1,...,x_r$, where $r = r_0 + r_1$.

(c) $\hat{h}$ is a continuous design, we have to discretize it. To do this, we find $H(x) = \int_0^x \hat{h}(t)dt$ and approximate it by a distribution function $\hat{G}(x)$ having jumps at x_i, $i=1,\ldots,r$. We take $\hat{G}$ to be a uniform approximation of H, i.e. $\hat{G}$ is the solution of the problem

$$\min \max_x |H(x)-G(x)|$$

where the minimum is taken over all step functions G having jumps at x_i, $i=1,\ldots,r$. It is easy to show that

$$\begin{aligned}\hat{G}(x) &= 0 &&, x < 0\\ &= \frac{1}{2}(H(x_{i-1})+H(x_i)), && x_{i-1} \le x < x_i,\ i=2,\ldots,r\\ &= 1 &&, x \ge 1.\end{aligned}$$

Note that $x_1=0$ and $x_r=1$. So now we have the discrete design as below:

$$\mu = \begin{cases} x_1,\ldots,x_r \\ \mu_1,\ldots,\mu_r, \end{cases}$$

where $\mu_1=\hat{G}(x_1)$, $\mu_i=\hat{G}(x_i)-\hat{G}(x_{i-1})$, $i=2,\ldots,r$.

(d) Now we have to allocate $n_0+n_1=n$ observations according to the design μ. This can be done by a scheme given in (Federov 1972, Section 3.1). We allocate $[(n-r)\mu_i]^+$ observations to point x_i, $i=1,\ldots,r$, where $[c]^+$ indicate the smallest integer satisfying the inequality $[c]^+ \ge c$. It is trivial to check that $\sum_{i=1}^{r} [(n-r)\mu_i]^+ \le n$. The remaining unrealized observations $n' = n-\sum_{i=1}^{r} [(n-r)\mu_i]^+$ can be added one-by-one up to the point where

$$(n-r)\mu_i \ge [(n-r)\mu_i]^+ - \frac{1}{2}.$$

We can also distribute n' remaining observations randomly among r points. This scheme for the distribution of observations works well if r is very small compared to n. Note that we should make sure that we took at least n_i^0 observations at points x_i^0, $i=1,\ldots,r_0$.

Steps IV and V can be implemented in a way similar to steps I and II respectively.

Step VI: Once we get the estimate of σ^2 and g'', we can immediately find $\hat{k}$ from (3.5),

$$\hat{k} = \int \hat{g}''^{4/9}\left\{\frac{n}{180\sigma^2} \cdot \frac{1}{\int \hat{g}''^{2/9}}\right\}^{1/5}$$

Also we have from (3.4)

$$\hat{p}(x) = (\hat{g}''(x)^{4/9}) / \int (\hat{g}''(y))^{4/9} dy$$

Now we obtain the partition Π_1 consisting of the $\hat{k}$ knots $\xi_1^1,\ldots,\xi_k^1$. The ξ_i^1's are obtained from the integral relationship

$$\int_0^{\xi_i^1} \hat{p}(x)dx = i/(\hat{k}+1), \; i=0,1,\ldots,\hat{k}+1.$$

ξ_0^1 and ξ_{k+1}^1 represent the end points.

Steps VII and VIII have been already discussed.

Let us make a few remarks concerning the algorithm. The estimates $\hat{h}$, $\hat{p}$ and $\hat{k}$ depend on the estimates of g'' which is discussed in step I. We have suggested three methods of finding an estimate to g'' and there are many other ways in which this can be done. Because of its simplicity we shall usually adopt the method indicated in (c) of step I. There the estimate of g'' is obtained from the linear spline approximation s_0. This s_0 should therefore be a good approximation to g, whatever that may mean. Thus, generally, the starting partition Π_0 must consist of a large number of knots and they must be distributed through the interval in a reasonable way. It is conceivable that a very bad choice of the partition Π_0 could result in a "misleading" spline approximation s_0 and that the resulting partition Π_1 be still worse. We did not confront such cases of instability during the testing of the algorithim.

The estimates $\hat{h}$, $\hat{p}$, and $\hat{k}$ obtained here are based on the asymptotic results, so if $\hat{k}$ obtained from (3.5) is small, it might not be a very good estimate of the number of knots. However it has been found useful to do the following. Instead of finding a partition consisting of $\hat{k}$ knots, we find five different partitions consisting of $\hat{k}-2$, $\hat{k}-1$, $\hat{k}$, $\hat{k}+1$, and $\hat{k}+2$ knots respectively. In each partition the knots are chosen according to the quantiles of same $\hat{p}$. We find the estimate of g(x) and the corresponding integrated mean square error using the knot sets of each of the five partitions. Now select the partition for which the IMSE is minimum. We call the points of this partition to be "good" knot set. It has also been found useful to iterate the algorithm a few times with a fixed number of knots. This allows the algorithm to base its resulting "good" set of knots on a "good" set of the same size. One or two iterations has usually been satisfactory.

Termination Criteria: We use here two termination criteria for the algorithm. The first is a simple bound on the number of complete cycles, i.e. we shall perform no more than "m" cycles. Usually m=5 or 6 is found to be a reasonable number. The second criteria is based on the test for the "lack of fit". This test is indicated in step II. If there is no lack of fit in the model, the cycles are terminated. Other than these two we can use some other termination criteria depending on our problem, e.g. if the IMSE at any stage is not decreased much compared to the IMSE at the previous stage, then also we can stop.

5. EXAMPLE

Here we shall illustrate the algorithm described in the last section by a numerical example. The algorithm is illustrated on the measurements of the function

$$g(x) = \cdot 125[(\cdot 1)^2+(2x-\cdot 3)^2]^{-1} + \cdot 125[(\cdot 12)^2+(2x-1\cdot 2)^2]^{-1} \quad (5.1)$$

For this function $g''(x)$ varies by a large amount in the interval $[0,1]$. We simulated the data errors by adding to $g(x_i)$ a number sampled from the normal distribution with mean zero and variance one hundred. We started with three equally spaced knots and took five observations at each knot and the end points. We call this cycle zero. We performed ten cycles and added one hundred observations at each cycle. This was done for both the estimates namely $\hat{\underset{\sim}{\theta}}_{LSE}$ and $\hat{\underset{\sim}{\theta}}_{BME}$. The linear spline fits obtained by the two estimates improved at each cycle. The results of cycle zero, cycle six, and last cycle for $\hat{\underset{\sim}{\theta}}_{LSE}$ and $\hat{\underset{\sim}{\theta}}_{BME}$ are shown in Figure A and Figure B respectively. The breaks (joints) in the graph are the knots. The algorithm has chosen the knots at the points where curve is taking turn, and it seems reasonable. Here we stopped after ten cycles. If we would have let the algorithm run for a few more cycles, the estimated response function would have been right on the top of the actual function and would not be able to see the difference between actual function and the estimated function. For the same reason we have shown in the figures only the result of three cycles.

We calculated the integrated mean square error for the two estimates at the end of each cycle. The results for LSE and BME are shown in Table a and Table b respectively. As expected the LSE did a good job in reducing the variance, while BME did a comparable job in reducing the bias. Actually, we notice that BME has done fairly good job in reducing the variance also.

Here at cycle zero i.e. start of the algorithm the integrated variance (V) was almost equal to integrated squared bias (B). If at the start of the algorithm B is larger than V then BME shows considerable superiority over LSE in the sense that it reduces B much faster than LSE does.

At the start of the algorithm if V is larger than B then the present set up does not show strong case for LSE minimizing V faster than BME does. This is apparently due to the fact that after two or three cycles and a few observations have been taken, these observations are dispersed, at least in this example, in a somewhat uniform manner so that LSE and BME operate on V in a similar manner.

One can compare the above procedures with a procedure, say, equally spaced knots and an equal number of observations at the knots. For a fixed number of observations this procedure is comparable to the above procedures when the number of knots used is close to the optimal, which is used by our procedures. In practice, however, the experimenter does not know this number.

In the example discussed, we took 100 observations at each cycle for ten cycles and then terminated. Some criterian for termination were discussed in Section 4. Other modifications are possible. For example we can aim at a fixed value for IMSE, say IMSE = 1 in the following manner. We note from (3.7) that IMSE $\simeq$ C $n^{-4/5}$. We can use this equation to estimate C at each cycle. After four or five cycles or if C appears to stabilize we can insert the estimated value of C and IMSE = 1 in equation (3.7) and solve it for n. This gives an estimate of the number of observations needed to reach the value of IMSE = 1. The estimated C, however, is usually too large so that the estimated n is also usually too large. For example here for the BME estimate with 525 observations the value of IMSE is 2·376. To reach IMSE = 1 we estimate n = 1548. We notice (Table b) that for n = 1025 we have already IMSE = 1.18.

Comparison of Estimates

Cycle Number	k	n	Integrated Variance V	Integrated Bias B	Integrated Mean Square Error IMSE
			a. LSE estimate		
0	3	25	9.711	10.693	20.404
1	3	125	3.996	5.740	9.736
2	3	225	2.122	5.477	7.599
3	4	325	2.139	3.857	5.996
4	5	425	1.428	2.489	3.917
5	5	525	1.284	2.486	3.770
6	5	625	1.107	2.491	3.598
7	6	725	1.075	1.465	2.540
8	6	825	0.957	1.465	2.422
9	6	925	0.842	1.448	2.290
10	6	1025	0.729	1.450	2.179
			b. BME estimate		
0	3	25	9.711	10.693	20.404
1	4	125	3.947	4.841	8.788
2	6	225	2.946	1.050	3.996
3	6	325	2.170	0.816	2.986
4	6	425	1.815	0.828	2.643
5	6	525	1.554	0.822	2.376
6	7	625	1.366	0.539	1.905
7	8	725	1.304	0.300	1.604
8	9	825	1.245	0.184	1.429
9	9	925	1.131	0.181	1.312
10	9	1025	1.039	0.141	1.180

FIGURE A: Function g(x) and LSE at end of cycles zero, six and ten.

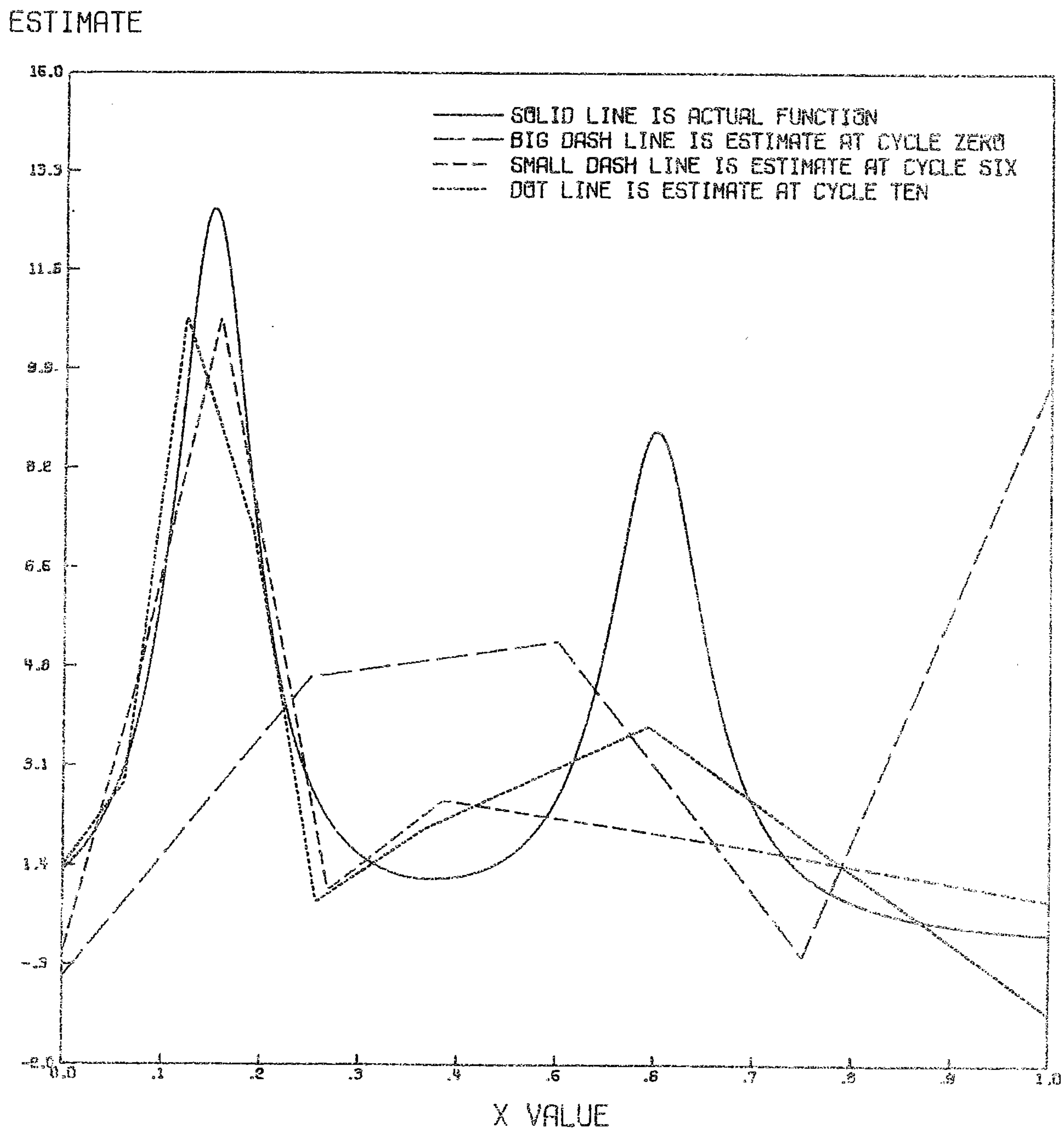
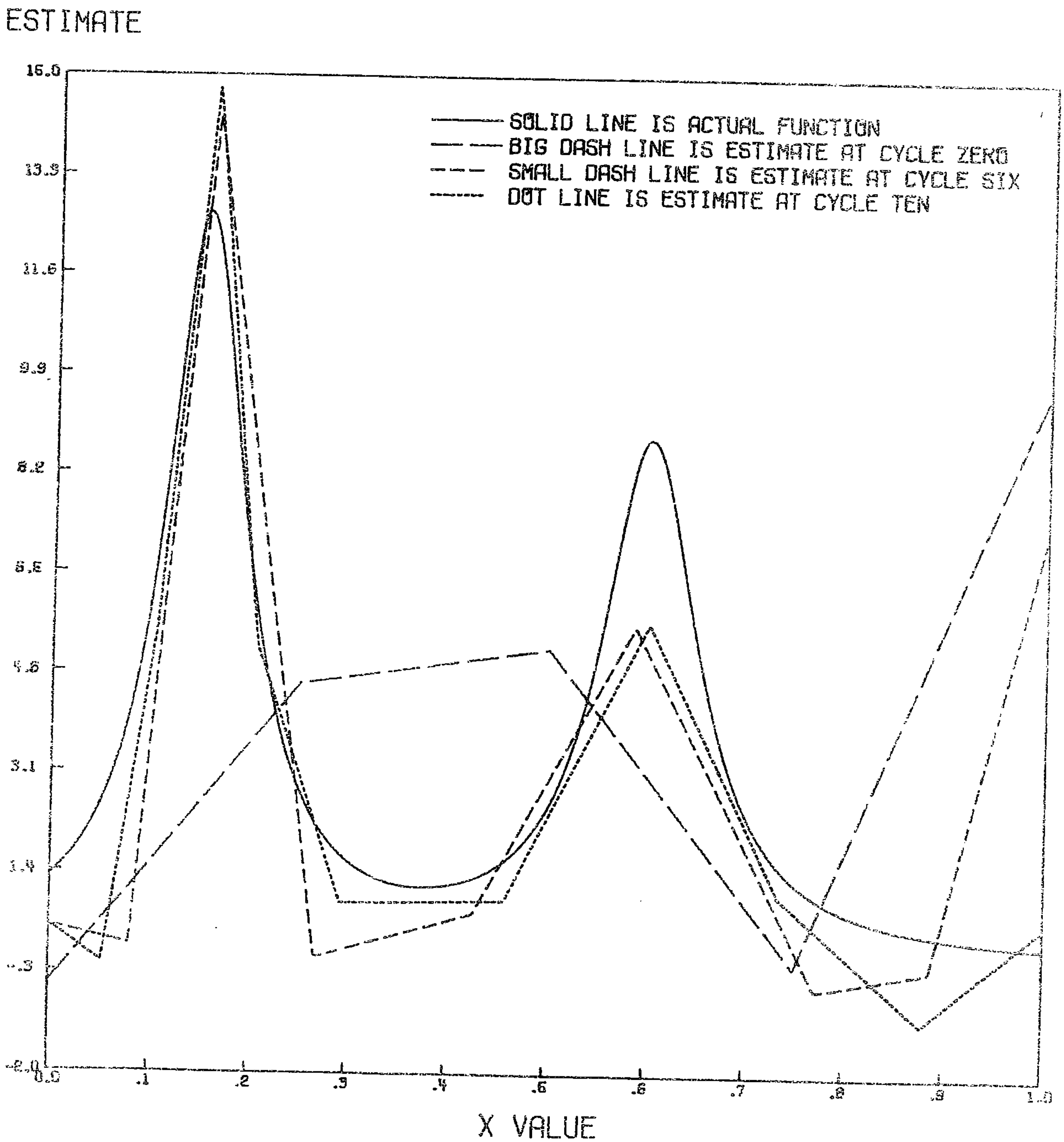

ESTIMATE
16.0
13.3
11.6
9.9
8.2
6.6
4.8
3.1
1.4
-.3
-2.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
SOLID LINE IS ACTUAL FUNCTION
BIG DASH LINE IS ESTIMATE AT CYCLE ZERO
SMALL DASH LINE IS ESTIMATE AT CYCLE SIX
DOT LINE IS ESTIMATE AT CYCLE TEN
X VALUE

FIGURE B: Function g(x) and BME at end of cycles zero, six and ten.


ESTIMATE
15.0
13.3
11.6
9.9
8.2
SOLID LINE IS ACTUAL FUNCTION
BIG DASH LINE IS ESTIMATE AT CYCLE ZERO
SMALL DASH LINE IS ESTIMATE AT CYCLE SIX
DOT LINE IS ESTIMATE AT CYCLE TEN
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
X VALUE

REFERENCES

Agarwal G. G. (1978), "Asymptotic Design and Estimation in Polynomial Spline Regression", Mimeograph Series #78-11, Department of Statistics, Purdue University, W. Lafayette.

_____, and Studden, W. J. (1978), "Asymptotic Design and Estimation using Linear Splines". To appear in Communications in Statistics.

de Boor, C., and Rice, J. R. (1968a), "Least Square Cubic Spline Approximation I-Fixed Knots", Purdue University Computer Science Technical Report No. 20.

____, and Rice, J. R. (1968b) "Least Square Cubic Spline Approximation II-Variable Knots", Purdue University Computer Science Technical Report No. 21.

Box, G. E. P., and Draper, N. R. (1959), "A Basis for Selection of Response Surface Design", Journal of the American Statistical Association, 54, 622-654.

Burchard, H. G. (1974), "Splines (with Optimal Knots) are Better", Applicable Analysis, 3, 309-319.

Dodson, D. S. (1972), "Optimal Order Approximation by Polynomial Spline Functions", unpublished Ph.D. Thesis, Department of Computer Science, Purdue University.

Federov, V. V. (1972), Theory of Optimal Experiment, Translated and Edited by E. M. Klimko and W. J. Studden, New York, Academic Press.

Greville, T. N. E. (1969), Theory and Application of Spline Functions, New York, Academic Press.

Guest, P. G. (1961), Numerical Methods of Curve Fitting, Cambridge, Cambridge University Press.

Holt, J. N. (1974), "A Self-Limiting Inversion Technique for Centre to Limb Data," Astronomy and Astrophysics, 30, 185-188.

Jupp, D. L. B. (1978), "Approximation to Data by Splines with Free Knots", 15, No. 2, 328-343.

____, and Stewart, Ian C. F. (1974), "A Piecewise Exponential Model for Seismic Well Logging Data, Journal of the Internation Association for Mathematical Geology, 6, No. 1, 33-45.

Karson, M. J., Manson, A. R., and Hader, R. J. (1969), "Minimum Bias Estimation and Experimental Design for Response Surfaces", Technometrics, 11, 461-475.

Prenter, P. M. (1975), Splines and Variational Methods, New York, John Wiley and Sons.

Rice, J. R. (1969), "On the Degree of Convergence of Nonlinear Spline Approximation", in Approximation Theory with Special Emphasis on Spline Functions (edited by Schoenberg, I. J.), New York, Academic Press.

Sacks, J., and Ylvisaker, D. (1970), "Designs for Regression Problems with Correlated Errors III", Annals of Mathematical Statistics, 41, 2057-2074.

Vozoff, K., and Jupp, D. L. B. (1975), "Joint Inversion of Geophysical Data", Geophysical Journal of the Royal Astronomical Society, 42, 977-991.

Wold, S. (1971), "Analysis of Kinetic Data by Means of Spline Functions", Chemica Scripta, 1, 97-102.

_______, (1974), "Spline Functions in Data Analysis", Technometrics, 16, 1-11.