View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 3, number 3

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-3

Publication Date: July 1974

Chapter Title: Optimal Response Surface Design in Monte Carlo Sampling Experiments

Chapter Author: John Conlisk

Chapter URL: http://www.nber.org/chapters/c10172

Chapter pages in book: (p. 17 - 28)

Annals of Economic and Social Measurement, 3/3, 1974

OPTIMAL RESPONSE SURFACE DESIGN IN MONTE CARLO SAMPLING EXPERIMENTS*

BY JOHN CONLISK

The theory of optimal experiment design is applied to the design of Monte Carlo sampling experiments. It is shown that prevailing practice in Monte Carlo sampling experimentation may lead to inefficient use of computer time. In particular, the practice of generating the same number of samples for each numerical specification of the target model is criticized. Alternative procedures are suggested.

I. INTRODUCTION

As Summers (1965) has pointed out, the increasing availability of computers argues for greater use of the "capital intensive" Monte Carlo approach to discovering small sample properties of estimators.[1] Of course, the capital should be used efficiently. Toward this end, Haitovsky and Jacobs (1972) presented in the initial issue of the *Annals* a general purpose program for efficient generation of Monte Carlo sampling data. Given such a program, further efficiency questions concern the number of Monte Carlo samples to generate, the sample size for each, the numerical specification of target model parameters, and related questions. These further efficiency questions may be linked with the statistical literature on optimal experiment design. This paper shows that efficient design of certain aspects of a Monte Carlo study may be viewed as a regression design problem. Application of regression design techniques suggests that some prevailing practices in Monte Carlo sampling experimentation use computer time inefficiently.

A recent Monte Carlo study by Orcutt and Winokur (1969) makes a useful illustration. Though it will be argued that the authors used computer time inefficiently, this may be no more than a trivial criticism of their study. Since their model was very simple, computer cost may have been a minor consideration. It is the simplicity and elegance of their model which makes it a convenient illustration. Nonetheless, computer cost is a serious consideration in many Monte Carlo sampling studies; and the points made carry over. Orcutt and Winokur were concerned with the autoregressive model

$$Y_t = \alpha Y_{t-1} + e_t, \quad \text{where } e_t \text{ is NID}(0, 1). \tag{1}$$

Here Y_t and e_t are scalar random variables, and t indexes time. As the authors note, no generality is lost by normalizing the intercept of (1) to zero and the variance of e_t to one.

Orcutt and Winokur wished to know the small sample properties of alternate estimators and test statistics associated with the model (1), especially the first two

* This research was supported by the National Science Foundation under grant GS-3201. Thanks are due to the editor and referees of the *Annals* for helpful comments on an earlier version of the paper.

[1] Monte Carlo methods are treated in the statistical literature at a more basic level than in most econometric applications. See references in the Naylor *et al.* survey (1967).

moments of these statistics. For example, they were interested in the functions

$$E(\alpha^*) = F(\alpha, N) \quad \text{and} \quad mse(\alpha^*) = G(\alpha, N). \tag{2}$$

Here $mse(.)$ denotes mean square error, and α^* denotes the ordinary least squares estimate of α based on an N-observation sample $(Y_1, \ldots, Y_N)$. Since the functions F and G could not be derived analytically, the authors estimated them empirically—that is, performed a Monte Carlo experiment. Their procedure can be described in terms of equation estimation (though the authors did not so describe the procedure).

Corresponding to the functions (2) to be estimated are two estimation equations:

$$\alpha_r^* = F(\alpha_r, N_r) + u_{Fr} \quad \text{and} \quad (\alpha_r^* - \alpha_r)^2 = G(\alpha_r, N_r) + u_{Gr}, \tag{3}$$

for $r = 1, \ldots, n$. Here the u's are error terms: they must have zero means in view of (2). The subscript r indexes observations. "Observation" needs explaining. The r-th observation was generated by the authors in three steps. (i) They specified values for α_r and N_r. (ii) They generated by computer a sample time series for Y_t of size N_r using the autoregressive model with $\alpha = \alpha_r$. (iii) They calculated an observation α_r^* on the estimator α^* (along with observations on all other estimators and test statistics under consideration). These three steps yield a value for the triplet $(\alpha_r^*, \alpha_r, N_r)$; repeating the procedure n times yields a sample of size n for estimating equations (3).

Viewing the authors' procedure in this way is useful in establishing a link with the experiment design literature. The authors had control over the independent variables α_r and N_r in the estimation equations (3). Their design problem was how to choose the n independent variable observations $(\alpha_1, N_1), \ldots, (\alpha_n, N_n)$. Such a choice problem is roughly speaking what the design literature is all about. Since n was a very large number (in the tens of thousands), the authors followed a standard design procedure by selecting a much smaller number, say m, of admissible values for a pair (α_r, N_r); and they then took repeated observations at each of these admissible pairs. Thus, the problem of choosing all $2n$ numbers $(\alpha_1, N_1), \ldots, (\alpha_n, N_n)$ reduced to the more tractable problem of choosing only m numbers—the numbers of observations at each of m admissible pairs. Call these numbers of observations $n_1, \ldots, n_m$; they sum to n.

Some widely used terminology is useful. The variables which an experimenter controls (α and N in the Orcutt–Winokur context) are called *design variables*. Their space is called the *design space*. The m admissible sets of values for the design variables are called *treatments* or *design points*. The numbers of observations per treatment $n_1, \ldots, n_m$ are called *treatment sample sizes*; and a set of values for the n_i is called a *design*. The functions to be estimated (like F and G from (2)) are called *response functions*. The object of a Monte Carlo sampling experiment is accurate estimation of the response functions; and the design question is how to choose a design which best serves this object. The most common design used in Monte Carlo sampling studies, including the Orcutt–Winokur study, is the equal treatment size design defined by $n_1 = \ldots = n_m$. (Few exceptions to the equal treatment size design were found in the literature; and these had no apparent optimal design rationalization.) It will be argued here that such a design uses computer time inefficiently. It

will be argued that the n_i usually should be made systematically unequal, according to guidelines suggested below.

An answer to the design question depends heavily on the functional form of response functions. Implicit in most Monte Carlo sampling studies is a disregard for the continuity of a response function. Orcutt and Winokur, for example, estimated the height of $F(\alpha_r, N_r)$ over each treatment point as the sample mean of the α_r^* observed at that point. Thus, in estimating the height, they made no formal use of available information about heights over adjacent treatments, as if F had no continuity. The spirit of this approach is captured by the step function specification

$$F(\alpha_r, N_r) = \begin{cases} \beta_1 \text{ if } (\alpha_r, N_r) \text{ falls at treatment 1,} \\ \beta_2 \text{ if } (\alpha_r, N_r) \text{ falls at treatment 2,} \\ \vdots \\ \beta_m \text{ if } (\alpha_r, N_r) \text{ falls at treatment } m. \end{cases} \tag{4}$$

In truth, a function like F is typically continuous; possibly F could be well approximated by the quadratic

$$F(\alpha_r, N_r) = \beta_1 + \beta_2\alpha_r + \beta_3 N_r + \beta_4\alpha_r^2 + \beta_5 N_r^2 + \beta_6\alpha_r N_r. \tag{5}$$

The two functional forms (4) and (5) will be used in the illustrative calculations below.

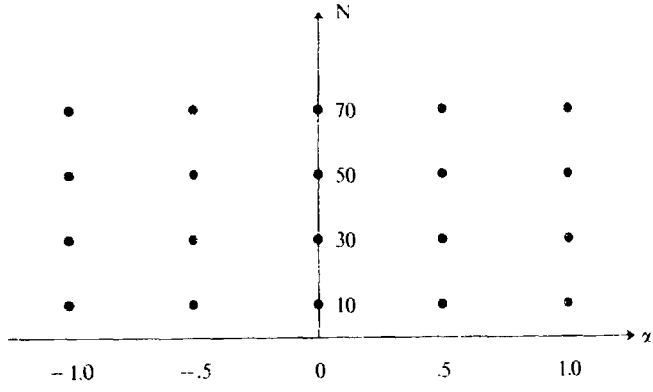

Figure 1

Orcutt and Winokur ran several experiments, using up to 48 treatments. For simplicity, illustrative calculations here will use $m = 20$ treatments, as described by the points on the design space of Figure 1. These treatments cover the range of stationarity ($-1 < \alpha < 1$) for the autoregressive model and a range of sample sizes ($10 \leq N \leq 70$) commonly faced in time series studies.

II. The Monte Carlo Model

A source of confusion in this context is the presence of two statistical models; they will be called the *target model* and the *Monte Carlo model*. The target model is the model of ultimate concern—the autoregressive model in the Orcutt–Winokur study. The Monte Carlo model is the estimation model—composed of equations like (3)—used to translate computer runs into inferences about the distributions of the target model estimators and test statistics. This section formally specifies the Monte Carlo model. Since each of the two models has its own parameters, estimators, and other components, one must continually delineate which model words like "parameter," "estimator," and "observation" refer to. Fortunately, the Monte Carlo model for the general case can be specified without specifying the target model for the general case.

In the Orcutt–Winokur illustration, there is a pair of design variables (α, N). For each Monte Carlo observation, the pair is assigned one of m admissible values. In the general case, let the target model have $D - 1$ parameters, which together with N, make D design variables. Further, let the m admissible values for the D-tuple of design variables be written as $(1, D)$ row vectors $z_1, \ldots, z_m$; so the z_i are the treatments.

The Monte Carlo experimenter's goal is to estimate a set of response functions using estimation equations like (3). For now, it is assumed that these equations satisfy the linearity and other assumptions of the following standard multi-equation regression model. The restrictiveness of this assumption, and how to relax it, are discussed in the next section.

$$
\begin{aligned}
&Y = X\beta + U, \\
(6)\qquad &E(U) = 0, \qquad \operatorname{var}(U_v) = V \otimes Q, \\
&Q = \operatorname{diag}(\underbrace{\sigma_1^2, \ldots, \sigma_1^2}_{n_1 \text{ times}} : \ldots : \underbrace{\sigma_m^2, \ldots, \sigma_m^2}_{n_m \text{ times}}).
\end{aligned}
$$

This is the Monte Carlo model. Here Y is an (n, p) matrix of observations on dependent variables, where p is the number of response functions, or regressions. The columns of Y correspond to different response functions and the rows to different Monte Carlo observations. In the Orcutt–Winokur context, for example, two of the columns of Y are $[\alpha_1^*, \ldots, \alpha_n^*]'$ and $[(\alpha_1^* - \alpha_1)^2, \ldots, (\alpha_n^* - \alpha_n)^2]'$, which are the dependent variable observations for equations (3).

The dependent variable matrix Y is composed of a systematic component $X\beta$ and an error component U. In the assumptions (6) on the (n, p) error matrix U, var(.) denotes "variance matrix of," $\otimes$ denotes Kronecker multiplication, and U_v denotes the $(np, 1)$ vector gotten by stacking the columns of U (first column on top, second column next, and so on). The assumptions on U say that U has zero mean, that its rows (observations) are uncorrelated, and that all rows have the same variance matrix V up to a scalar multiple which may vary by treatment. This variance multiple for the i-th treatment is σ_i^2; the σ_i^2 allow a limited form of heteroskedasticity. As is apparent from (6), observations are assumed to be ordered by treatment from first treatment to last.

In the systematic component $X\beta$ of $Y = X\beta + U$, the matrix X is an (n, k) regressor matrix, where k is the number of regressors; and β is a (k, p) matrix of regression coefficients. The elements of X are functions of the design variables and are thus subject to experimental control. The n rows of X correspond to the n Monte Carlo observations. The value of a given row of X depends on which treatment the observation is taken at and what the regression functional form is. A row of X taken at treatment z_i will be given by some $(1, k)$ row vector of functions of z_i, call the function $f(z_i)$. Thus, the regressor matrix X takes the form:

$$X = \begin{bmatrix} f(z_1) \\ \vdots \\ f(z_1) \\ \hline \vdots \\ \hline f(z_m) \\ \vdots \\ f(z_m) \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_1 \text{ times} \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_m \text{ times} \end{matrix} \tag{7}$$

To illustrate the regression functional form f, the step form (4) above for the response function F can be represented by a dummy variable regression, with the i-th of $k = m$ dummy regressors equal to one if the observation is at the i-th treatment (and zero otherwise). The implied f-function is

$$f(z_i) = [i\text{-th } (1, m) \text{ unit vector}] \quad \text{(step form)}. \tag{8}$$

As another illustration, the quadratic form (5) above leads to the f-function

$$f(z_i) = [1, \alpha, N, \alpha^2, N^2, \alpha N] \quad \text{with } (\alpha, N) = z_i \text{ (quadratic form)}. \tag{9}$$

There are no necessary relations in general among the number of design variables D, the number of regressor variables k, the number of treatments m, and the number of observations n; although $D < k \leq m < n$ is common in practice.

The best linear unbiased estimate B of β and its variance matrix are given by

$$B = (X'Q^{-1}X)^{-1}X'Q^{-1}Y, \quad \text{var}(B_v) = V \otimes (X'Q^{-1}X)^{-1}. \tag{10}$$

Here B_v is the $(kp, 1)$ vector gotten by stacking the columns of B. Derivations can be found for example in Goldberger (1964, pp. 201–12), though they must be modified slightly to allow for heteroskedasticity. In view of (6) and (7), var (B_v) may be rewritten

$$\text{var}(B_v) = V \otimes \left[\sum_{i=1}^{m} n_i \sigma_i^{-2} f(z_i)' f(z_i)\right]^{-1}. \tag{11}$$

III. The Design Model

The Monte Carlo experimenter's goal may now be viewed as accurate estimation of the matrix β of Monte Carlo model parameters. He wishes to choose a design $n_1, \ldots, n_m$ which facilitates this estimation. The general experiment design literature suggests a choice procedure, called the *design model* here. The specific

formulation is that of Watts and Conlisk (1969), which builds on the work of many authors [see Fedorov (1972) and references there for a good entry to the literature]. There are two major components to the design model – the choice set and the choice criterion.

The choice set. It is assumed here that the scope of the Monte Carlo sampling experiment is limited by a budget constraint $\sum_i c_i n_i \leq C$, where c_i is the cost of one Monte Carlo observation at the i-th treatment and C is the available budget. The choice set is then the set of all non-negative m-tuples $(n_1, \ldots, n_m)$ which satisfy the budget constraint.

The choice criterion. The matrix $\operatorname{var}(B_v)$ is a matrix measure of the experimenter's error in estimating the Monte Carlo model coefficients β by B. It is natural to specify a scalar error measure defined on $\operatorname{var}(B_v)$ as the experimenter's objective function to minimize. Some conventional specifications are the determinant (generalized variance), maximum eigenvalue, and weighted trace of $\operatorname{var}(B_v)$ [see Fedorov (1972, 52–3) on these and other objective functions]. The discussion and illustrations here will use a weighted trace criterion, though one of the other standard criteria could easily be substituted.

The experimenter is interested in estimating β. More specifically, assume he is interested in estimating the height of each of the p response functions over a representative sprinkling of points in the design space; and assume he specifies the treatment points $z_1, \ldots, z_m$ as these representative points. The best linear unbiased estimate of these points is $X_0 B$, where X_0 is the (m, k) matrix of rows $f(z_1), \ldots, f(z_m)$. The (i, j)th element of $X_0 B$ is the estimated height of the j-th response function over the i-th treatment point. The experimenter is assumed to select as an overall measure of estimation error a weighted sum of variances of the elements of $X_0 B$—specifically, $\sum_i \sum_j w_i t_j \operatorname{var}[(X_0 B)_{ij}]$. Here $w_1, \ldots, w_m$ are weights associated with treatments and $t_1, \ldots, t_p$ are weights associated with response functions. The relative sizes of the w_i and t_j reflect the relative importances to the experimenter of the treatments and response functions. In view of (11), this estimation error measure may be manipulated as follows, where $W = \operatorname{diag}(w_1, \ldots, w_m)$, $T = \operatorname{diag}(t_1, \ldots, t_p)$, and a subscript v denotes the vector gotten by stacking the columns of the matrix on which the v appears.

$$
\begin{aligned}
\textstyle\sum_{i=1}^{m} \sum_{j=1}^{p} w_i t_j \operatorname{var}[(X_0 B)_{ij}] &= \operatorname{tr}\{(T \otimes W) \operatorname{var}[(X_0 B)_v]\} \\
&= \operatorname{tr}\{(T \otimes W)(I \otimes X_0) \operatorname{var}(B_v)(I \otimes X_0')\} \\
&= \operatorname{tr}[(I \otimes X_0')(T \otimes W)(I \otimes X_0)\{V \otimes \textstyle\sum_{i=1}^{m} n_i \sigma_i^{-2} f(z_i)' f(z_i)\}^{-1}\}] \\
&= \operatorname{tr}[(TV) \otimes \{(X_0' W X_0)[\textstyle\sum_{i=1}^{m} n_i \sigma_i^{-2} f(z_i)' f(z_i)]^{-1}\}] \\
&= \operatorname{tr}^k(TV) \operatorname{tr}^p \{(X_0' W X_0)[\textstyle\sum_{i=1}^{m} n_i \sigma_i^{-2} f(z_i)' f(z_i)]^{-1}\}.
\end{aligned}
\tag{12}
$$

This weighted trace function is assumed to be the experimenter's choice criterion. Using it, he can rank alternative designs in the choice set.

The design model. In words, the experimenter wishes to minimize the choice criterion (12) over the choice set. In symbols the design model is this:

$$\begin{aligned} &\underset{n_1,\ldots,n_m}{\text{minimize}}\ \operatorname{tr}\{X_0' W X_0 [\textstyle\sum_{i=1}^{m} n_i \sigma_i^{-2} f(z_i) f(z_i)']^{-1}\} \\ &\text{subject to } \textstyle\sum_{i=1}^{m} c_i n_i \le C, n_1 \ge 0, \ldots, n_m \ge 0. \end{aligned} \tag{13}$$

This is a well behaved mathematical programming problem: the objective function is convex and the constraints linear. Strictly speaking, (13) is an integer programming problem, since the n_i are treatment sample sizes. If the sample is large, however, little will be lost by the conventional practice of treating the n_i as continuous in solving (13) and then rounding off. The list of specifications required of the experimenter to apply the model are as follows:

(14)
The treatment points $z_1, \ldots, z_m$.
The corresponding variance weights $\sigma_1^2, \ldots, \sigma_m^2$.
The corresponding costs per observations $c_1, \ldots, c_m$ and the available budget C.
The corresponding treatment importance weights $w_1, \ldots, w_m$.
The regression functional form f (which together with the z_i determines X_0).

It may be noted that only the relative sizes of the σ_i^2 are important; a multiplicative change of scale would leave the solution to (13) unchanged. The same holds for the w_i; and it holds for the c_i plus C.

Computationally, solution algorithms for (13) are made easier by the fact that there is only one constraint in addition to the non-negativity constraints. Let $H(n_1, \ldots, n_m)$ denote the objective function in (13) and H_i its i-th partial. The following iterative solution algorithm is based on the idea of letting the relative sizes of the H_i/c_i determine how the n_i shift up and down from iteration to iteration:

$$n_i^{t+1} = (C/c_i)[H_i(n_1^t, \ldots, n_m^t)/c_i]^{\sigma} / \textstyle\sum_{j=1}^{m} [H_j(n_1^t, \ldots, n_m^t)/c_j]^{\sigma}.$$

Here t indexes the iteration, and σ is a positive convergence parameter to be set by the experimenter. My experience suggests setting $\sigma = 1$ at first. If the n_i do not bounce up and down from iteration to iteration, $\sigma = 1$ will do reasonably well. If the n_i do bounce up and down, reduce σ until the bouncing stops. This algorithm has been used to solve problems with $m = 224$ and $k = 50$ in less than half an hour on a CDC 3600. For initial n_i set at $n_i^0 = C/mc_i$, 50 iterations usually produced good convergence. The major computational effort of a given iteration is the summation and inversion of the cross products matrix in brackets in (13). Thus, one iteration is roughly equivalent to the work of running one regression with m observations and k regressors. In another context, Conlisk and Watts (1969) used a model identical to (13) except that there was more than one constraint in addition to the non-negativity constraints. This required a substantially more complicated solution procedure. A gradient projection algorithm programmed by Kreuser (1968) required about 45 minutes on a Burroughs 5500 to solve problems with $m = 54$, $k = 13$, and five constraints in addition to the non-negativity constraints.

An explicit solution formula to the design model is available when f takes the step function form (8); in this case, $X_0 = I_m$ and the matrix to be inverted in (13) is

diagonal. The solution formula is

$$c_i n_i / C = (w_i \sigma_i^2 c_i)^{1/2} / \sum_j (w_j \sigma_j^2 c_j)^{1/2} \qquad (i = 1, \ldots, m). \tag{15}$$

When the w_i, σ_i^2, and c_i vary over treatments, the optimal n_i computed from (15) will vary over treatments (except in the rare case when all the $w_i \sigma_i^2 / c_i$ happen to coincide). As intuition would suggest, a larger treatment importance weight w_i or variance weight σ_i^2 leads to a larger treatment size n_i; and a larger treatment cost c_i leads to a lower n_i.

Generalizing the criterion function. A convenient feature of the criterion function derivation (12) was that the magnitude $tr(TV)$ factored out and could thus be deleted from the design problem (13). Since T and V were the only components in (12) which reflected the existence of more than one response function, their deletion meant that the experimenter might as well have been working with only one response function to start with. The key assumptions in getting this simplification were—(i) that all response functions have the same functional form f; (ii) that the variance matrices $\sigma_1^2 V, \ldots, \sigma_m^2 V$ associated with different treatments were identical up to a multiplicative constant; (iii) that the experimenter was interested in estimating response surface heights over the same design space points for every response function; and (iv) that the weights in the criterion function $\sum_i \sum_j w_i t_j$ var$[(X_0 B)_{ij}]$ took the simple multiplicative form $w_i t_j$. More intuitively, these four assumptions say that the experimenter treats all response functions symmetrically; so there is no reason for the multiplicity of them to alter the ranking of designs.

The four assumptions are restrictive and may often be objectionable in practice. Fortunately, easy generalizations are available. For example, suppose asymptotic theory or other considerations lead the Monte Carlo experimenter to believe that different response functions, such as F and G in (2), have substantially different functional forms. He might then wish to specify a separate Monte Carlo model like (6) for each functional form, in which case he could define a separate sub-objective function like (12) for each functional form. Since each sub-objective would be a weighted sum of variances, it would be natural to define as grand objective function a weighted sum of the sub-objectives. Aside from the additional specifications and computations required, the generalization would be straightforward. Subject to the same qualification, generalizations of the other three assumptions listed would also be straightforward.

IV. Optimal Design Illustration

The Orcutt–Winokur context was described in the opening section; and treatments $z_1, \ldots, z_{20}$ were specified (Figure 1). To apply the design model, specifications are needed for the σ_i^2, the c_i, C, and the w_i. There is reason to specify unequal variance weights σ_i^2. The major dependent variables to a Monte Carlo experimenter will typically be estimators; and, in contexts where analytical results are available, estimators often have variances inversely proportional (roughly) to the target model sample size N. For this reason, the σ_i^2 were set equal to the inverses of the corresponding N-values in the illustration. For the budget constraint magnitudes c_i and C, it was noted that a given Monte Carlo observation involves

generation and processing of a target model sample of size N. With this in mind, the cost c_i for one Monte Carlo observation at the i-th treatment was made proportional to the corresponding N-value. The total budget C was set so that an equal treatment size design with all $n_i = 1.000$ would just exhaust the budget. The design model has the property that the optimal ratios among the n_i are invariant to the size of C: so C is not an important specification for the following discussion. The treatment importance weights w_i should reflect the experimenters' interests. The weights specified are presented on the first row of Table 1: they assume greater interest in positive α-values (positive serial correlation in the target model) than negative, and equal interest in alternate N-values.

TABLE 1

OPTIMAL n_i FOR ORCUTT-WINOKUR ILLUSTRATION

		α-Value				
	N	−1.0	−0.5	0	0.5	1.0
Specification of w_i	All N	0.5	0.5	0.7	1.0	0.7
Optimal Design for Step f	10	3,460	3,460	4,094	4,893	4,094
	30	1,153	1,153	1,365	1,631	1,365
	50	692	692	819	979	819
	70	494	494	585	699	585
Optimal Design for Quadratic f	10	7,185	0	8,524	0	8,824
	30	1,402	0	2,036	0	1,718
	50	674	0	1,555	0	864
	70	1,066	0	1,138	0	1,300

It only remains to specify the regression functional form f. Optimal designs are presented on Table 1 for two f-specifications, the step form (8) and the quadratic form (9). The optimal step function design departs greatly from the equal treatment size design common in Monte Carlo studies; the reasons are apparent from the step function solution formula (15). The optimal quadratic function design departs even further from the equal treatment size design: the reason is continuity. Since the quadratic function is continuous, inferences about response function heights over all treatments can be made even if there are no observations at some treatments: so the model is free to put all stress on cheap or geometrically well placed treatments. The classic example of this phenomenon would occur if the response functional form was simple linear and the c_i were equal: then the optimal design would allocate all observations to the corner treatments, for much the same reason that a table on a shaky floor will be stablest if the legs are at the corners.

If Orcutt and Winokur felt confident that the functional flexibility of a quadratic was adequate to their context (and if they accepted the other illustrative specifications), the quadratic design would be appropriate. If they thought a quadratic was too restrictive, but still wished to exploit prior beliefs about continuity, they might compute a design for a more flexible, but still continuous, f. If they felt uncomfortable with the restrictiveness of any continuity assumption, the

step function design would be appropriate. After the fact, it is possible to see from Orcutt and Winokur's results that a quadratic f would have been reasonable for the response functions they were estimating.[2] Of course, Orcutt and Winokur could not know this prior to their study.

The question of how to specify f is typically a knotty one [see Conlisk (1973)]. However, the important point here is that no assumption about f is likely to lead to an equal treatment size design. Since the criterion function of the design model is a variance magnitude, the relative efficiency of two designs in the usual variance sense can be measured by the ratio of the two criterion function values. When the step f is specified, the efficiency of the equal treatment size design relative to the optimum design is only 0.59, indicating that the equal treatment size design loses 41 percent on the dollar of computer time. When the quadratic form for f is specified, this efficiency is 0.58. Similar low efficiencies for the equal treatment size design hold under other specifications of f.

V. Conclusion

Though Monte Carlo sampling experimenters seldom use optimal design analysis, the application is straightforward once the goals of a Monte Carlo experiment are formulated in terms of estimating response functions (whose dependent variables are usually the first two moments of target model estimators and test statistics). More specifically, if the response functions are treated as regression equations, regression design theory leads to optimal designs determined by the well behaved programming problem (13). In specifying the inputs (14) to the programming problem, the experimenter tailors the optimal design to his Monte Carlo context.

A distinctive feature of Monte Carlo (and other computer simulation) contexts is the presence of multiple response functions [see Naylor *et al.* (1967), section 7.4]. Orcutt and Winokur, for example, table results for 16 functionally independent mean and mean square error response functions. As discussed in section III, the multiple response case collapses to the single response case under certain simplifying assumptions. Since the simplifying assumptions may often be objectionable, the generalization of the objective function discussed in section III may be needed.

One often hears the comment that Monte Carlo results are no substitute for general analytical results, because Monte Carlo results apply only to the limited target model specifications chosen by the experimenter. The inappropriateness of this criticism is apparent once a Monte Carlo experimenter's goal is seen as estimation of response functions like the mean function $F(\alpha, N)$ and mean square error function $G(\alpha, N)$ from (2). If analytical results were available, all they could give us would be functions like $F(\alpha, N)$ and $G(\alpha, N)$. But this is exactly what a Monte Carlo study can give us (subject to controllable approximation error). The fact that Monte Carlo observations on a function like $F(\alpha, N)$ can be generated for only a

[2] For example, their Table V presents the (approximate) contours of six response functions—mean and mean square error functions [like $F(\alpha, N)$ and $G(\alpha, N)$ in (2) above] for three estimators of α. The table presents the heights of each response function over a 15 point grid of (α, N) combinations. Least squares fits of the quadratic f for each of the six functions yield six R^2's ranging from 0.97 to 0.99.

limited number of (α, N) combinations does not prevent us from exploiting continuity to estimate the whole function $F(\alpha, N)$.

The Orcutt and Winokur illustration (and others not reported) suggest that potential efficiency gains from optimal designing are sizable relative to the equal treatment design typically used in Monte Carlo studies. Four reasons for the inefficiency of the equal treatment size design can be listed. *First*, the costs per treatment c_i will vary among treatments with varying target model sample sizes (N-values) and may vary for other reasons (as when a design variable being zero simplifies computations). Unequal c_i promote unequal n_i. *Second*, the variance weights σ_i^2 are likely to vary by treatment, since the response function dependent variables are likely to have variances which decline with target model sample size. Unequal σ_i^2 also promote unequal n_i. *Third*, the treatment importance weights w_i are likely to vary by treatment, since a Monte Carlo experimenter will typically have greater interest in some ranges of target model parameters than others: this also promotes unequal n_i. *Fourth*, continuity of the regression functional form f promotes unequal n_i. Even if the c_i, σ_i^2, and w_i were equal across treatments, a continuous f would promote unequal n_i, since the model would seek outlying and other geometrically well placed treatments, often to the complete exclusion of some treatments ($n_i = 0$ for some i).

All of this should be qualified by consideration of the cost of implementing the design model. Solution to the programming problem (13) can be costly. If this solution cost is important, the experimenter might wish to use the optimal design for the step form of f, since the explicit solution formula (15) makes computational cost trivial in this case. Though the step f-form does not exploit continuity, it is a definite improvement over the equal treatment size design.

University of California at San Diego

Submitted July 1972
Revised August 1973

REFERENCES

Conlisk, J. (1973), "Choice of Response Functional Form in Designing Subside Experiments." *Econometrica*. Vol. XLI. No. 4. pp. 643–656.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Academic Press. New York.

Goldberger, A. (1964), *Econometric Theory*, Wiley, New York.

Haitovsky, Y. and S. Jacobs (1972). "REGEN—Computer Program to Generate Multivariate Observations for Linear Regression Equations," *Annals of Economic and Social Measurement*. Vol. I. No. 1, pp. 43–57.

Naylor, T. H., D. S. Burdick, and W. E. Sasser (1967). "Computer Simulation Experiments with Economic Systems: The Problem of Experimental Design." *Journal of the American Statistical Association*. Vol. LXII, No. 320. pp. 1315–1337.

Kreuser, J. (1968). *Users and Programmers Manuals for GPM*, mimeo writeup for Social Systems Research Institute, University of Wisconsin.

Orcutt, G. and H. Winokur (1969), "First Order Autoregression: Inference, Estimation, and Prediction." *Econometrica*, Vol. XXXVII, No. 1. pp. 1–14.

Summers, R. (1965). "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," *Econometrica*, Vol. XXXIII, No. 1, pp. 1–41.

Watts, H. and J. Conlisk (1969), "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," *American Statistical Association Proceedings. Social Statistics Section*, pp. 150–156.