EFFICIENCY PROBLEMS IN POLYNOMIAL ESTIMATION

By Paul G. Hoel

University of California, Los Angeles

1. Summary. Using the generalized variance as a criterion for the efficiency of estimation, the best choice of fixed variable values within an interval for estimating the coefficients of a polynomial regression curve of given degree is determined for the classical regression model. Using this same criterion, some results are obtained on the increased efficiency arising from doubling the number of equally spaced observation points

(i) when the total interval is fixed and

(ii) when the total interval is doubled. Measures of the increased efficiency are found for the classical regression model and for models based on a particular stationary stochastic process and a pure birth stochastic process.

2. Introduction. In the classical theory of regression, a set of values $x_1, x_2, \cdots, x_n$ of a variable x is selected and observations are made on a related variable y corresponding to those selected x values. If y_i denotes the y value corresponding to x_i, it is then assumed that $y_1, y_2, \cdots, y_n$ are uncorrelated variables with a common variance σ^2. Now if it is assumed that the means of the y's lie on a polynomial curve of degree k, that is, that

$$E(y_i) = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k \tag{1}$$

then a basic problem in statistics is how best to estimate the β's.

There are two aspects to this estimation problem. One is to determine the best method for using the information given by a set of n observtions $y_1, y_2, \cdots, y_n$. The other is to determine the best method for choosing the x values at which to take observations.

Although much research has gone into studying the first aspect of the problem, considerably less has been done on the second. Many years ago, K. Smith [1] was able to determine those x values within a fixed interval that minimize the maximum variance of a single estimated ordinate for polynomials up to degree six. More recently, De La Garza [2] was able to show that just as much information is obtained from observations made at certain $k + 1$ points in the interior of an interval as from n distinct points in that interval. Elfving [3], Chernoff [4], Daniels [5], and Ehrenfeld [6] have also made contributions toward this and other closely related problems.

In this paper an optimum solution based on the generalized variance is given for the problem of how to choose the x values in an interval for the classical regression model. In addition, a beginning is made on the more general problem

Received January 24, 1958; revised June 18, 1958.

of how to choose x values for efficient polynomial estimation when one drops the assumption that the y's are uncorrelated.

3. Estimation methods. When a number of parameters are to be estimated simultaneously, the volume of the ellipsoid of concentration of the estimates is often used as a measure of the efficiency of the estimates. Since the square of the volume of the ellipsoid of concentration is proportional to the generalized variance of the estimates, one can just as well use the generalized variance as a measure of efficiency. This is the measure that will be used in this paper for making comparisons of different sets of estimates.

Suppose one wished to estimate the function $\lambda_0\beta_0 + \lambda_1\beta_1 + \cdots + \lambda_k\beta_k$, where the λ's are an arbitrary set of real numbers, by means of a linear estimate, $c_1y_1 + c_2y_2 + \cdots + c_ny_n$. Suppose further that the estimate is to be unbiased and possess minimum variance. Then it can be shown that the resulting estimates for the β's are given by the matrix formula

$$\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}y \tag{2}$$

where S is the covariance matrix of the y's and X is the matrix

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix}.$$

Furthermore, it can also be shown that the generalized variance of these estimates is given by the determinant formula

$$\text{G.V.} = |X'S^{-1}X|^{-1} \tag{3}$$

These same formulas will be obtained if one assumes that the y's possess a multivariate normal distribution and then finds the maximum likelihood estimates of the β's.

The advantage of the estimates given by formula (2) lies in the fact that it can be shown that among all linear unbiased estimates of the β's, the estimates given by this formula possess a minimum generalized variance. Thus, if one restricts himself to linear estimates, these are optimum estimates. All the comparisons to be made in the following sections will assume that the estimates are those given by formula (2), and hence that the generalized variance is given by formula (3).

4. Classical regression. Since the classical regression model assumes that $y_1, y_2, \cdots, y_n$ are uncorrelated with a common variance σ^2, the covariance matrix S is a diagonal matrix with elements σ^2.

Now De La Garza [2] has shown that the same information matrix, $X'S^{-1}X$, and hence the same value of the generalized variance, can be obtained by replacing a given set of n observations at the points $x_1, x_2, \cdots, x_n$ by a total of n

observations made at $k+1$ properly selected points in the interval from x_1 to x_n. These points will be denoted by $t_1, t_2, \cdots, t_{k+1}$ and the number of observations to be made at t_i will be denoted by n_i, where $\sum_{i=1}^{k+1} n_i = n$. In terms of these substitute observations, the matrices in (3) are all square matrices and therefore the determinant of their product can be obtained by taking the product of their determinants. As a result, (3) will assume the form

$$\frac{1}{\text{G.V.}} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix} \begin{vmatrix} \frac{n_1}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{n_2}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{n_{k+1}}{\sigma^2} \end{vmatrix} \begin{vmatrix} 1 & t_1 & \cdots & t_1^k \\ 1 & t_2 & \cdots & t_2^k \\ \vdots & \vdots & & \vdots \\ 1 & t_{k+1} & \cdots & t_{k+1}^k \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix}^2 \frac{\prod_{i=1}^{k+1} n_i}{\sigma^{2k+2}}.$$

But this determinant is a Vandermonde determinant with value $\prod_{i<j} (t_i - t_j)$; consequently

$$\frac{1}{\text{G.V.}} = \frac{1}{\sigma^{2k+2}} \prod_{i<j}^{k+1} (t_i - t_j)^2 \prod_{i=1}^{k+1} n_i \tag{4}$$

Since $\prod_{i=1}^{k+1} n_i$, subject to the restriction $\sum_{i=1}^{k+1} n_i = n$, is maximized when $n_1 = n_2 = \cdots = n_{k+1}$, it follows that the generalized variance will be minimized for a fixed set of values when the same number of observations is taken at each of the t values. This assumes that n will be chosen to make $n/(k+1)$ an integer.

Now consider the maximization of $\prod_{i<j} (t_j - t_j)^2$, subject to the restriction that $x_1 \leqq t_i \leqq x_n$, $i = 1, \cdots, k+1$. If x is transformed linearly so that this restriction assumes the form $-1 \leqq t_i \leqq 1$, $i = 1, \cdots, k+1$, then it is known [7] that the set of t values that maximizes $\prod_{i<j} (t_i - t_j)^2$ is given by the zeros of a polynomial which is the integral of one of the Legendre polynomials. These zeros can be obtained from the proper tables [8].

It is clear from inspecting the function $\prod (t_i - t_j)^2$ that the end points of the interval will always be chosen as two of the t values. It is also clear that the greater the range of x values, the smaller will be the generalized variance.

In view of the preceding results, it follows that optimum linear estimates of the coefficients of classical polynomial regression are obtained by using the estimates given by formula (2), choosing as large a range of x values as possible, taking observations at the $k+1$ points in this range given by means of the zeros of a tabulated polynomial, and repeating the experiment as many times as the total set, n, of observations will permit, with n chosen to make $n/(k+1)$ an integer.

The preceding optimum manner of choosing x values assumes that the generalized variance of the estimates of the coefficients of the regression polynomial is the proper measure of efficiency to use. If the sample regression polynomial curve is to be used exclusively for estimating ordinates of the theoretical regression polynomial curve, then one might prefer a measure of efficiency based on the variances and covariances of such estimated values. From this point of view, let $\tau_1, \cdots, \tau_{k+1}$ denote $k+1$ arbitrary points chosen in the given interval. Further, let α_i and $\hat{\alpha}_i$ denote the ordinate, and its estimate, of the polynomial regression curve at τ_i. Thus,

$$\alpha_i = \beta_0 + \beta_1\tau_i + \cdots + \beta_k\tau_i^k, \qquad i = 1, \cdots, k+1$$

and

$$\hat{\alpha}_i = \hat{\beta}_0 + \hat{\beta}_1\tau_i + \cdots + \hat{\beta}_k\tau_i^k, \qquad i = 1, \cdots, k+1.$$

Calculations will yield the covariance formula

$$m_{ij} = E(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) = \sum_{r=0}^{k}\sum_{s=0}^{k}\sigma_{rs}\tau_i^r\tau_j^s$$

where σ_{rs} is the covariance of $\hat{\beta}_r$ and $\hat{\beta}_s$. Since the generalized variance is the determinant of the covariance matrix, the generalized variance of the $\hat{\alpha}$'s will be equal to the determinant $|m_{ij}|$. But it will be observed that the matrix (m_{ij}) can be written in the form

$$(m_{ij}) = \begin{bmatrix} 1 & \tau_1 & \cdots & \tau_1^k \\ 1 & \tau_2 & \cdots & \tau_2^k \\ \vdots & \vdots & & \\ 1 & \tau_{k+1} & \cdots & \tau_{k+1}^k \end{bmatrix} \begin{bmatrix} \sigma_{00} & \cdots & \sigma_{0k} \\ \vdots & & \vdots \\ \sigma_{k0} & \cdots & \sigma_{kk} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tau_1 & \tau_2 & \cdots & \tau_{k+1} \\ \vdots & \vdots & & \vdots \\ \tau_1^k & \tau_2^k & \cdots & \tau_{k+1}^k \end{bmatrix}$$

Since $|\sigma_{rs}|$ is the generalized variance of the β's, it follows that

$$\text{G.V.}(\hat{\alpha}) = \text{G.V.}(\hat{\beta})\prod_{i<j}^{k+1}(\tau_i - \tau_j)^2$$

This result shows that the generalized variance of the estimates of the ordinates of a polynomial regression curve at $k+1$ arbitrary points in an interval will be minimized when the generalized variance of the estimates of the coefficients of the polynomial regression curve is minimized[1].

A recent paper by Guest [11], which was published after this paper had been submitted, has generalized the results of Smith [1] to polynomials of any degree. He shows that the values of $t_1, t_2, \cdots, t_n$ that minimize the maximum variance of a single estimated ordinate are given by means of the zeros of the derivative of a Legendre polynomial. It is easily seen that this set of values is the same set which minimizes the generalized variance above. Thus, whether one is interested

[1] I am indebted to Professor John Tukey for suggesting this relationship.

in efficient estimation of regression coefficients, or in efficient ordinate estimation, either at $k + 1$ points or one point, the optimum choice of t values is the same.

5. Comparison methods. When the assumption that the y's are uncorrelated is dropped, the problem of how best to choose the x's becomes very difficult. The choice will depend in a complicated manner upon the covariance matrix S. As a consequence, comparisons will be made only for equally spaced sets of points and only for three classes of covariance matrices. The sets of points that were selected for consideration are the following:

(1) n equally spaced points in the interval $(0, l)$

(2) $2n$ equally spaced points in the interval $(0, l)$

(3) $2n$ equally spaced points in the interval $(0, 2l)$

(4) two sets of observations of type (1).

A comparison of the relative advantages of choices (2), (3), and (4) over (1) will be made by comparing their generalized variances. Letting δ denote the interval between consecutive x values, these generalized variances will be denoted by G.V. (n, δ), G.V. $(2n, \delta/2)$, G.V. $(2n, \delta)$, and G.V. (2 runs), respectively.

The three classes of covariance matrices that will be studied are the following:

(a) uncorrelated variables, common variance

(b) $\rho(y_i, y_j) = e^{-\alpha|x_i - x_j|}$, $a > 0$, common variance

(c) covariance matrix of a pure birth stochastic process.

The first of these is the classical regression model considered in the preceding section. The second is the covariance matrix of a particular stationary stochastic process. The third was selected because it represents a stochastic process of the non-stationary type and in which the covariances grow as x increases. These three covariance matrices cover a rather wide range of correlation relationships and therefore conclusions obtained from them should have a rather wide range of application.

For comparison purposes it is advantageous to consider the following three ratios:

$$R_1 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta/2)}\right]^{1/(k+1)}$$

$$R_2 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta)}\right]^{1/(k+1)} \tag{5}$$

$$R_3 = \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (m) \text{ runs}}\right]^{1/(k+1)}$$

The reason for these choices is that it is easily shown that R_3 has the value m; consequently if the value of R_1, for example, should turn out to be m, it can be concluded that m runs of the basic experiment are needed to yield the same efficiency of estimation as that obtained by doubling the number of equally spaced observation points in the given interval. All comparisons will be made

in this manner, that is, by stating the number of runs of the experiment needed to yield the same efficiency as the choice of x values being considered.

6. Uncorrelated variables. It will be assumed that $n > k + 1$; consequently the X matrix in (3) will not be a square matrix and formula (4) will not be applicable. Under equal spacing in the interval $(0, l)$, the x values will be chosen as $x_i = i\delta$. As a result, the X matrix will assume the form

$$X = \begin{bmatrix} 1 & \delta & \cdots & \delta^k \\ 1 & 2\delta & \cdots & (2\delta)^k \\ \vdots & \vdots & & \vdots \\ 1 & n\delta & \cdots & (n\delta)^k \end{bmatrix} \tag{6}$$

Since S^{-1} is a diagonal matrix with elements $1/\sigma^2$, it is easily seen that (3) reduces to

$$\frac{1}{\text{G.V. } (n, \delta)} = \frac{\delta^{k\,(k+1)}}{\sigma^{2k+2}} \begin{vmatrix} n & \sum_1^n i & \cdots & \sum_1^n i^k \\ \sum_1^n i & \sum_1^n i^2 & \cdots & \sum_1^n i^{k+1} \\ \vdots & \vdots & & \vdots \\ \sum_1^n i^k & \sum_1^n i^{k+1} & \cdots & \sum_1^n i^{2k} \end{vmatrix}. \tag{7}$$

The value of this determinant is known [10] to be the polynomial displayed in (8); hence

$$\frac{1}{\text{G.V. } (n, \delta)} = \frac{\delta^{k\,(k+1)}}{\sigma^{2k+2}} An^{k+1}(n^2 - 1^2)^k(n^2 - 2^2)^{k-1} \cdots (n^2 - k^2) \tag{8}$$

where $A = (1!\,2! \cdots k!)^4 / (1!\,2! \cdots (2k + 1)!)$. The value of R_1 given in (5) then becomes

$$R_1 = \frac{1}{2^k}\left[\frac{(2n)^{k+1}(4n^2 - 1^2)^k \cdots (4n^2 - k^2)}{n^{k+1}(n^2 - 1^2)^k \cdots (n^2 - k^2)}\right]^{1/k+1}. \tag{9}$$

Using (8) and (5), it follows readily that

$$R_2 = 2^k R_1. \tag{10}$$

Now consider the limiting values of R_1 and R_2 as $n \to \infty$. The resulting values may be considered as asymptotic measures of efficiency. From (9) and (10) it follows that

$$\lim_{n\to\infty} R_1 = 2 \quad \text{and} \quad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

The first result implies that if one has a large number of equally spaced points in a fixed interval at which observations are made, then two runs of the experi-

ment will yield the same efficiency of estimation as doubling the number of equally spaced points in that interval. The second result implies, for example, that if the polynomial regression curve is of degree 4, then 32 runs of the experiment will be needed to yield the same efficiency of estimation as doubling the number of points by doubling the interval over which observations are to be made. It is clear from this second result that the higher the degree of the polynomial the more important it is to extend the range of x values as far as possible.

7. Stationary process model. Denoting the correlation between y_i and y_j by ρ_{ij}, it follows under equal spacing that the correlation function for model (b) will assume the form

$$\rho_{ij} = e^{-a|x_i - x_j|} = e^{-a\delta|i-j|}.$$

Letting $w = e^{-a\delta}$ and setting $\sigma^2 = 1$, since it will always cancel out in the R ratios, it will be seen that the covariance matrix here is given by

$$S = \begin{bmatrix} 1 & w & w^2 & \cdots & w^{n-1} \\ w & 1 & w & \cdots & w^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ w^{n-1} & w^{n-2} & w^{n-3} & \cdots & 1 \end{bmatrix}$$

Calculations will show that the inverse of S is given by

$$S^{-1} = \frac{1}{1-w^2} \begin{bmatrix} 1 & -w & 0 & \cdots & 0 & 0 \\ -w & 1+w^2 & -w & \cdots & 0 & 0 \\ 0 & -w & 1+w^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -w & 1 \end{bmatrix}$$

If S^{-1} is written as the sum of several matrices and then premultiplied by X' and postmultiplied by X, and finally brought together again into one matrix, it will be found that (3) assumes the form

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{(1-w^2)^{k+1}} \, |B(n, w)| \tag{11}$$

where $B(n, w)$ is the matrix whose element in row $p+1$ and column $q+1$ is given by

$$\begin{aligned} b_{p+1\,q+1} = {} & (w^2 + 1) \sum_1^n i^{p+q} \\ & - w \sum_2^n [i^p(i-1)^q + i^q(i-1)^p] - w^2[n^{p+q} + 1]. \end{aligned} \tag{12}$$

Since $w = e^{-a\delta}$, the value of G.V. $(2n, \delta/2)$ can be obtained by replacing n by $2n$, δ by $\delta/2$, and w by $\sqrt{w}$ in (11). As a result, it will follow that

$$R_1 = \frac{1 + w}{2^k} \frac{|\, B(2n, \sqrt{w}) \,|^{1/(k+1)}}{|\, B(n, w) \,|^{1/(k+1)}}.$$

Similarly,

$$R_2 = \frac{|\, B(2n, w) \,|^{1/(k+1)}}{|\, B(n, w) \,|^{1/(k+1)}}.$$

Now allow $n \to \infty$. From (12) it will be observed that the dominating part of $b_{p+1\,q+1}$ is $(w - 1)^2 \sum i^{q+q}$. As a result, the asymptotic value of the determinant $|\, B(n, w) \,|$ is

$$\begin{vmatrix} (w-1)^2 n & (w-1)^2 \sum i & \cdots & (w-1)^2 \sum i^k \\ (w-1)^2 \sum i & (w-1)^2 \sum i^2 & \cdots & (w-1)^2 \sum i^{k+1} \\ \vdots & \vdots & & \vdots \\ (w-1)^2 \sum i^k & (w-1)^2 \sum i^{k+1} & \cdots & (w-1)^2 \sum i^{2k} \end{vmatrix}.$$

But this is merely $(w - 1)^{2k+2}$ times the determinant in (7), which in turn has the asymptotic value $An^{(k+1)^2}$. From the preceding results, it follows that

$$\lim_{n\to\infty} R_1 = \frac{w + 1}{2^k} \frac{(\sqrt{w} - 1)^2 2^{k+1}}{(w - 1)^2} = \frac{2(w + 1)}{(\sqrt{w} + 1)^2}$$

and

$$\lim_{n\to\infty} R_2 = 2^{k+1}$$

For the purpose of seeing the implications of these formulas, consider the numerical value $w = e^{-a\delta} = .64$. This value implies that the correlation coefficient between neighboring y values is .64. Calculations yield the values

$$\lim_{n\to\infty} R_1 = 1.01 \qquad \text{and} \qquad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

Thus, doubling the number of observation points in a given interval, when there are already a large number of such points, gives practically no additional estimation information. The value of R_2, however, shows that the same asymptotic efficiency is gained here as in the case of uncorrelated variables. For correlated variables like those being considered in this section, it is clear that the interval over which observations are to be made should be extended as far as possible, but that if it can't be extended, repeating the experiment is far more efficient than taking additional observation points.

8. Pure birth process model. Although a pure birth process is a discrete process with an exponential regression curve, it was selected only for its covariance matrix properties which are quite different from those of the two preceding models.

If b denotes the constant asymptotic birth rate, y_0 the population size at time t_0, and y the population size at time $t > t_0$, then the conditional probability function for y, given y_0, is

$$P\{y_0, y; t_0, t\} = \binom{y-1}{y_0-1} e^{-by_0(t-t_0)}[1 - e^{-b(t-t_0)}]^{y-y_0}.$$

Using this formula, expected value calculations will show that the covariance of y_i and y_j, $j \geqq i$, is given by

$$\sigma_{ij} = y_0 e^{b(t_j-t_0)}[e^{b(t_i-t_0)} - 1].$$

Under equal spacing as before, $t_0 = 0$ and $t_i = i\delta$; hence letting $z = e^{b\delta}$,

$$\sigma_{ij} = y_0 z^j(z^i - 1).$$

From this formula it follows that

$$\sigma_{ij+m} = z^m \sigma_{ij} \quad \text{and} \quad \sigma_{jj} = \frac{z^j(z^j-1)}{z^i(z^i-1)}\sigma_{ii} \tag{13}$$

As a result, the covariance matrix S assumes the form

$$S = \begin{bmatrix} \sigma_{11} & z\sigma_{11} & \cdots & z^{n-1}\sigma_{11} \\ z\sigma_{11} & \sigma_{22} & \cdots & z^{n-2}\sigma_{22} \\ \vdots & \vdots & & \vdots \\ z^{n-1}\sigma_{11} & z^{n-2}\sigma_{22} & \cdots & \sigma_{nn} \end{bmatrix}$$

The second of formulas (13) enables this matrix to be expressed as the product of the following two matrices.

$$\begin{bmatrix} \frac{\sigma_{11}}{z-1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_{22}}{z^2-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_{nn}}{z^n-1} \end{bmatrix} \begin{bmatrix} z-1 & z^2-z & \cdots & z^n - z^{n-1} \\ z-1 & z^2-1 & \cdots & z^n - z^{n-2} \\ \vdots & \vdots & & \vdots \\ z-1 & z^2-1 & \cdots & z^n - 1 \end{bmatrix}$$

Some rather lengthy calculations will show that the inverse matrix is given by

$$S^{-1} = \frac{1}{y_0(z-1)} \begin{bmatrix} z+1 & -z & \cdots & 0 & 0 \\ -1 & z+1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & \cdots & z+1 & -z \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{z} & 0 & \cdots & 0 \\ 0 & \frac{1}{z^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{z^n} \end{bmatrix}$$

Additional lengthy computations, similar to those employed in the preceding section, will show that (3) assumes the form

$$\frac{1}{\text{G.V.}(n, \delta)} = \frac{\delta^{k(k+1)} \mid C(n, z) \mid}{y_0^{k+1}(z-1)^k z^n}$$

where $C(n, z)$ is the matrix whose element in row $p+1$ and column $q+1$ is given by

$$c_{p+1\,q+1} = \frac{(2^p - 1^p)(2^q - 1^q)}{z} + \frac{(3^p - 2^p)(3^q - 2^q)}{z^2} + \cdots + \frac{(n^p - (n-1)^p)(n^q - (n-1)^q)}{z^{n-1}}.$$

For $p = 1$ or $q = 1$, c_{p+1q+1} is defined by $c_{11} = z^n / (z-1) - 1$, and $c_{i1} = c_{1i} = n^{i+1} - 1$, $i > 1$.

When $n \to \infty$, the elements of this matrix exclusive of those in the first row and first column, converge to functions of z, for $z > 1$. Let

$$g_{pq}(z) = \lim_{n\to\infty} c_{p+1\,q+1}.$$

Since, for $z > 1$, c_{11} dominates c_{i1}, $i > 1$, the determinant $\mid c(n, z) \mid$ will possess the asymptotic value

$$\frac{z^n}{z-1} \begin{vmatrix} g_{11}(z) & \cdots & g_{1k}(z) \\ \vdots & & \vdots \\ g_{k1}(z) & \cdots & g_{kk}(z) \end{vmatrix}$$

For $k < 5$ it has been shown that the preceding determinant has the value

$$\frac{cz^{k(k-1)/2}}{(z-1)^{k^2}}$$

where c depends on k but not on z. Using these results the asymptotic value of the generalized variance is given by

$$\frac{1}{\text{G.V.}(n, \delta)} = \frac{\delta^{k(k+1)} c z^{k(k-1)/2}}{y_0^{k+1}(z-1)^{k^2+k+1}}. \tag{14}$$

From this result it is easily shown that

$$\lim_{n\to\infty} R_1 = \frac{(\sqrt{z}+1)^{(k^2+k+1)/(k+1)}}{2^k(\sqrt{z})^{(k_2-k)/(2k+2)}}.$$

Since (14) does not involve n, it follows that

$$\lim_{n\to\infty} R_2 = 1.$$

As a numerical illustration here, let $z = e^{b\delta} = 10/9$. This value implies that the correlation between y_1 and y_2 is approximately .7 and increases between

neighboring y values as one moves out on the axis. Calculations here yield the following limiting values for R_1.

k	1	2	3	4
$\lim_{n\to\infty} R_1$	1.47	1.32	1.25	1.20

These limiting values of R_1 show that some additional estimation information is gained by doubling the number of points in a fixed interval but that repeating the experiment yields considerably more information. The limiting value of R_2 would seem to indicate that no additional information is gained by extending the interval. This limiting result, however, is not realistic for small samples as will be seen in the next section.

9. Numerical results. Since the asymptotic measures of estimation efficiency obtained in the preceding sections may not be very realistic for small numbers of observations, some numerical computations were made with the assistance of high speed computing equipment. The values of $w = .64$ and $z = 10/9$ used previously were used in these computations. Values of $n = 5$ and $n = 10$ were chosen but only the results for $n = 10$ are given because some of the $n = 5$ values appeared questionable and because there were only moderate differences between the two sets of values. The limiting values of R_1 and R_2 are shown in parentheses adjacent to the computed values. In these computations, adjustments were made in the values of R_1 and R_2 to allow for the fact that doubling the number of points in an interval extends the total interval spanned by the points when the first point is located at $x = \delta$. These adjustments essentially kept the spanned interval unchanged. This was accomplished by replacing $\delta/2$ by $\delta(n-1)/(2n-1)$ in the denominator of R_1 and δ by $\delta(2n-2)/(2n-1)$ in the denominator of R_2.

k	Model (a)	Model (b)	Model (c)
1	1.90 (2)	1.03 (1.01)	1.43 (1.47)
2	1.81 (2)	1.02 (1.01)	1.25 (1.32)
3	1.72 (2)	1.02 (1.01)	1.19 (1.25)
4	1.64 (2)	1.03 (1.01)	1.18 (1.20)

k	Model (a)	Model (b)	Model (c)
1	3.80 (4)	2.91 (4)	1.83 (1)
2	7.24 (8)	5.01 (8)	1.85 (1)
3	13.76 (16)	8.94 (16)	3.21 (1)
4	26.24 (32)	16.41 (32)	6.15 (1)

It will be observed that the asymptotic values of R_2 are poor approximations for models (b) and (c). These results seem to indicate that in general one should always attempt to extend the range over which observations are to be taken as far as possible and the higher the degree of polynomial the greater is the advantage. They also seem to indicate that if the range can't be extended, it is considerably more efficient to replicate the experiment than double the number of observations, particularly if the variables are strongly correlated.

REFERENCES

[1] K. SMITH, "On the standard deviations of adjusted and interpolated values of an observed polynomial function," *Biometrika*, Vol. 12 (1918), pp. 1–85.

[2] A. DE LA GARZA, "Spacing of information in polynomial regression," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 123–30.

[3] G. ELFVING, "Optimum allocation in linear regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–62.

[4] H. CHERNOFF, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.

[5] H. E. DANIELS. "A distribution free test for regression parameters," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 499–513.

[6] S. EHRENFELD, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.

[7] I. SCHUR, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koefficienten," *Math. Zeit.*, Vol. 1 (1918), p. 378.

[8] G. PREVOST, *Tables et Formules de Fonctions Spheriques*, Gauthier-Villars, Paris, 1933.

[9] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 494.

[10] T. MUIR, *Treatise on the Theory of Determinants*, Longmans, Green and Co. (1933), p. 440.

[11] P. G. GUEST, "The spacing of observations in polynomial regression," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 294–299.