A MODEL FOR OPTIMIZING EXPERIMENTAL DESIGNS FOR ESTIMATING RESPONSE SURFACES

John Conlisk, University of California, San Diego
Harold Watts, University of Wisconsin

I. INTRODUCTION

Among many problems and challenges coming from the Graduated Work Incentive Experiment, one in particular has provided an opportunity for the authors, as econometricians, to round out their statistical education; and, perhaps, to make a contribution to the literature on experimental design. The problem we faced is a familiar economic one--how to get the most of some desirable output from limited inputs of financial and other resources, while observing various additional constraints.

For the graduated Work Incentive Experiment, the desirable output was precision in estimating the effects of alternative income maintenance policies. The principal scarce input was money; but, because of earlier choices regarding the scale and structure of the experiment, the number of families included in the experiment was also treated as a scarce input. Among the additional constraints on the problem were bounds on the guarantee levels, marginal tax rates, and pre-experiment income levels of the families in the experiment. It was also decided that an index of policy-makers' interest in alternative income maintenance policies should affect the design. These factors, together with specifications of an appropriate behavioral response function, have been fitted into a tractable mathematical model from which optimal allocation of the scarce experimental funds and families can be derived.

The model itself is not specific to the Graduated Work Incentive Experiment; rather it may be viewed as a general experimental design model for regression analysis. The model is both operational and quite flexible. It is operational in the sense that it reduces to a problem of minimizing a convex non-linear objective function subject to a set of linear constraints, a problem for which efficient computer solutions are available. The model is flexible in the sense that it can handle problems of arbitrary size, arbitrary regions of observation, alternative response functional forms, multiple response functions, varying experimental objectives, unequal costs per observation, unequal error variances, multiple constraints on the design, and other variations.

Section II sets out the mathematical bare bones of the model. Section III discusses the application of its various components to the Graduated Work Incentive Experiment.

II. THE MATHEMATICAL MODEL

Suppose an N-observation sample must be designed for estimting a response function

$$(1)\quad y_r = f(z_{r1},\ldots,z_{rk}) + e_r \quad r = 1,\ldots,N$$

where the context is as follows. The z_{ij} are design variables; they are subject to experimental control either by stratification (as when families of given income level are chosen in a tax experiment) or by direct control (as when payment levels are set in such an experiment). The y_r are the observable responses; and the e_r are (at least partially) unobservable random errors. The cost of a given observation may vary with the levels of $(z_{r1},\ldots,z_{rk})$ for that observation; and there is a maximum budget C that can be spent. Observations are restricted to a given region in the k-dimensional design space of the design variables. Finally, the function f is linear in parameters, and tractable assumptions for the e_r are allowed; so (1) is a standard regression equation.

A. The Basic Model

The building blocks of the basic model are 1) the regression model, 2) the admissable regressor rows, 3) the objective function, and 4) the budget constraint.

1. The regression model for equation (1) is

$$(2)\quad \begin{aligned} y &= X\beta + e \\ E(e) &= 0 \qquad V(e) = \sigma^2 I \\ b &= (X'X)^{-1}X'y \qquad V(b) = \sigma^2(X'X)^{-1} \end{aligned}$$

where y is the dependent variable vector, X the regressor variable matrix, e the error vector, β the coefficient vector, b the least squares estimate of β, and V(.) the variance matrix operator. It is assumed that X is of full column rank. The regressor matrix X depends, row for row, on the design matrix $Z = [z_{ij}]$, but is not in general the same. For instance, with k = 2, it might be that $(x_{r1}, x_{r2}, x_{r3}) = (1, z_{r1}, \log(z_{r1}/z_{r2}))$.

2. The admissable regressor rows may be introduced as follows. The design problem is to choose Z, and thus X, in some optimal way. Each row of Z represents an observation on the design variables. Complete freedom in choosing rows is not allowed. Instead, it will be assumed that there are a fixed number of admissable rows, each of which may be represented a number of times in Z. Corresponding to each admissable row for Z is an admissable row for X. Let m be the number of such admissable rows for X, x_i be the ith of them, and n_i be the number of times x_i is represented in X. Then X is composed of n_1 rows like x_1, n_2 rows like x_2, and so on. The total sample size N and the regressor cross product matrix X'X are given by

$$(3)\quad N = \sum_{i=1}^{m} n_i \qquad X'X = \sum_{i=1}^{m} n_i x_i' x_i .$$

So the design problem of choosing Z, and thus X, is simplified to the problem of choosing the m non-negative integers $n_1,\ldots,n_m$, given a set of admissable rows. The admissable rows for the design matrix Z represent points in the design space, called design points. The experimenter chooses these design points so as to give the relevant region of the design space adequate coverage. For X to be of full rank, the x_i, when stacked into an m-row matrix, must be of full rank (and, as discussed below, an appropriate number of the n_i must be positive).

3. An objective function to optimize in choosing the n_i is required. Suppose the experimenter's goal is accurate estimation of a vector $P\beta$ of linear combinations of the elements of β. The best linear unbiased estimate of $P\beta$ is Pb. It is assumed that the experimenter wishes to minimize a weighted sum of the variances of the elements of Pb. This objective function may be written tr[WV(Pb)], where tr(.) is the trace operator and W is a diagonal weight matrix whose diagonal elements indicate the policy importances to the experimenter of the elements of $P\beta$. Substituting from (2) and (3) and multiplying by the constant σ^{-2} gives the objective function as used, call it ϕ. Letting D = P'WP,

$$\begin{aligned}\phi(n_1,\ldots,n_m) &= \sigma^{-2}\mathrm{tr}[WV(Pb)]\\ &= \sigma^{-2}\mathrm{tr}[P'WPV(b)]\\ &= \sigma^{-2}\mathrm{tr}[P'WP\sigma^2(X'X)^{-1}]\\ &= \mathrm{tr}[D(\sum_{i=1}^{m} n_i x_i' x_i)^{-1}]\ .\end{aligned}$$

4. The budget constraint of the basic design model is $\Sigma_{i=1}^{m} c_i n_i \le C$, where c_i is the cost of one observation at the ith design point (that is, with regressor row x_i) and C is the total available budget.

Given these building blocks, the basic design model may be simply stated as follows. With D = P'WP,

$$\text{minimize}\quad \phi(n_1,\ldots,n_m) = \mathrm{tr}[D(\sum_{i=1}^{m} n_i x_i' x_i)^{-1}] \tag{4}$$
$$\text{subject to}\quad \sum_{i=1}^{m} c_i n_i \le C,\quad n_1 \ge 0,\ \ldots,\ n_m \ge 0.$$

Strictly speaking, this is an integer programming problem, since the n_i are integers. Practically speaking, however, little will be lost in practice by treating the n_i as continuous in solving (4) and then rounding off. Economists may see the design problem (4) as analagous to a utility-maximization problem from consumer choice theory, where ϕ corresponds to an inverse measure of utility, the n_i to amounts of m goods, the c_i to prices, and C to available income.

Summarizing then, the basic design model for regression analysis requires the experimenter to specify 1) a regression model, 2) a set of design points and the corresponding regressor rows x_i, 3) two objective function matrices P and W, and 4) the costs c_i and budget C. Then he must solve the programming problem (4).

B. Further Discussion of the Model

1. Consider the derivatives and convexity of $\phi = \phi(n_1,\ldots,n_m)$. Let $S = (\Sigma_{i=1}^{m} n_i x_i' x_i)^{-1}$. Then

$$\begin{aligned}\partial\phi/\partial n_i &= \partial\,\mathrm{tr}(DS)/\partial n_i = \mathrm{tr}(D\partial S/\partial n_i)\\ &= -\,\mathrm{tr}(DSx_i'x_iS) = -\,\mathrm{tr}(x_iSDSx_i')\\ &= -\,x_iSDSx_i'\ ,\end{aligned}\tag{5}$$
$$\begin{aligned}\partial^2\phi/\partial n_i\partial n_j &= -x_i(\partial S/\partial n_j)DSx_i' - x_iSD(\partial S/\partial n_j)x_i'\\ &= (x_iSx_j')(x_jSDSx_i') + (x_iSDSx_j')(x_jSx_i')\\ &= 2(x_iSx_j')(x_iSDSx_j')\end{aligned}$$

Since $SDS = (W^{1/2}PS)'(W^{1/2}PS)$ is non-negative definite, then the first partials $\partial\phi/\partial n_i$ are non-positive, as expected; an increase in the sample can do no harm. (Note that $\partial\phi/\partial n_i$ can be zero, since D may be of less than full rank.) The convexity of ϕ can be proved by showing the matrix of second partials to be non-negative definite. Letting $\otimes$ denote Kronecker multiplication, $\partial^2\phi/\partial n_i\partial n_j$ can be conveniently restated

$$\begin{aligned}\partial^2\phi/\partial n_i\partial n_j &= 2(x_iSx_j')\otimes(x_iSDSx_j')\\ &= 2(x_i\otimes x_i)[S\otimes(SDS)](x_j\otimes x_j)'.\end{aligned}$$

Hence the entire matrix of second partials may be written

$$2\begin{pmatrix} x_1\otimes x_1 & & O\\ & \ddots & \\ O & & x_m\otimes x_m\end{pmatrix} \times \begin{pmatrix} S\otimes(SDS) & \cdots & S\otimes(SDS)\\ \vdots & & \vdots\\ S\otimes(SDS) & \cdots & S\otimes(SDS)\end{pmatrix} \times \begin{pmatrix} x_1\otimes x_1 & & O\\ & \ddots & \\ O & & x_m\otimes x_m\end{pmatrix}' \tag{6}$$

This product is non-negative definite if the central matrix is. But the central matrix may be alternately written $(uu')\otimes S\otimes(SDS)$ where u is a column of ones; and $(uu')\otimes S\otimes(SDS)$ is non-negative definite because the Kronecker product of non-negative definite and positive definite matrices is non-negative definite.

2. Solving the design problem (4) is not difficult since the Kuhn-Tuchker first order minimization conditions take the simple form

$$\begin{aligned}(\partial\phi/\partial n_i)/c_i &= \lambda \text{ for all } i \text{ with optimal } n_i > 0,\\ (\partial\phi/\partial n_i)/c_i &> \lambda \text{ for all } i \text{ with optimal } n_i = 0.\end{aligned}\tag{7}$$

Equations (7) say that all design points included positively in the optimal design have the same marginal effectiveness per dollar of cost, call it λ, in reducing ϕ; while all design points excluded have lesser effectiveness. λ is the shadow price of the budget constraint; it is negative and equals $\partial\phi/\partial C$, evaluated at the optimum. The budget constraint will of course hold with equality. The first order conditions (7) indeed assure a global minimum since ϕ is convex; though the minimum may not be unique, since ϕ is not strictly convex. A simple iterative solution procedure may be based on the idea of letting the relative sizes of the $(\partial\phi/\partial n_i)/c_i$ determine how the n_i shift up and down from iteration to iteration.

3. Four useful scale properties of the model may be stated. Noting first that the objective function ϕ is homogeneous of degree minus one in the n_i, consider the relations

$$\begin{aligned} n_i &= \alpha_i N , \\ N &= C/\sum_{i=1}^{m} \alpha_i c_i , \qquad (8) \\ \phi &= (1/C)(\sum_{i=1}^{m}\alpha_i c_i)\mathrm{tr}[D(\sum_{i=1}^{m}\alpha_i x_i' x_i)^{-1}]. \end{aligned}$$

The first of these defines the fractions α_i, which give the proportional allocation of the total sample N over the design points. The second uses the first to rewrite the budget constraint (with equality holding). The third uses the first two to rewrite the objective function. The scale properties are: First, for a given proportional allocation $(\alpha_1,\ldots,\alpha_m)$, a change in C will cause an equiproportionate change in ϕ. Second, for given α_i, an equiproportionate change in all the c_i will result in the same proportionate change in ϕ, and the same inversely proportionate changes in N and the n_i. Third, for given α_i, equiproportionate changes in C and all the c_i will leave ϕ, N, and the n_i unchanged. Fourth, the optimal α_i are independent of the value of C.

4. In constructing the P-matrix, there are many sensible choices the experimenter might make. Two rather neutral examples are

$$(9) \qquad P = I \quad \text{and} \quad P = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{pmatrix} .$$

The first choice would imply that the experimenter was interested in estimating the elements of $P\beta = \beta$ themselves. The second choice would imply that he was interested in estimating the heights of the response function over the m design points. In making the choice, no firm constraint need be put on the number of rows in P, though several things about this may be noted. If P is not of full row rank, then it may always be condensed row-wise until it is of full row rank; so P need never have more rows than columns. That is, a P_o with full row rank and a corresponding W_o may always be found such that $P'WP = P_o'W_oP_o = D$ in (4), regardless of the rank or the number of rows in P. If the condensed matrix P_o has fewer rows than columns, there can be trouble. For example, suppose the one-row P-matrix $P = P_o = x_1$. Choice of this P would imply that the experimenter was solely interested in estimating the height $P\beta = x_1\beta$ of the response function over the first design point. In this case, the optimal design would be to put all observations at the first design point; so $n_1 = N$ and $n_2 = \ldots = n_m = 0$. This would mean, for an x_1 with more than one element, that the inverse in (4) would not exist; so the objective function would break down. This breakdown is not necessary whenever P_o has fewer rows than columns; it may or may not happen, depending on the values of P_o and the x_i. When it does happen, it is essentially because the experimenter has specified a regression form more complicated than he really wishes to estimate, as indicated by his choice of P. So the solution is to simplify the regression form and/or to increase the row rank of P. For either of P-specifications (9), of course, the compacted matrix P_o is non-singular; so the breakdown will not occur. (The P_o of this paragraph is only for discussion purposes; it need not be solved for in practice.)

5. An explicit solution for a one-way analysis of variance model is available. Suppose the x_i take the form

$$x_1 = (1,0,\ldots,0), \quad x_2 = (0,1,\ldots,0), \quad \ldots, \quad x_m = (0,0,\ldots,1).$$

Then the model is a one-way analysis of variance model where β_i and n_i are the mean of and the number of observations allocated to the ith cell, respectively. Letting $P = I$ as in the first of examples (9), and letting w_i be the ith diagonal element of W, the explicit solution to (4) for the optimal n_i is

$$(10) \quad n_i = (w_i/c_i)^{1/2} C/\Sigma_j (w_j c_j)^{1/2}$$

$$i = 1,\ldots,m.$$

6. Orthogonality of the regressor matrix X is not an optimality condition for the design model of this paper; though there is a well known theorem in the design literature giving conditions under which orthogonality is optimal. It is useful to review this theorem and see why the conditions may not be met. For this discussion, suppose the columns of X are not functionally related; so orthogonality of X is at least possible. (For example, one column of X may not contain square terms of another column.) Further suppose the elements of X are expressed as deviations from a point in the center of the region of interest (except possibly for a column of ones). Now consider three assumptions. First, all observations cost the same amount c; so the total sample is fixed at $N = C/c$. Second, the size of the region of interest is determined by putting upper bounds a_j on the mean squares of the regressor variables. (That is, for all j, suppose the jth diagonal element of $X'X/N$ may not exceed a_j.)

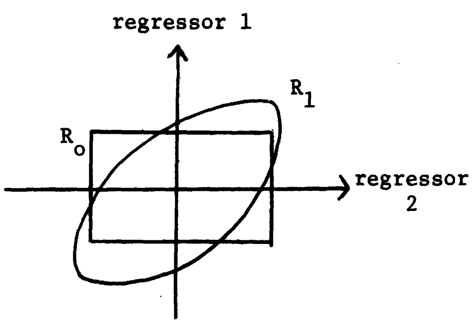
Third, minimization of the variances of the elements of b is the objective. Given these assumptions, the following derivation due to Tocher (1952) applies. Letting T be an upper triangular matrix such that $X'X = T'T$, and letting a double subscript on a matrix symbol denote the corresponding element, then

$$
\begin{aligned}
&\mathrm{var}(b_j) = \sigma^2(X'X)^{-1}_{jj} = \sigma^2(T'T)^{-1}_{jj} \\
&= \sigma^2(T^{-1}T'^{-1})_{jj} = \sigma^2\Sigma_i(T^{-1})^2_{ij} \\
(11)\quad &\geq \sigma^2(T^{-1})^2_{jj} = \sigma^2/T^2_{jj} \geq \sigma^2/\Sigma_i T^2_{ij} \\
&= \sigma^2/(T'T)_{jj} = \sigma^2/(X'X)_{jj} \geq \sigma^2/Na_j \ .
\end{aligned}
$$

It may be seen that the equalities in (11) will hold--and thus $\mathrm{var}(b_j)$ will achieve its lower bound--if and only if $X'X$, $T'T$, and T are diagonal, and the $(X'X/N)_{jj}$ are pushed to their bounds a_j. So optimality in terms of minimum $\mathrm{var}(b_j)$ requires orthogonality of X, given the three assumptions listed.

By seeing how the assumptions may not be met in the context of this paper, we may see how orthogonality may be non-optimal. First, the result will not hold in general if the cost of an observation varies with the values of the regressor variables. For example, if there are just two regressor variables (in addition possibly to a column of ones in X) as shown on Figure 1, and if the costs per observation are higher in the second and fourth quadrants than in the first and third quadrants, then the model will tend to set up a positive collinearity between the variables at the optimum. (A similar effect may result when further constraints (such as those discussed below) are added to the model.) Second, the result will not hold in general if the region of interest is determined in another way. The assumption above, by limiting each regressor variable separately, sets up a rectangular region like R_o in Figure 1; whereas the experimenter may in fact be faced by a region of interest

Figure 1



like R_1. In the latter case, the model will tend to set up a negative collinearity between the two regressor variables. Third, the result will not hold in general if the experimenter's objective is not minimization of the $\mathrm{var}(b_j)$ separately. For example, the experimenter may be more interested in $\mathrm{var}(b_1-b_2)$ than $\mathrm{var}(b_1)$ and $\mathrm{var}(b_2)$ separately; this interest can easily be expressed by appropriate choice of P. In such a case, the model would tend to set a negative correlation between regressor variables 1 and 2.

C. Extensions of the Model

The following extensions are considered singly, but they can easily be used in combination.

1. The choice of a specific regression functional form for the response function (1) is crucial to the character of the optimal design. For instance, if (1) were first degree polynomial, the optimal design would tend to concentrate only on design points around the boundary of the relevant design space region. If, however, (1) were third degree polynomial, the optimal design would tend to pick up some design points in the interior for identifying curvature. Unfortunately, the experimenter seldom knows ahead of time what the appropriate functional form is. Suppose the experimenter is considering q candidate functional forms, and suppose $\phi_j(n_1,\ldots,n_m)$ for $j = 1,\ldots,q$ are the corresponding objective functions. It is helpful to think of $\phi_j(n_1,\ldots,n_m)$ as a loss function where j indexes states of nature (functional forms) and $(n_1,\ldots,n_m)$ represents actions (sample designs). One way to investigate the sensitivity of the model to alternate functional forms is then to construct a qXq loss table, where the q states of nature are the q functional forms, and the q actions are the optimal designs for the q functional forms taken one at a time. In generating a design for ultimate use, the experimenter might use the Bayes strategy of minimizing the expected loss

$$(12)\quad G(n_1,\ldots,n_m) = \sum_{j=1}^{q}\pi_j\phi_j(n_1,\ldots,n_m)$$

where the π_j are prior probabilities for the states of nature, or functional forms. Replacing $\phi(n_1,\ldots,n_m)$ by $G(n_1,\ldots,n_m)$ in the design problem (4) above does not greatly complicate things. The sum of convex functions is convex; so G is convex. The derivatives (5) must be replaced by a π_j-weighted sums of expressions like (5). With appropriate minor modifications, all the other comments of the last section carry over.

2. Non-linearity in the parameters of the response function (1) may be simply handled by replacing the function with its Taylor series linearization about some guessed parameter values. This approach has been used by Box and Lucas (1959). Similarly, one might view the vector $P\beta$ as a linear approximation to some non-linear vector-valued function of β.

3. Multiple objectives in an experiment may be handled by a simple extension of the model. Suppose the experimenter wishes to estimate a set of response functions. (For instance, a medical experimenter might wish to estimate various physiological responses to controlled dieting.) Let the corresponding regression models be stated compactly as $Y = X\boldsymbol{\beta} + E$; where Y, $\boldsymbol{\beta}$, and E are matrices whose columns correspond to different regressions; and where the rows of E are independent with zero means and identical variance matrices $\sigma^2 U$. It is known (see Goldberger (1964)

pp. 201-12, 246-8) that $B = (X'X)^{-1}X'Y$ is the best linear unbiased estimate of $\boldsymbol{\beta}$; and that $V(b_B) = \sigma^2 U \otimes (X'X)^{-1}$, where b_B is the vector gotten by stacking the columns of B in order from the first down to the last. By analogy to the objective function construction above, suppose the experimenter's estimate of interest is the vector Pb_B of linear combinations of the elements of b_B; and suppose he wishes to minimize the weighted sum $tr[WV(Pb_B)]$ of the variances of elements of Pb_B, where W is a diagonal weight matrix. The objective function, call if J, may then be written (with $D = P'WP$)

$$(13) \quad J(n_1,\ldots,n_m) = tr[D(U \otimes (\sum_{i=1}^{m} n_i x_i' x_i)^{-1})] \quad .$$

Replacing $\phi(n_1,\ldots,n_m)$ by $J(n_1,\ldots,n_m)$ in the design problem (4) does not substantively change the programming problem to be solved. The experimenter, in setting up the problem now has the one additional task of specifying a value for U. If U is diagonal (which seems unlikely in practice), then the various response functions to be estimated are independent; and (13) can be put in the form of (12) with the ϕ_j representing objective functions for the separate response functions, and the π_j equalling the diagonal elements of U.

4. Alternative forms for the objective function are available, typically based on some function of the variance matrix $V(b)=\sigma^2(X'X)^{-1}$ (that is, typically based on some function of the "information matrix" X'X). Two convenient possibilities are the determinant, or "generalized variance," function $|V(b)|$ and the trace function $tr(V(b))$ used here (neglect the weight matrix D momentarily). These functions are convenient because they are continuous, differentiable, convex, and so on. (See Kiefer (1959) for a discussion of these and other possibilities.) Minimizing $|V(b)|$ may be conveniently rationalized by noting that, if b is normal, the volume of a confidence ellipsoid around b, for given probability of containing β, is proportional to $|V(b)|$. Minimizing $tr(V(b))$ may be conveniently rationalized by noting that $tr(V(b)) = E(b-\beta)'(b-\beta)$ (the expected value of a quadratic loss function) or by noting, as above, that $tr(V(b))$ is the sum of the variances of the regression coefficient estimates. In a sense, $|V(b)|$ and $tr(V(b))$ are not very different objectives. Let $\lambda_1,\ldots,\lambda_s$ be the eigenvalues of $V(b)$; they must all be positive. Then it is known that $|V(b)| = \lambda_1\ldots\lambda_s$ and $tr(V(b)) = \lambda_1+\ldots+\lambda_s$. So minimizing $|V(b)|$ is equivalent to minimizing the geometric mean of a set of positive numbers, and minimizing $tr(V(b))$ is equivalent to minimizing the arithmetic mean of the same numbers.

A determinant objective function is quite capable of carrying the design problems discussed here. However, the trace function is actually used because it is easier to weight. The trace function is a linear function of elements of $V(b)$ and thus takes simple linear weights; whereas the determinant function is multiplicative and thus does not take linear weights. More specifically, it makes sense to replace $tr(V(b))$ by $tr(DV(b)) = E(b-\beta)'D(b-\beta)$; whereas it does not make sense to replace $|V(b)|$ by $|DV(b)|$. D must be square or $|DV(b)|$ is not defined; and D must be nonsingular or $|DV(b)| = 0$. But, with D nonsingular, $|DV(b)| = |D||V(b)|$; so minimizing $|DV(b)|$ is the same as simply minimizing $|V(b)|$. On the other hand, if the experimenter does not wish to use weights, and if computational ease is important, then $|V(b)| = |\sigma^2(X'X)^{-1}| = 1/|\sigma^{-2}X'X|$ may be more appropriate because it does not involve matrix inversion.

5. Unequal error variances in the regression model can easily be handled. Suppose the error variance differs from design point to design point such that the error variance corresponding to the ith regressor row x_i is $\sigma^2 v_i$. It may be shown that the objective function then becomes $\phi(n_1,\ldots,n_m) = tr[D(\Sigma_{i=1}^m n_i v_i^{-1} x_i' x_i)^{-1}]$, which introduces a very minor change indeed in the design problem (4). Of course, the experimenter must specify the v_i.

6. Attrition of observations from the sample can arise when, for example, some families which initially agree to be part of a cross-family sample later drop out; or when, for example, some observations in a laboratory experiment are unusable due to experimenter error. Suppose the attrition fraction varies by design point such that the attrition fraction corresponding to the ith regressor row x_i is μ_i. This can be easily handled in the design model (4) by replacing the n_i by $\mu_i n_i$ in the objective function. If attrition also affects the costs c_i, then they must be adjusted. For example, a family which drops out of a cross-family experiment may cost less than a family which stays the duration. If c_i' and c_i'' are the costs of observations which do and do not stay in the sample, respectively, then the appropriate cost to enter in the design model (4) is $c_i = (1-\mu_i)c_i' + \mu_i c_i''$.

7. Budget minimization subject to a maximum error constraint may sometimes be the experimenter's problem rather than error minimization subject to a maximum budget constraint. In such a case, the problem corresponding to (4) would be

$$\begin{aligned} &\text{minimize} \quad C = c_1 n_1 + \ldots + c_m n_m \\ (14) \quad &\text{subject to} \\ &\phi(n_1,\ldots,n_m) = tr[D(\sum_{i=1}^{m} n_i x_i' x_i)^{-1}] \le \phi_o \\ &n_1 \ge 0 \; , \; \ldots \; , \; n_m \ge 0 \end{aligned}$$

where ϕ_o is a pre-selected maximum admissable error. Given a solution procedure for the original problem (4), solution of the new problem (14) is easy. Note that the optimization conditions (7) apply to (14) as well as (4), and that these m-1 conditions determine the m-1 independent $\alpha_i = n_i/N$. Also, recall that the optimal $\alpha_i = n_i/N$ are independent of C. Then a solution to (4) for any value of C will yield the optimal α_i for (14). And, given the optimal fractional allocation $(\alpha_1,\ldots,\alpha_m)$ for (14), it is easy to find the optimal absolute allocation $(n_1,\ldots,n_m)$ and budget C which make $\phi = \phi_o$.

8. Additional constraints on the design problem (4) may often be required. The simplest sort of addition would be to replace some or all of the zero bounds $n_i \geq 0$ by positive bounds $n_i \geq n_i^o > 0$. For example, the experimenter may start with observations numbering $(n_1^o, \ldots n_m^o)$ from some previous round of experimentation. This kind of additional constraint on (4) can be handled with only trivial modifications to the simple optimization conditions (7); so the design problem is still computationally straight-forward. More substantive additional constraints may also arise. For example, a fixed experimental capacity may add a total sample constraint $\Sigma_{i=1}^m n_i \leq N_o$. (Of course, the budget constraint may all along have been interpreted as a total sample constraint, where $c_1 = \ldots = c_m = 1$ and C is the maximum total sample. The current discussion refers to joint imposition of a budget and a total sample constraint.) Or a bureaucratic ruling may place a sub-budget constraint $\Sigma_{i=1}^p c_i n_i \leq C_p$ on the first $p < m$ design points. And so on. Such substantive additional constraints make the programming problem to be solved much more difficult. Nonetheless, if all the constraints are linear, efficient computer routines are available. Some types of non-linear constraints are, of course, also tractable.

III. THE APPLICATION

In the Graduated Work Incentive Experiment we are concerned with the response of family earnings to the changed alternatives produced by introduction of a negative income tax. Accordingly, we have specified as the dependent or response variable for a given family (the y_r in equation (1)) the ratio of the family's actual earnings during the experiment to a pre-experiment estimate of "normal" earnings. Calling the response variable R and dropping the subscript:

$$R = \frac{\text{actual earnings}}{\text{pre-experiment normal earnings}} .$$

The three independent or design variables specified were:

$$g = \frac{\text{maximum benefit (paid when earnings are zero)}}{\text{poverty level income}} .$$

$$t = \text{marginal tax rate}$$
$$= \text{reduction in benefit per dollar earned.}$$

$$w = \frac{\text{pre-experiment normal earnings}}{\text{poverty level income}} .$$

So the response function was of the form $R = f(g,t,w)$ (neglecting the error term). The first two design variables g and t are subject to direct experimental control for each family in the sample; they are parameters of the linear negative tax the family is faced with. The third design variable w must be controlled by stratification; families must be screened until ones of desired w-level are found.

There are assuredly more variables than g, t, and w which may affect the response variable R; although many variables which come to mind may operate principally through g, t, and w. For example, family size operates through g and w by affecting poverty level income in the denominators of g and w. Also, family size, as well as education, race, age, and so on, operate through the normal earnings variable w. Insofar as variables excluded from f(g,t,w) can be randomized by careful sampling procedures, they can be lumped into the error term (the e_r of equation (1)). So the regression model used for sample design is a very abbreviated version of the model one might eventually apply to the data produced by the experiment. The practical limit on the number of variables that can be handled in the design is far more stringent than the practical limit on the number of variables that can be measured for eventual analysis. Even screening enough families to get the desired stratification by w-level turned out to be quite difficult. In summary, it seemed to us that w was clearly the most important stratification variable to control, and that once it was controlled there did not seem to be any second variable of comparable importance.

The problem, then, was one of specifying a sample in the three dimensional design space of (g,t,w)-triplets. Sampling was restricted to a region within the design space which provided substantial variation in (g,t,w), but which kept to (g,t,w)-combinations of actual policy interest. Within this region of interest, twenty seven design points were selected (so m = 27). There were nine (g,t)-combinations or treatments (one "control" combination with g = t = 0 and eight non-zero combinations) at each of three w-levels. So the design problem reduced to finding optimal numbers $n_1, \ldots, n_{27}$ of families to allocate to each design point.

A crucial part of defining optimality of a design is specification of a regression functional form. In the Graduated Work Incentive Experiment, numerous alternative transforms and combinations of g, t, and w were used to provide f(g,t,w) functions which had both linearity-in-parameters and varying degrees of non-linear flexibility in g, t, and w. The functional forms used had from 6 to 13 parameters; so a substantial degree of non-linearity in g, t, and w was allowed for.

For a given regression functional form, the variance matrix $V(b) = \sigma^2(\Sigma_i n_i x_i' x_i)^{-1}$ of the parameter estimates is easily obtained for any specific allocation of families to design points (that is, any choice of $n_1, \ldots, n_{27}$). The remaining problem in defining optimality of design lies in specifying a scalar-valued function of V(b) to optimize. In terms of the optimand $\phi(n_1, \ldots, n_m) = \sigma^{-2} tr[P'WPV(b)]$ introduced above, this requires specification of the matrices P and W. Recall that $P\beta$ is the assumed vector of magnitudes-to-be-predicted and thus Pb is the estimate of interest.

In the Graduated Work Incentive Experiment, the assumed objective was taken to be estimation of the incremental treasury cost of a linear negative tax due to induced reduction of work effort and earnings--that is, the difference between the cost assuming zero work reduction and the cost given the actual work reduction. (The possibility of negative work reduction--work increase--is fully allowed for.) The cost referred to is the cost for the entire country of a national negative tax. Given an estimate of the response function parameters, the work response and thus the desired incremental cost can be estimated for any individual family. By summing over all families, the national cost can be estimated. Such a cost estimate depends on the specific negative tax parameters g and t assumed as well as on the parameter estimate b; so the cost might be denoted H(g,t,b). (The variable w has "integrated out" in the summation.) With a few approximative tricks, this estimate can be expressed as a linear function of b, call it h(g,t)b where h(g,t) is a row vector depending on g and t. The rows of P were set equal to the values of h(g,t) for various policy-relevant combinations of g and t. So the elements of the vector Pb are estimated incremental treasury costs due to induced earnings response for various (g,t)-combinations. A policy-importance weight was specified for each (g,t)-combination, and thus each element of Pb; these weights were arrayed in the diagonal matrix W. This completed the specification of the objective function $\phi(n_1,\ldots,n_m) = \sigma^{-2}\mathrm{tr}[P'WPV(b)]$.

This objective function is to be minimized by appropriate allocation of families to the 27 design points--that is, by appropriate choice of $n_1,\ldots,n_{27}$. But this allocation was constrained by several further considerations. The principal constraint was the budget constraint $\Sigma_i c_i n_i \leq C$. The costs c_i were composed of administrative costs, which are relatively constant over design points (costs of screening families, administering questionnaires, mailing checks, and so on), and of negative tax payment costs, which vary widely over design points. A curious feature of this experiment is that the cost of payments to families is both an ingredient of the design problem and in essence what the experiment is designed to estimate. Nevertheless, we have some well-founded notions about the relative sizes of these costs, which do not vary widely even when assumed earnings response patterns do vary widely.

In addition to the budget constraint, there were sample size constraints, both on the total sample and on sub-samples for various w-levels. There was also a constraint on the fraction of the budget that could be allocated to a particular high-benefit (g,t)-combination. Finally, most of the design points had positive rather than zero lower bounds ($n_i \geq n_i^o > 0$) since a subgroup of families had been allocated to the design points prior to the use of the optimization model.

Allowance was also made for attrition from the sample of families at low payment design points, and for higher error variances of families at high payment design points.

These specifications combine to make up a programming problem involving 27 variables, a nonlinear objective function, and typically five substantive linear constraints in addition to the 27 lower bound constraints. A computer routine due to Kreuser (1968) solved this problem in about 10 to 15 minutes on a Burroughs 5500. Of course, the problem was solved many times over as specifications were polished and sensitivities to changed specifications were tested. In addition to solutions for this problem, a small number of solutions were generated for a larger problem. A distinction was made between two locations at which families were being sampled; so there were 27 design points for each of two locations, or 54 design points and corresponding n_i altogether. The number of constraints and the number of terms in the regression function were also increased for this enlarged problem. This raised the solution time on the Burroughs 5500 to the neighborhood of 45 minutes.

The final solution of the Graduated Work Incentive design had the following noteworthy characteristics. (i) The optimal allocation of experimental families produces a decidedly non-orthogonal design. (ii) Several of the design points were allocated no observations or the minimum possible observations (given the positive lower bounds). (iii) The majority of the budget was allocated to a few high-payment design points; and the majority of the total sample was allocated to low-payment design points. (iv) The optimal designs generated by the model were substantially more efficient than various intuitive designs discussed before the optimization model was used. (Since the objective function ϕ is a variance magnitude, the relative efficiency of two designs can be measured in the usual sense by taking the ratio of their ϕ-values.)

Summing up our experience with the model in this, its first, application, we have found it very useful for incorporating explicitly a great many considerations which affect an experimental sample allocation. Indeed, it is difficult to see how these matters could be properly reflected by any less formal (or rule of thumb) procedure. The basic model will be used in the design of several upcoming social experiments, and its generality is such that it can be adapted to many experimental situations which are characterized by several dimensions of experimental control, and where continuity of response suggests some form of estimated regression surface.

ACKNOWLEDGMENTS

The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin pursuant to the provisions of the Economic Opportunity Act of 1964. Thanks for helpful comments on this material are due to numerous members of the Institute for Research on Poverty and of Mathematica Inc. (of Princeton, New Jersey), as well as to Richard Day and Arthur Goldberger. Thanks are also due to Felicity Skidmore for editorial help and to Jerome Kreuser for non-linear programming help.

REFERENCES

[1] G. E. P. Box and H. L. Lucas (1959), "Design of Experiments in Non-Linear Situations," *Biometrika*, Vol. 46, pp. 77-90.

[2] Arthur Goldgerger (1964), *Econometric Theory*, New York: Wiley, 1964.

[3] J. Kiefer (1959), "Optimum Experimental Designs," *Journal of the Royal Statistical Society Series B*, Vol. 21, pp. 273-319.

[4] Jerome Kreuser (1968), *Users and Programmers Manuals for GPM*: University of Wisconsin, Social Systems Research Institute Programming Series, 1968.

[5] K. D. Tocher (1952), *Biometrika*, Vol. 39.