Open access • Journal Article • DOI:10.1080/01621459.1971.10482358

Robust Procedures for Estimating Polynomial Regression — Source link

Corwin L. Atwood

Published on: 01 Dec 1971 - Journal of the American Statistical Association (Taylor & Francis Group)

Topics: Polynomial regression and Estimator

ROBUST PROCEDURES FOR ESTIMATING POLYNOMIAL REGRESSION*

by

Corwin L. Atwood

Technical Report No. 143

University of Minnesota
Minneapolis, Minnesota

August, 1970

*Research supported in part by NSF grant GP 11021.

ABSTRACT

Robust Procedures for Estimating Polynomial Regression

For estimating a regression function $g(x)$ which is a polynomial of degree s on a compact interval, let U_s be the least squares estimator and ξ_s the optimal design. Suppose now that the regression function is a polynomial of degree s plus a small polynomial of higher degree k. We use an estimator of the form $(1-\alpha)U_s + \alpha U_k$ and a design of the form $(1-\beta)\xi_s + \beta\xi_k$, with $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. Our criterion for an optimal procedure is to minimize the maximum (over x) mean square error. For $k \leq s + 2$, $k \leq 10$ we find values of (α, β) which, compared to the standard procedures $\alpha = \beta = 0$ and $\alpha = \beta = 1$, show desirable robustness when the true regression function deviates from a polynomial of degree s by only a moderate amount.

Robust Procedures for Estimating Polynomial Regression[1]

by

Corwin L. Atwood

In this paper we seek good estimators and designs for estimating a regression function, when the function is assumed to be a polynomial of given degree plus a "small" polynomial of given higher degree. We propose a procedure and verify that in the cases computed our procedure shows desirable robustness properties when compared to the standard procedure.

Model.

For x in some closed interval $[a, b]$, suppose uncorrelated random variables Y_x are observable, each with variance σ^2 and mean $g(x)$, the regression function, where $g(x)$ is a polynomial of degree k for some given $k \geq 2$. For some s, $1 \leq s < k$, let $h(x)$ be a polynomial of degree s which best approximates $g(x)$, in the sense that for any polynomial $h_1(x)$ of degree s,

$$\max_x |g(x) - h(x)| \leq \max_x |g(x) - h_1(x)|.$$

By standard approximation theory (Meinardus (1967), p. 16), h is unique. Here and throughout this paper, $\max_x$ means the maximum for $a \leq x \leq b$. Define $\epsilon = \max_x |g(x) - h(x)|$. If an experimenter believes that the regression function is "approximately" a polynomial of degree s, with the deviation from this polynomial equal to a polynomial of degree k, then he believes the assumptions of our model

[1]Research supported in part by NSF grant GP 11021.

with ε "small." This will be the situation of interest to us.

This model can be parametrized in various ways. By a proper change of variables we can assume $-1 \leq x \leq 1$, and will do so. We can write $g(x) = f'(x)\theta = f^{(1)'}(x)\theta^{(1)} + f^{(2)'}(x)\theta^{(2)}$, where f is any column vector of $k + 1$ linearly independent polynomials each of degree $\leq k$, partitioned into $f^{(1)}$ and $f^{(2)}$, with $f^{(1)}$ consisting of $s + 1$ polynomials each of degree $\leq s$. The vector θ of $k + 1$ unknown parameters is partitioned into $\theta^{(1)}$ and $\theta^{(2)}$, of dimensions $s + 1$ and $k - s$ respectively. Whatever functions are used in the vector f, note that $\varepsilon = 0$ if and only if $\theta^{(2)} = 0$, i.e., if and only if $g(x) = f^{(1)'}(x)\theta^{(1)}$. Eventually we will use $f'(x) = (1, x, \ldots, x^k)$.

The experimenter observes Y_x at various x and wishes to estimate $g(x)$. A design ξ specifies the proportion of observations to take at each x. If U_x is the estimator used at any particular x, we wish to minimize the maximum mean square error, $\max_x E(U_x - g(x))^2$. By a procedure we will mean a design and an estimator used together.

Previous results.

If $\varepsilon = 0$, i.e., the regression function is $f^{(1)'}(x)\theta^{(1)}$, then the best linear unbiased estimator and the optimal design are well known (Scheffe (1959), Guest (1958) and Hoel (1958).) We will call this the "standard procedure." When $\varepsilon = 0$ it has zero bias and maximum variance $= (s+1)\sigma^2/n$ (Kiefer and Wolfowitz (1960)). We will give procedures which are robust for $\varepsilon > 0$. That is they have maximum mean square error almost as small as that of the standard procedure when ε is close to 0, and smaller otherwise. We give numerical results for the cases $k \leq s + 2$, $k \leq 10$.

Box and Draper (1959) and (1963) consider a similar problem. Their x ranges over a multidimensional spherical region R, their vector of regression functions f consists of all monomials of degree ≤ 3 (resp. 2 in the 1959 paper) and $f^{(2)}(x) = x^3$ (resp. x^2). The only estimator U_x considered is the best linear unbiased estimator of $f^{(1)'}(x)\theta^{(1)}$ assuming $\theta^{(2)} = 0$. They seek a rotatable design which minimizes the integrated mean square error $\int_R E(U_x - f'(x)\theta)^2 dx$. Thus the Box-Draper assumptions differ somewhat from ours. However their results differ more markedly. The optimal designs obtained look quite similar to each other (in a sense made precise) for large values of $\theta^{(2)}$ (close to the "all-bias" situation). In contrast,we are chiefly concerned with the case when ε is small, hence $\theta^{(2)}$ is small. Also while we do not ever find optimal procedures, we do compare the mean square errors of several procedures in order to determine their relative efficiencies for various values of ε. Nothing like this is done by Box and Draper.

Karson, Manson and Hader (1969) follow Box and Draper, but use an estimator of the form $f^{(1)'}(x)b$, where b is a linear function of the observed Y values chosen to minimize the integrated squared bias $\int_R E[f^{(1)'}(x)b - f'(x)\theta]^2 dx$. They show that in several examples it is possible to use this estimator and an appropriate design to achieve smaller integrated mean square error than is possible using the Box-Draper estimator.

Stigler (1970) meets the problem of an imperfectly known model by defining new design criteria. He uses the traditional estimators in conjunction with the designs thus obtained.

Procedure to be used.

Let ξ_s denote the optimal design for estimating $f^{(1)'}(x)\theta^{(1)}$ assuming $\theta^{(2)} = 0$, and let ξ_k denote the optimal design for estimating $f'(x)\theta$ when θ is completely unknown. These designs take an equal number of observations at $s + 1$ and $k + 1$ points, respectively, where the points to use were found by Guest (1958) and Hoel (1958). Although these points can be found in principle from existing tables, the numerical values do not seem to have been computed and written down. They are therefore given in Appendix II. Throughout, we will assume that the total number of observations n is such that the required number of observations at each point can be taken. If we take n observations at (not necessarily distinct) points $x_1, \ldots, x_n$, let A be the $n \times (k+1)$ matrix with ith row $f'(x_i)$. Assume that A is of full rank, as it will be in all our applications. We let A_1 and A_2 be the matrices consisting of the first $s + 1$ and last $k - s$ columns of A, and let Y be the n-vector with components Y_{x_i}. Then $EY = A\theta = A_1\theta^{(1)} + A_2\theta^{(2)}$. If $\theta^{(2)} = 0$ the best linear unbiased estimator of $f'(x)\theta$ is $f^{(1)'}(x)(A_1'A_1)^{-1}A_1'Y$. If no assumption is made about $\theta^{(2)}$, the best linear unbiased estimator of $f'(x)\theta$ is $f'(x)(A'A)^{-1}A'Y$.

It seems reasonable to use an estimator and design approximating $f^{(1)'}(A_1'A_1)^{-1}A_1'Y$ and ξ_s if ε is small, and approximating $f'(A'A)^{-1}A'Y$ and ξ_k if ε is large. (We suppress x, the argument of f, here and below.) We will consider estimators of the form

$$U_x(\alpha) = (1-\alpha)f^{(1)'}(A_1'A_1)^{-1}A_1'y + \alpha f'(A'A)^{-1}A'Y$$

and designs of the form $\xi_\beta = (1-\beta)\xi_s + \beta\xi_k$, for parameters α and β with $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$.

For a given design ξ and a given point x, let us compute the mean square error of $U_x(\alpha)$. Direct computation shows that

$$EU_x(\alpha) = f^{(1)'}\theta^{(1)} + (\alpha f^{(2)'} + (1-\alpha)f^{(1)'}M_1^{-1}M_2)\theta^{(2)}$$

where we write $M = n^{-1}A'A$, and partition M as

$$M = \begin{bmatrix} M_1 & M_2 \\ M_2' & M_3 \end{bmatrix}$$

with M_1 and M_3 of sizes $(s+1) \times (s+1)$ and $(k-s) \times (k-s)$ respectively. Thus the bias is given by

$$(1) \qquad (1-\alpha)(f^{(1)'}M_1^{-1}M_2 - f^{(2)'})\theta^{(2)}.$$

We will return to this after dealing with the variance.

To calculate the variance write

$$(2) \qquad U_x(\alpha) = f^{(1)'}(A_1'A_1)^{-1}A_1'Y + \alpha(f'(A'A)^{-1}A' - f^{(1)'}(A_1'A_1)^{-1}A_1')Y.$$

We show as follows that the two terms on the right side of the equation are uncorrelated. If $\theta^{(2)} = 0$ then $Ef^{(1)'}(A_1'A_1)^{-1}A_1'Y = f^{(1)'}\theta^{(1)}$. Also, $Ef'(A'A)^{-1}A'Y = f'\theta$, which equals $f^{(1)'}\theta^{(1)}$ if $\theta^{(2)} = 0$. Equating the two expectations when $\theta^{(2)} = 0$ gives

$$f'(A'A)^{-1}A'A_1\,\theta^{(1)} = f^{(1)'}(A_1'A_1)^{-1}A_1'A_1\theta^{(1)} \text{ for all } \theta^{(1)}.$$

That is

$$(f'(A'A)^{-1}A' - f^{(1)'}(A_1'A_1)^{-1}A_1')A_1 = 0.$$

Therefore the two terms on the right side of equation (2) have covariance

$$(f'(A'A)^{-1}A' - f^{(1)'}(A_1'A_1)^{-1}A_1')\text{cov } Y(A_1(A_1'A_1)^{-1}f^{(1)})$$

which equals 0 because the covariance matrix of Y is $\sigma^2 I$. We conclude that

$$\text{(3)} \qquad \text{var } U_x(\alpha) = \text{var } f^{(1)'}(A_1'A_1)^{-1}A_1'Y + \alpha^2 \text{var}(f'(A'A)^{-1}A' - f^{(1)'}(A_1'A_1)^{-1}A_1')Y.$$

To evaluate this, set $\alpha = 1$. Then $U_x(1) = f'(A'A)^{-1}A'Y$, so

$$\text{var } f'(A'A)^{-1}A'Y = \text{var } f^{(1)'}(A_1'A_1)^{-1}A_1'Y + \text{var}(f'(A'A)^{-1}A' - f^{(1)'}(A_1'A_1)^{-1}A_1')Y.$$

We solve for the right hand term and substitute it into (3). Direct computation then gives

$$\text{var } U_x(\alpha) = \sigma^2 n^{-1}(1-\alpha^2)f^{(1)'}M_1^{-1}f^{(1)} + \sigma^2 n^{-1}\alpha^2 f'M^{-1}f.$$

If the design ξ_β is used, then this quantity also depends on β, and we write

$$\text{var } U_x(\alpha, \beta) = \sigma^2 n^{-1} V_x(\alpha, \beta)$$

defining $V_x(\alpha, \beta)$.

To carry out a numerical search for suitable α and β, it is useful to examine the behavior of $\max_x V_x(\alpha, \beta)$ in the region $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$. This is done in Appendix I.

We have expressed the bias in (1) in terms of $\theta^{(2)}$. Our assumptions specify ϵ rather than $\theta^{(2)}$, so let us express the bias in terms of ϵ. The expression depends on the parametrization used, so from now on suppose $f'(x) = (1, x, \ldots, x^k)$.

Suppose first that $k = s + 1$. In this case it is well known (Meinardus (1967), p. 31ff.) that the best degrees s approximation to the regression function g differs from g by T_k, the Chebyshev polynomial of degree k, multiplied by the maximum deviation from g. In our case the deviating polynomial is $\pm\epsilon\, T_k$, and $\theta^{(2)}$ is its leading coefficient, namely $\pm\epsilon\, 2^{k-1}$. Therefore from (1), the bias equals $\pm\, \epsilon\, B$, where

$$B = 2^{k-1}(1-\alpha)(f^{(1)'}M_1^{-1}M_2 - f^{(2)'}).$$

We have suppressed the dependence of B on x, α and β.

Now suppose $k = s + 2$. In this case $\theta^{(2)}$ has two components, and our assumption about the single quantity ϵ is not enough to determine $\theta^{(2)}$. Let the components of $\theta^{(2)}$ be θ_k and θ_{k+1} and let $\tau = -\theta_k/\theta_{k+1}$, possibly infinite. We may assume $\tau \geq 0$, for if it is not we can work with $g(-x)$, which has its $\tau \geq 0$. The best degree s approximation to g deviates from g by a constant times a so-called Zolotarev polynomial Z_k, which depends on τ. (See Meinardus (1967), p. 41ff.) If $\tau \leq k \tan^2(\pi/2k)$ then the Zolotarev polynomial is expressible in terms of a Chebyshev polynomial:

$$Z_k(x) = 2^{-k+1}(1+\tau/k)^k T_k\left(\frac{x-\tau/k}{1+\tau/k}\right).$$

This polynomial has maximum $2^{-k+1}(1+\tau/k)^k$. If $\tau > k \tan^2(\pi/2k)$ then the polynomial can be expressed in principle in terms of elliptic functions, but it is not readily usable. The one exception is if τ

is infinite, in which case $\theta_{k+1} = 0$ and the problem reduces to that discussed in the previous paragraph. Rather than obtaining results which are valid for all τ, we will choose 0, $k \tan^2(\pi/2k)$ and ∞ as convenient and hopefully representative values of τ and confine our results to these three cases. Argument similar to that in the last paragraph shows that the bias equals $\pm \varepsilon B$, where

$$B = 2^{k-1}(1+\tau/k)^{-k}(1-\alpha)(f^{(1)'}M_1^{-1}M_2 - f^{(2)'})\binom{-\tau}{1} \quad \text{if } 0 \le \tau \le k \tan^2(\pi/2k)$$

$$B = 2^{k-2}(f^{(1)'}M_1^{-1}M_2 - f^{(2)'})\binom{1}{0} \quad \text{if } \tau = \infty .$$

Because of the difficulty presented by the approximation theory, the cases $k > s + 2$ are not considered.

We now have expressions for the variance and bias. Denoting the ratio $n\varepsilon^2/\sigma^2$ by r, the mean square error written to show dependence on α, β, r and x is

$$n^{-1}\sigma^2 MSE_x(\alpha, \beta, r) = n^{-1}\sigma^2(V_x(\alpha, \beta) + rB_x^2(\alpha, \beta)).$$

Note that the relative importance of the bias in the mean square error depends not on ε alone, but on $r = n\varepsilon^2/\sigma^2$. This has a satisfying intuitive interpretation. If $\sigma n^{-1/2}$ is large compared to ε (typically if n is small) then we should try hardest to reduce the variance rather than the bias. If $\sigma n^{-1/2}$ is small compared to ε, as when n is large, then we should use some of our observations to reduce the bias.

The (α, β) pairs to use were found by numerical search based on the following criteria. We want procedures which have high efficiency

(say 95%, 90% and 80%) relative to the standard procedure when $\epsilon = 0$. (It turns out that a 99% efficient procedure behaves so much like the standard procedure that it's not worth the trouble to use it.) Thus we want (α, β) with $(s+1)/\max_x V_x(\alpha, \beta) = .95$, etc. In addition, for moderate ϵ we would like our procedure to have smaller mean square error than the standard procedure. To make this precise, when $k = s + 1$ we arbitrarily let r^* be that value of r for which the standard procedure $\alpha = \beta = 0$ has maximum mean square error $n^{-1}\sigma^2(k+1)$. (Thus at $r = r^*$, $\alpha = \beta = 1$ gives the same maximum mean square error as the standard procedure.) If $k = s + 2$ the mean square error depends on τ, so we get values $r^*(\tau)$. Then we let r^* be the smallest of $r^*(\tau)$ for $\tau = 0$, $k \tan^2(\pi/2k)$ and ∞. Among those (α, β) pairs which give the desired efficiency for $\epsilon = 0$, we choose that (α, β) which minimizes $\max_x(\alpha, \beta, r^*)$ when $k = s + 1$, and which minimizes $\max_\tau \max_x \mathrm{MSE}_x(\alpha, \beta, r^*)$ when $k = s + 2$. Other ways of deciding on an r^* would probably also give good results. All that is desired is a small mean squared error for moderate values of r. Details of the numerical methods used are described in Appendix I.

Numerical results.

We graph a few cases in Figures 1 and 2. In Figure 1 we treat the case $s = 1$, $k = 2$, and graph $\max_x \mathrm{MSE}_x(\alpha, \beta, r)$ as a function of r. For clarity only the standard procedure and the two procedures with efficiencies .95 and .80 at $r = 0$ are considered. The corresponding graphs are labeled 1.00, .95 and .80 respectively. The procedure $\alpha = \beta = 1$ has $\max_x \mathrm{MSE}_x = 3$ for all r, and thus is preferable to

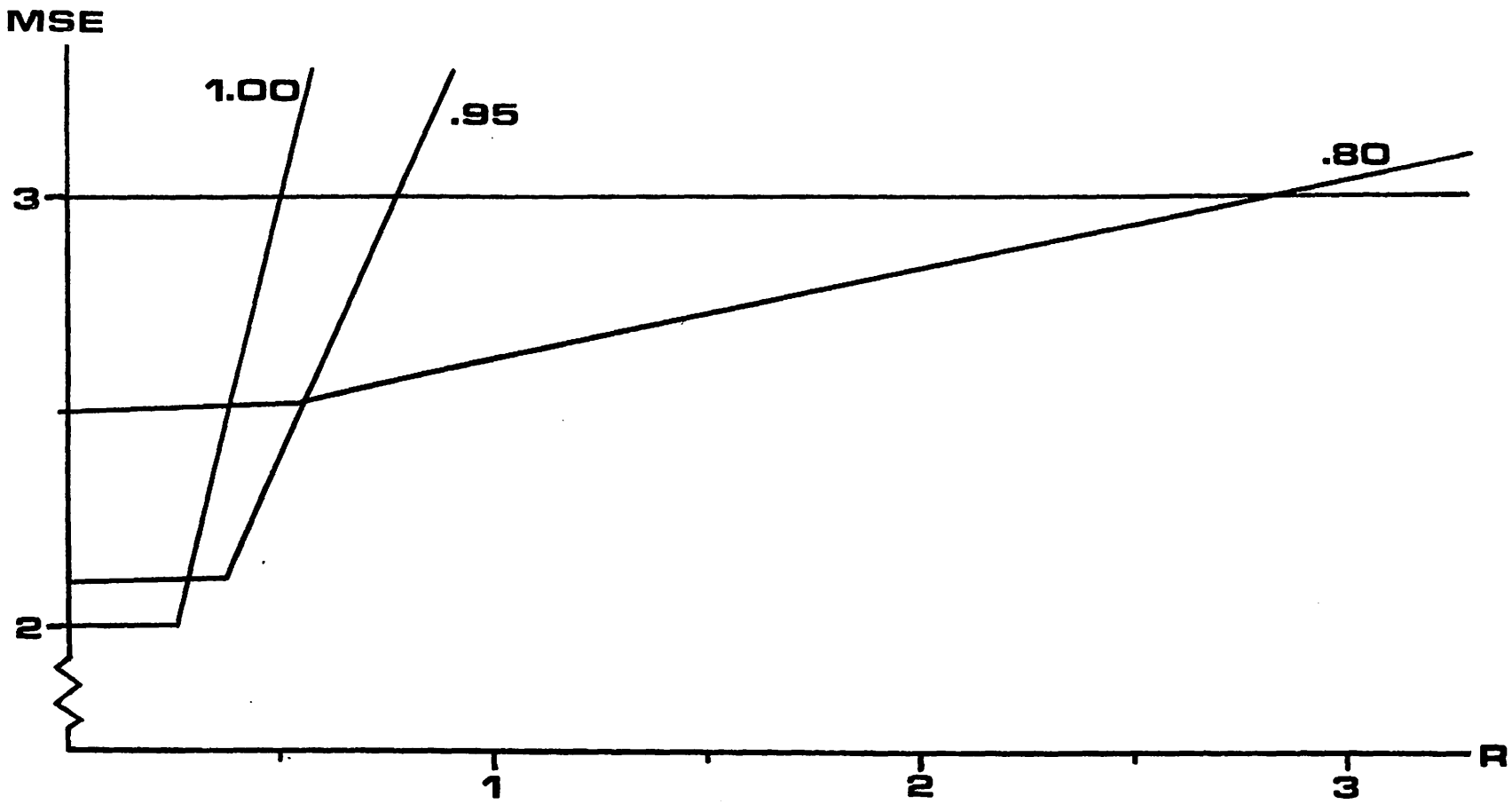

Figure 1: s = 1, k = 2

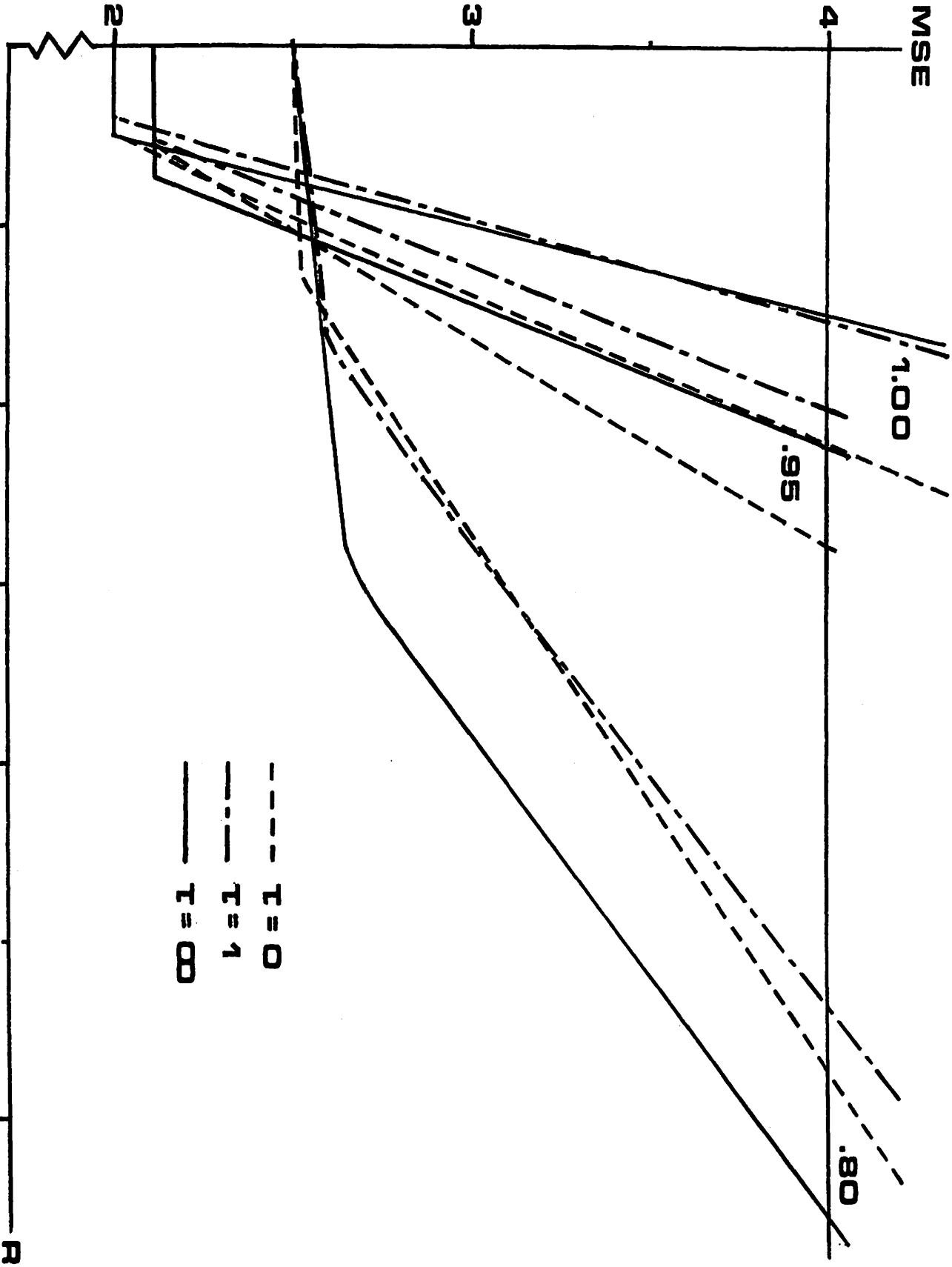

Figure 2: s = 1, k = 3

any of the other procedures if r is large. The line of constant height 3 is therefore shown for comparison.

In Figure 2 we treat the case $s = 1$, $k = 3$ and as in Figure 1 we consider the standard procedure and the two procedures with efficiencies .95 and .80 at $r = 0$. However in this figure each procedure has three graphs, corresponding to the three values of τ under consideration. (Note that $3 \tan^2(\pi/6) = 1$.) When $\tau = \infty$ the regression function is exactly that when $s = 1$, $k = 2$. It is interesting to note that the non-standard procedures of Figure 2 do not do as well against $\tau = \infty$ as the corresponding procedures of Figure 1; this is to be expected, since the procedures in Figure 2 are trying to cope with several values of τ simultaneously. The procedure $\alpha = \beta = 1$ has $\max_x MSE_x = 4$ for all r and τ, and thus is preferable to any of the other procedures if r is large. This line is shown for comparison.

The results for all the cases are summarized in Tables 1 and 2. We write MSE for $\max_x MSE_x$. For a given α and β, we write $e(r) = MSE(0, 0, r)/MSE(\alpha, \beta, r)$. The notation is suggestive of efficiency. However for $r > 0$ the ratio does not have the usual efficiency interpretation as a ratio of sample sizes, since for fixed α and β and fixed $\epsilon > 0$ the maximum mean square error is not a linear function of n^{-1}. In the tables, for α and β not both 0, r_0 is the point at which $MSE(0, 0, r_0) = MSE(\alpha, \beta, r_0)$. For α and β not both 1, r_1 is the point at which $MSE(\alpha, \beta, r_1) = k + 1$. Thus for r between r_0 and r_1 the (α, β) procedure under

Table 1: $k = s + 1$

s	α	β	$MSE(\alpha,\beta,0)$	$e(0)$	r_0	$MSE(0,0,r_0) = MSE(\alpha,\beta,r_0)$	r_1	$MSE(0,0,r_1)$	$e(r_1)$
1	0	0	2.000	1.00	-	-	.50	3.00	1.00
	.139	.281	2.105	.95	.28	2.11	.74	3.98	1.33
	.365	.492	2.222	.90	.31	2.23	1.17	5.69	1.90
	.689	.759	2.500	.80	.38	2.51	2.78	12.11	4.04
2	0	0	3.000	1.00	-	-	.82	4.00	1.00
	.189	.297	3.158	.95	.42	3.16	1.19	4.86	1.21
	.408	.546	3.333	.90	.52	3.35	1.89	6.51	1.63
	.845	.873	3.750	.80	.71	3.75	7.68	20.22	5.06
3	0	0	4.000	1.00	-	-	.68	5.00	1.00
	.173	.340	4.211	.95	.38	4.22	1.04	5.91	1.18
	.517	.617	4.444	.90	.47	4.45	1.92	8.18	1.64
4	0	0	5.000	1.00	-	-	.78	6.00	1.00
	.205	.385	5.263	.95	.45	5.27	1.25	7.06	1.17
	.630	.693	5.556	.90	.58	5.56	2.75	10.47	1.74
5	0	0	6.000	1.00	-	-	.73	7.00	1.00
	.212	.432	6.316	.95	.44	6.32	1.25	8.19	1.17
	.722	.772	6.667	.90	.59	6.67	3.41	13.21	1.89
6	0	0	7.000	1.00	-	-	.78	8.00	1.00
	.297	.478	7.368	.95	.49	7.37	1.49	9.57	1.12
	.821	.849	7.778	.90	.68	7.78	5.52	18.43	2.30
7	0	0	8.000	1.00	-	-	.76	9.00	1.00
	.394	.522	8.421	.95	.50	8.43	1.64	10.98	1.22
	.912	.925	8.889	.90	.71	8.89	10.37	30.41	3.38
8	0	0	9.000	1.00	-	-	.78	10.00	1.00
	.471	.566	9.474	.95	.54	9.48	1.90	12.41	1.24
9	0	0	10.000	1.00	-	-	.77	11.00	1.00
	.532	.611	10.526	.95	.55	10.53	2.07	13.83	1.26

Table 2: $k = s + 2$

s	α	β	$MSE(\alpha,\beta,0)$	$e(0)$	r_0	$MSE(0,0,r_0) = MSE(\alpha,\beta,r_0)$	r_1	$MSE(0,0,r_1)$	$e(r_1)$
1	0	0	2.000	1.00	-	-	1.12	4.00	1.00
							.76	4.00	1.00
							.75	4.00	1.00
	.128	.233	2.105	.95	.38	2.28	1.41	4.69	1.17
					.24	2.11	1.02	4.95	1.24
					.27	2.11	1.12	5.47	1.37
	.185	.435	2.222	.90	.36	2.23	1.70	5.37	1.34
					.27	2.25	1.37	6.24	1.56
					.31	2.25	1.54	7.15	1.79
	.405	.698	2.500	.80	.48	2.52	2.89	8.20	2.04
					.36	2.55	2.70	11.10	2.77
					.39	2.54	3.28	14.13	3.53
2	0	0	3.000	1.00	-	-	.78	5.00	1.00
							.71	5.00	1.00
							1.25	5.00	1.00
	.112	.222	3.158	.95	.32	3.16	1.13	6.37	1.27
					.29	3.16	.95	6.06	1.21
					.42	3.16	1.62	5.87	1.17
	.176	.426	3.333	.90	.37	3.34	1.57	8.16	1.63
					.33	3.35	1.29	7.57	1.51
					.53	3.36	2.06	6.90	1.38
	.533	.693	3.750	.80	.47	3.76	3.84	17.22	3.44
					.43	3.77	2.95	14.88	2.98
					.72	3.78	4.56	12.81	2.56
3	0	0	4.000	1.00	-	-	.80	6.00	1.00
							.79	6.00	1.00
							1.07	6.00	1.00
	.122	.238	4.211	.95	.36	4.21	1.21	7.65	1.28
					.35	4.22	1.10	7.31	1.22
					.38	4.21	1.49	7.06	1.18
	.191	.458	4.444	.90	.42	4.46	1.78	10.02	1.67
					.41	4.47	1.51	9.05	1.51
					.47	4.46	1.99	8.35	1.39
	.634	.744	5.000	.80	.56	5.01	5.59	25.76	4.29
					.55	5.02	4.11	19.97	3.33
					.69	5.02	6.56	20.05	3.34
4	0	0	5.000	1.00	-	-	.93	7.00	1.00
							.93	7.00	1.00
							1.22	7.00	1.00
	.147	.258	5.263	.95	.43	5.27	1.46	8.93	1.28
					.43	5.27	1.36	8.60	1.23
					.45	5.27	1.71	8.11	1.16
	.221	.497	5.556	.90	.53	5.57	2.22	11.78	1.68
					.53	5.58	1.89	10.59	1.51
					.59	5.58	2.32	9.51	1.36

Table 2 (cont.)

s	α	β	MSE(α,β,0)	e(0)	r_0	MSE(0,0,r_0)= MSE(α,β,r_0)	r_1	MSE(0,0,r_1)	e(r_1)
4	.731	.806	6.250	.80	.73	6.26	9.51	39.04	5.58
					.72	6.26	6.90	29.36	4.19
					.89	6.26	9.74	26.31	3.76
5	0	0	6.000	1.00	-	-	1.11	8.00	1.00
							1.11	8.00	1.00
							1.16	8.00	1.00
	.169	.281	6.316	.95	.46	6.32	1.76	10.05	1.26
					.61	6.68	1.34	8.73	1.09
					.44	6.32	1.74	9.35	1.17
	.263	.539	6.667	.90	.61	6.68	2.57	12.70	1.59
					.62	6.69	2.01	10.87	1.36
					.60	6.69	2.54	11.20	1.40
	.812	.875	7.500	.80	.95	7.50	12.49	45.39	5.67
					.95	7.50	9.57	35.80	4.47
					.95	7.51	13.94	37.67	4.71
6	0	0	7.000	1.00	-	-	.96	9.00	1.00
							.96	9.00	1.00
							1.23	9.00	1.00
	.162	.305	7.368	.95	.43	7.37	1.55	10.82	1.20
					.43	7.37	1.30	10.04	1.12
					.49	7.37	1.85	10.35	1.15
	.370	.575	7.778	.90	.56	7.79	2.81	14.67	1.63
					.57	7.79	2.29	13.08	1.45
					.68	7.79	3.21	13.36	1.48
7	0	0	8.000	1.00	-	-	.91	10.00	1.00
							.91	10.00	1.00
							1.21	10.00	1.00
	.165	.330	8.421	.95	.44	8.43	1.55	12.14	1.21
					.44	8.43	1.36	11.49	1.15
					.50	8.43	1.87	11.49	1.15
	.473	.611	8.889	.90	.58	8.90	3.27	17.86	1.79
					.58	8.90	2.70	15.98	1.60
					.71	8.90	3.93	16.07	1.60
8	0	0	9.000	1.00	-	-	.92	11.00	1.00
							.92	11.00	1.00
							1.25	11.00	1.00
	.175	.354	9.474	.95	.47	9.48	1.65	13.47	1.22
					.47	9.48	1.48	12.88	1.17
					.54	9.48	1.98	12.58	1.14
	.536	.651	10.000	.90	.63	10.01	3.98	21.39	1.94
					.63	10.01	3.32	19.14	1.74
					.79	10.01	4.54	18.11	1.65

consideration is preferable to both $\alpha = \beta = 0$ and $\alpha = \beta = 1$. If $(s+1)/(k+1) \geq .80$ then the procedure $\alpha = \beta = 1$ has $e(0) \geq .80$; in this case no procedure is given with $e(0) = .80$. Table 2 has the same format as Table 1, except that when α and β are not both 0 and r is not 0, three lines are needed, for the cases $\tau = 0$, $k \tan^2(\pi/2k)$ and ∞, given in that order. Whenever $e(r)$ is calculated, the same τ is used for both terms of the ratio. Entries are left blank if they merely duplicate the entries immediately above.

The nonstandard procedures all have a minimum value of $e(r)$ slightly less than $e(0)$. This is because $MSE(0, 0, r)$ is always constant for r very small, while for $\alpha > 0$, $\beta > 0$ we characteristically have $MSE(\alpha, \beta, r)$ rising gently for r near 0. Thus $e(r)$ dips slightly before rising. But in no case is this dip as great as .01.

An experimenter will usually be able only to approximate a tabulated β. In order to still have a desired value for $e(0)$, if his β differs from the tabulated β he may wish to use an α different from the tabulated α. Aids for doing this are in Appendix I.

Acknowledgments.

I am grateful for several very helpful conversations with Jack Kiefer. Portions of the computer programming were done by Paul Steinback and David Doren.

Appendix I

In this section we prove the properties of $\max_x V_x(\alpha, \beta)$ which are used in the numerical search, and describe the search.

We show first that for fixed α, $\max_x V_x(\alpha, \beta)$ is convex in β. For any square nonsingular matrix M, the derivative $d(MM^{-1})$ equals 0, and therefore $dM^{-1} = -M^{-1}(dM)M^{-1}$. Therefore for fixed x and α, and $0 < \beta < 1$

$$(4) \qquad \frac{\partial^2}{\partial\beta^2} V_x(\alpha, \beta) = 2(1-\alpha^2)f^{(1)'}(x)M_1^{-1}(dM_1)M_1^{-1}(dM_1)M_1^{-1}f^{(1)}(x) + 2\alpha^2 f'(x)M^{-1}(dM)M^{-1}(dM)M^{-1}f(x),$$

where dM and dM_1 denote first derivatives with respect to β. Since M^{-1} and M_1^{-1} are positive definite the expression (4) is nonnegative and $V_x(\alpha, \beta)$ is convex. To show convexity of $\max_x V_x(\alpha, \beta)$ consider any β_0, β_1 and t, all between 0 and 1, and let $\beta_t = (1-t)\beta_0 + t\beta_1$. Let x_t be a point where $V_x(\alpha, \beta_t)$ attains its maximum. Then

$$\max_x V_x(\alpha, \beta_t) = V_{x_t}(\alpha, \beta_t) \le (1-t)V_{x_t}(\alpha, \beta_0) + tV_{x_t}(\alpha, \beta_1) \le (1-t)\max_x V_x(\alpha, \beta_0) + t \max_x V_x(\alpha, \beta_1).$$

We now show that for fixed β, $\max_x V_x(\alpha, \beta)$ is increasing in α. We have

$$V_x(\alpha, \beta) = f^{(1)'}M_1^{-1}f^{(1)} + \alpha^2(f'M^{-1}f - f^{(1)'}M_1^{-1}f^{(1)}).$$

The expression in parentheses equals

$$(f^{(2)} - M_2'M_1^{-1}f^{(1)})'(M_3 - M_2'M_1^{-1}M_2)^{-1}(f^{(2)} - M_2'M_1^{-1}f^{(1)})$$

which is nonnegative. Thus for any x and β, $V_x(\alpha, \beta)$ is increasing in α. We get from this to $\max_x V_x(\alpha, \beta)$ by an argument similar to that in the last paragraph.

By the optimality of the standard procedures it is immediate that $\max_x V_x(0, \beta)$ has a minimum of $s + 1$ at $\beta = 0$, and $\max_x V_x(1, \beta)$ has a minimum of $k + 1$ at $\beta = 1$.

Thus for each α there is a β (if it is not unique, choose the smallest, say) which minimizes $\max_x V_x(\alpha, \beta)$, and these pairs (α, β) form a line which curves from $(0, 0)$ to $(1, 1)$. This divides the square $0 \le \alpha \le 1$, $0 \le \beta \le 1$ into two regions. We will refer to the region containing the line $\beta = 1$ as the "upper region." We are interested in the curves where $\max_x V_x(\alpha, \beta)$ is constant. On such a level curve in the upper region, β decreases as α increases. On such a curve in the lower region, β increases as α increases.

Of interest to us are the level curves $(s+1)/\max_x V_x(\alpha, \beta) = .95$, .90 and .80. We search along these curves for (α, β) minimizing $\max_x MSE(\alpha, \beta, r^*)$. A preliminary exploration of the lower region indicates that we can do better in the upper region, so a thorough search is made there. For each α, the β which lies on the appropriate curve is found by an iteration process. Along the curve, $\max_x MSE_x(\alpha, \beta, r^*)$ seems to be convex, so another iteration process is used to find the approximate minimum. Along this curve, $\max_x MSE_x(\alpha, \beta, r^*)$ does not vary radically, so it is not critical that we find the exact minimum (typically near the minimum $\max_x MSE_x(\alpha, \beta, r^*)$ varies by less than .1 as α varies by .1).

To avoid compounding too many iterations, $\max_x$ is always computed as the maximum for x on a discrete grid of points spaced .01 units apart. Inspection of the functions used indicates that we cannot guarantee our computed maxima to be accurate to more than two decimal places. The one case in which we can get better computed accuracy is if $r = 0$, for $V_x(\alpha, \beta)$ always attains its maximum at a point of the grid, namely at 1. We compute $(s+1)/\max_x V_x(\alpha, \beta)$ accurate to within $\pm$ .0001. Accuracy to two decimal places for $r > 0$ is really adequate, because the experimenter will only have a rough idea of the magnitude of r. Our results then give adequate guidance on what procedure to use.

If the experimenter can only approximate a tabulated β, he may wish also to modify the tabulated α in order to stay on a desired level curve of $\max_x V_x(\alpha, \beta)$. Our calculations indicate that this will not change $\max_x MSE_x$ greatly. For this purpose, Table 3 gives selected points on the level curves of $\max_x V_x(\alpha, \beta)$.

This table gives β values on the level curves of $\max_x V_x(\alpha, \beta)$ in the upper region, for α at multiples of .100. We define $e(0) = (s+1)/\max_x V_x(\alpha, \beta)$, and define (α_0, β_0) to be the point at which the level curve leaves the upper region.

Table 3

$k = s + 1$

s	e(0)	.000	.100	.200	.300	.400	.500	.600	.700	.800	.900	α_0	β_0
1	.95	.285	.283	.275	.264							.324	.261
	.90	.545	.541	.528	.508	.482						.471	.462
	.80	1.000	.993	.974	.943	.903	.857	.806	.754			.707	.750
2	.95	.302	.300	.296	.290							.355	.286
	.90	.583	.580	.573	.562	.547	.529					.540	.521
	.80							.966	.930	.891		.851	.870
3	.95	.343	.342	.339	.334							.366	.330
	.90	.660	.659	.654	.645	.633	.619					.586	.605
4	.95	.388	.387	.385	.380	.376						.414	.375
	.90	.747	.745	.741	.734	.724	.712	.697				.671	.685
5	.95	.435	.435	.433	.428	.424						.442	.422
	.90	.836	.834	.830	.824	.815	.803	.790	.775			.744	.769
6	.95	.483	.483	.481	.477	.473						.487	.468
	.90	.924	.923	.919	.913	.905	.895	.882	.868	.852		.830	.848
7	.95	.530	.530	.528	.525	.521	.516					.520	.512
	.90					.994	.984	.973	.959	.944	.927	.912	.925
8	.95	.578	.578	.576	.573	.569	.564					.565	.562
9	.95	.626	.626	.624	.621	.617	.613	.607				.602	.604

Table 3 (cont.)

$k = s + 2$

s	e(0)	.000	.100	.200	.300	.400	.500	.600	.700	.800	.900	α_0	β_0
1	.95	.238	.235	.225								.219	.223
	.90	.454	.448	.431	.406							.323	.399
	.80	.833	.823	.795	.752	.701						.493	.650
2	.95	.224	.223	.217								.241	.214
	.90	.436	.433	.423	.408							.366	.396
	.80	.824	.818	.802	.777	.744	.706					.580	.673
3	.95	.240	.238	.234								.260	.231
	.90	.467	.464	.457	.445	.430						.405	.429
	.80	.882	.878	.866	.847	.821	.791	.756				.662	.734
4	.95	.261	.260	.256								.271	.253
	.90	.507	.505	.499	.489	.476						.437	.470
	.80	.956	.953	.943	.927	.905	.879	.850	.817			.747	.801
5	.95	.284	.283	.280								.282	.275
	.90	.551	.549	.544	.535	.524						.464	.515
	.80					.991	.968	.941	.911	.879		.820	.873
6	.95	.308	.307	.304	.300							.303	.300
	.90	.596	.595	.590	.582	.571	.558					.504	.558
	.80									.972	.940	.910	.937
7	.95	.332	.331	.329	.324							.328	.323
	.90	.642	.641	.636	.629	.619	.607					.551	.598
8	.95	.356	.356	.353	.349							.342	.348
	.90	.688	.686	.682	.675	.666	.655					.589	.640

Appendix II

Optimal Designs for Polynomial Regression on [-1, 1]

The optimal design in each case is uniform on the points of support given. (Guest (1958) and Hoel (1958)). The values have been rounded off to 6 decimal places.

Degree of Polynomial	Points of Support
1	± 1
2	$0, \pm 1$
3	$\pm .447214, \pm 1$
4	$0, \pm .654654, \pm 1$
5	$\pm .285232, \pm .765055, \pm 1$
6	$0, \pm .468849, \pm .830224, \pm 1$
7	$\pm .209299, \pm .591700, \pm .871740, \pm 1$
8	$0, \pm .363117, \pm .677186, \pm .899758, \pm 1$
9	$.165279, \pm .477925, \pm .738774, \pm .919534, \pm 1$
10	$0, \pm .295758, \pm .565235, \pm .784483, \pm .934001, \pm 1$

REFERENCES

Box, G. E. P. and Draper, N. R. (1959). A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54 622-654.

Box, G. E. P. and Draper, N. R. (1963). The choice of a second order rotatable design. Biometrika 50 335-352.

Guest, P. G. (1958). The spacing of observations in polynomial regression. Ann. Math. Statist. 29 294-299.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Statist. 29 1134-1145.

Karson, M. J., Manson, A. R. and Hader, R. J. (1969). Minimum bias estimation and experimental design for response surfaces. Technometrics 11 461-475.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canad. J. Math. 14 363-366.

Meinardus, G. (1967). Approximation of Functions: Theory and Numerical Methods. Springer-Verlag, New York.

Scheffe, H. (1959). The Analysis of Variance. Wiley, New York.

Stigler, S. M. (1970). Optimal experimental design for polynomial regression. Technical Report No. 225, Dept. of Statistics, Univ. of Wisconsin. (To appear in J. Amer. Statist. Assoc.)