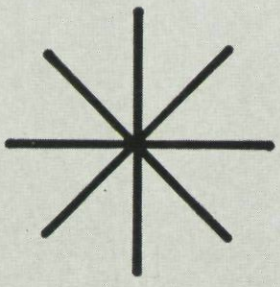
....COMPUTER APPLICATIONS



EDITOR: *I. J. Abrams*

SUBEDITORS: *Ronald E. Frank*
Paul E. Green
Charles Kadushin
David Montgomery
Kenneth M. Warwick

THEODORE A. VAN WORMER and DOYLE L. WEISS*

Empirical validation and testing of simulation models have been hampered by the problems associated with estimating parameters for their nonlinear structures. This article discusses direct search techniques as a methodology for fitting parameters of models too complex for linear regression analysis.

Fitting Parameters to Complex Models by Direct Search

The modern computer has had at least two fundamental effects on researchers and policy makers in marketing. First, its power as a symbol manipulation device provides researchers with a new flexibility for model construction, and second, computer powered information systems provide hope that when complex models and theories are developed they may also be empirically tested and validated.

Attempts to employ rigorous methods for testing complex models (such as statistically estimating their parameters) present many difficulties. Even in small models, there are many parameters and the structure is usually complex, extremely nonlinear, and highly interdependent. For some models, however, subparts of the structure can be isolated and examined independently. Even in these cases, however, estimating parameters is often difficult because of the substructure's nonlinear form. Often the form is so complex that no known transformation can make it linear and thus suitable for regression analysis. This article reports on the application of a procedure for estimating parameters of nonlinear structures when linear regression analysis is not appropriate.

* Theodore A. Van Wormer is Vice President with the CiMS Group, Inc., Pittsburgh, Pennsylvania. Doyle L. Weiss is Associate Professor of Industrial Management, Krannert Graduate School, Purdue University.

Journal of Marketing Research,
Vol. VII (November 1970), 503–12

DIRECT SEARCH

The usual method of estimating a model's parameters from a set of data involves minimization of a sum of squares, a "least squares" problem. The squares to be summed and minimized are formed by squaring differences between the observed value of the dependent variable and its estimated value resulting from the model.

The classical method of solving least squares problems which result in nonlinear normal equations involves the use of Taylor approximations about a trial solution for the parameters to reduce residuals to linear form [9]. Levenberg [5] notes that this method fails if the subsequent application of least squares to the linear approximations produces parameter estimates considerably distant from trial solutions. Failure occurs because the method neglects higher order terms, which become larger and more important as distance increases. Levenberg outlines procedures for mitigating this difficulty for certain classes of least squares problems. Such work stems directly from Newton's method for finding roots for sets of nonlinear simultaneous equations and is typical of the classical solutions to such problems.

In contrast to these classical techniques, the increasing availability and power of modern computers have caused some researchers to devise methods for searching directly for solutions to complex problems in least squares. The difference between such direct search methods and classical mathematical methods of optimization is largely in the former's use of a simple searching strategy instead of sophisticated theoretical development. Operationally, direct search techniques examine trial solutions of the function being optimized and compare these to earlier ones to determine what, if anything, should be done next to improve the critical value. The power of direct search rests (1) in its built-in strategy for deciding where on the response surface to search next for improvements and (2) the speed of the data processing equipment used. Clearly, one can trade sophistication in searching procedures against speed in evaluating trial solutions.

There are other areas where such problems exist. In addition to the problem of obtaining estimates of parameters for complex simulation models, many scientific models are nonlinear, and some cannot be made linear. For example, if a researcher collects 25 pairs of observations (x, y) on a process, its model is:

$$\hat{y} = A \sin (BX + C) + D \operatorname{sech} (EX + C).$$

One can easily form the sum of the squared deviations from predicted and observed values of y as:

$$\sum_i (y_i - \hat{y}_i)^2 = f(A, B, C, D, E),$$

which one may regard as a function of the model's parameters. The classical approach to minimizing the above sum is to take first derivatives and set them equal to zero. However, this creates a more difficult problem than previously existed.

Models in the biological sciences often attempt to predict growth of various organisms or populations in certain environments. These models' rules usually state that the rate of growth depends partly upon the current population size. Although these rules can sometimes be linearized, this often results in transforming other linear rules into nonlinear form, which increases the problem's complexity.

In general, direct search techniques view the function to be minimized (or maximized) as a black box. The box allows certain inputs to be set (trial values of the function's arguments) and responds with a single output value (the value of the function at that point).

In this article, we shall discuss a few of the issues and approaches involved in using direct search as a technique to fit parameters to complex models [7, 8]. First, the function to be minimized must be defined. Although function definition must be independent of the parameter estimation technique used, with a linear system using regression analysis a direct statement of the function to be minimized is not always made.

The problem of selecting the function to minimize is not always trivial. For example, with data on three variables (X, Y, Z), one would predict X as a function of Y and Z:

$$\hat{X}_i = f(Y_i, Z_i).$$

If one believes that the function has parameters (A, B, and C) the prediction equation is, more explicitly:

$$\hat{X}_i = f(Y_i, Z_i, A, B, C).$$

In calculating $\hat{X}_i$ for any values of A, B, and C, what function should be minimized to estimate A, B, and C? First, perhaps one should minimize:

$$g(X_i, X_i) = \sum_i (X_i - X_i)^2.$$

This is certainly one possibility. However, if the observed X_i vary widely, say from .001 to 1000, one might minimize h, rather than g:

$$h(X_i, \hat{X}_i) = \sum_i (1.0 - \hat{X}_i/X_i)^2; X_i > 0.$$

A full discussion of the implications of the alternative minimization criterion indicated above is beyond the scope of this article. We merely wish to indicate that the choices necessary in any estimation problem are often a function of the situation and are almost never trivial.

Consider another problem in function definition. A chemical engineer has data on X, Y, and Z, all greater than zero. He believes that they are related as follows:

$$Ax_i + By_i = \exp(-C/z_i).$$

He wishes to find A, B, and C so that his model best fits the observed data. He may form the usual least

squares error function:

$$f(A, B, C) = \sum_i [Ax_i + By_i - \exp(-C/z_i)]^2$$

and attempt to minimize f, which has no minimum. Consider its components as follows: if A and B are both zero, then their contribution to the sum will be zero. But no matter how large C is, the exponential term may be made smaller by making C even larger. The limiting values for A, B, and C then become:

$$A = 0, \qquad B = 0, \qquad C = \infty.$$

Such an error definition is not very useful to the engineer, who must rethink his definition of fitting the data.

As a methodology, direct search covers a vast array of techniques. Of necessity, we can touch on only a few of these, and as a result we shall state some limitations on rationales employed and search environments selected.

First, we shall examine techniques which make inferences from the behavior of the response surface in the local vicinity of the point to be examined. This immediately rules out such techniques as random selection of argument values or imposing a grid on the space involved and checking all grid points.

Second, we shall limit ourselves to the cases where there is no advantage in making more than one function evaluation concurrently (running a number of experiments simultaneously with different values for the exogenous variable). Thus, we restrict our interest to sequential search.

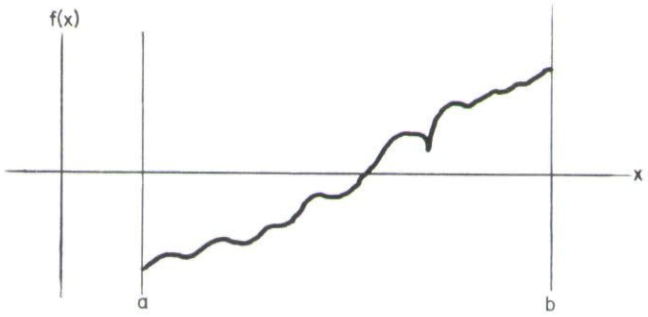
Third, our interest will be only in deterministic, rather than stochastic, search problems. There may be some random error in the observed data, but in our problems evaluating the function which defines the response surface more than once with the same inputs always gives the same output. This rules out search procedures where random (or unforeseen) elements enter into the function evaluation procedure.

Fourth, we shall be concerned with functions defined for all values of the exogenous variable. If a function is undefined outside of some region, we shall arbitrarily define it to yield some very large positive number (if we are minimizing). This rules out, among other things, functions defined only at discrete points.

Finally, we shall be concerned with unimodal response surfaces. Thus in searching for and finding a minimum, it is arbitrarily defined as the minimum minimorum. While some response surfaces are not unimodal, many are. In fact, the sum of squares of predicted values of the dependent variable minus observed values is unimodal if our prediction function is well behaved.

These restrictions rule out a large number of problems and techniques of theoretical and practical interest. Fortunately, they still leave a substantial number of problems of concern to researchers working with marketing data.

Figure 1
RESPONSE SURFACE: EXAMPLE I



The following are rather elementary examples. Recent users of this methodology may not be familiar with the more elementary methods, however, these examples may assist some readers in familiarizing themselves with the conceptual issues involved in more complex techniques.

EXAMPLE I

We wish to find the value of the independent variable where a function of this variable equals zero. The value of the variable is known to be within the interval (a, b). To the left of the zero the function is known to be negative and to the right, it is positive. (See Figure 1.)

It is not important that this function is not necessarily differentiable or even continuous, but only that it crosses the y-axis at only one point, to the left of which it is negative and to the right, positive.

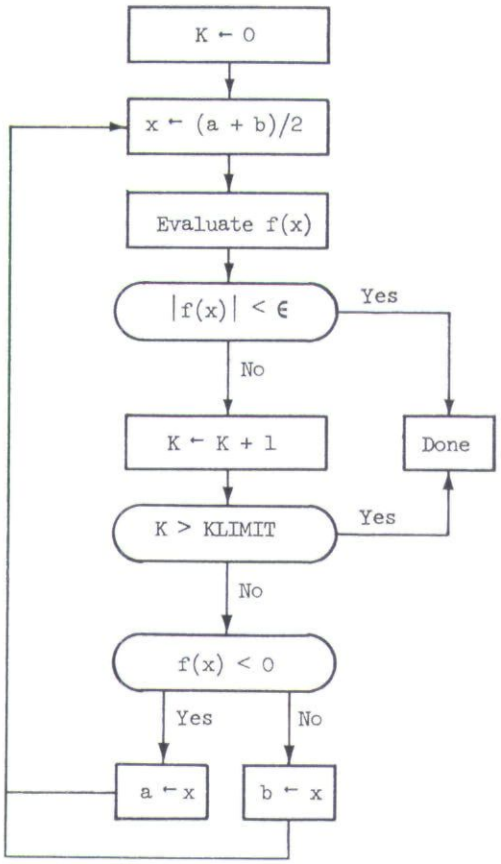
This is a case of "derivative information," where we may employ Bolzano's method, as follows. First, we cut the interval of uncertainty (a, b) in half and evaluate the function at the point $x = (a + b)/2$. For a positive result, we replace b with x; for a negative result, we replace a with x. We then iterate, recompute x, and evaluate the function at the new value, using the updated value of a or b. A likely termination criterion would be that the function's value be sufficiently close to zero or that enough evaluations have been made. The flow chart for this procedure is presented in Figure 2. In the case illustrated by the flow chart, search is terminated after n evaluations. The interval of uncertainty will have been reduced by a factor of 2^n, from $b - a$ to $(b - a)/2^n$.

EXAMPLE II

Consider a function of one variable known to be strongly unimodal and concave downward. To discover the variable's value at which the function takes on its minimum, consider the data X_i, Y_i, Z_i for $i = 1, \cdots, n$ and the following model:

$$Z_i = \alpha X_i + Y_i/\alpha$$

Figure 2
FLOW CHART: BOLZANO'S METHOD



We must find α such that:

$$f(\alpha) = \sum_i [Z_i - (\alpha X_i + Y_i/\alpha)]^2$$

is minimal. If $f(\alpha)$ takes on its minimum value between $a < \alpha < b$, we evaluate the function $f(\alpha)$ at $\alpha = a$ and $\alpha = b$, obtaining $f(a)$ and $f(b)$. Next, searching factors $g_2 = (\sqrt{5} - 1)/2$ and $g_1 = 1 - g_2$ are computed. In rough terms, $g_2 = .618$ and $g_1 = .382$. These factors have the property that $g_1/g_2 = g_2/1$, and they are the divisors of a line which results in the "golden section." We now begin *golden section search.*

We evaluate function f at $f(a + g_1(b - a))$ and $f(a + g_2(b - a))$. Abbreviating these two new evaluation points as $f(p_1)$ and $f(p_2)$ where $p_1 = a + g_1(b - a)$ and $p_2 = b - g_1(b - a)$, we have four values of function $f(a), f(p_1), f(p_2), f(b)$.

Since our function is strongly unimodal, one of these four values is less than or equal to all of the others. If the minimal value is at one of the interior points, p_1 or p_2, we retain this point and its two neighbors. For example, if $f(p_1)$ is minimal, the desired minimum lies between a and p_2, so we can ignore the interval between p_2 and b. If the minimum value is at an end point (a or b), then we save its two nearest neighbors (p_1 and p_2). In either event, we have reduced the interval of uncertainty by a factor of g_1. We next compute the value of the function at a new point.

To select the new point, we first replace the old end points (a and b) with the new end points (for example a and p_2). Next we evaluate f at $f(a + g_1(p_2 - a))$ and $f(p_2 - g_1(p_2 - a))$; only one new computation is necessary because we already have the second value. We reduce our interval of uncertainty by .382 each iteration. The flow chart for *golden section search* is in Figure 3.

EXAMPLE III

Here is a case where there is more than one parameter to estimate, again with data X_i, Y_i, and Z_i and the model:

$$\hat{Z}_i = \alpha X_i + \beta X_i Y_i .$$

The sum to be minimized is:

$$f(\alpha, \beta) = \sum_i [Z_i - \hat{Z}_i]^2.$$

Again, we assume that f takes on its minimum value in the region ($a_0 \leq \alpha \leq a_1$, $b_0 \leq \beta \leq b_1$). We must now select values for α and β to evaluate f, finding the best value for one parameter at a time. After the best value is found for one parameter, the value of f might be improved by searching for better values of the other parameter. This procedure is called *univariate search* since it involves manipulating only one variable at a time. Selecting a set of arbitrary starting values α_0 and β_0, we perform golden section search on α, holding β constant at β_0. After obtaining the best value of α (after some number of function evaluations) at α_1, we now look for the best value of β, holding α constant at α_1. After finding β_1 such that $f(\alpha_1, \beta_1)$ is minimal for different values of β, we now attempt again to improve the value of α holding β constant at β_1. This procedure continues until some arbitrary termination criterion is met. Univariate search works well when there is little interaction between the parameters. However, if there is strong interaction between α and β, it is rather ineffective. Figures 4 and 5 illustrate contour maps of these two extreme cases.

Notice that in Figure 5 the procedure reaches the point (α_1, β_0) and no apparent improvement is possible, although we are clearly not at the minimum of the function. The difficulty lies in the necessity to search in a direction which represents some combination of changing α and β, as considered next.

EXAMPLE IV

Again assuming a function of two parameters $f(\alpha, \beta)$ to be minimized (in some region), we start at an arbitrary (α_0, β_0). We now consider two new points $(\alpha_0 + \Delta, \beta_0)$ and $(\alpha_0, \beta_0 + \Delta)$ where Δ is an arbitrary number.

We now have three values of our function: $f(\alpha_0, \beta_0)$, $f(\alpha_0 + \Delta, \beta_0)$, and $f(\alpha_0, \beta_0 + \Delta)$. These three points determine a plane, and by suitable mathematics we can compute the combination of concurrent changes needed in α and β to gain the maximum reduction in f per unit change in α and β. This technique is known as *steepest descent* (or steepest ascent, if we are maximizing).

In the example, the direction of steepest descent will be found in the direction:

$$K_1\alpha + K_2\beta,$$

where K_1 and K_2 can be found by solving the equations:

$$\frac{K_1}{K_2} = \frac{f(\alpha_0 + \Delta, \beta_0) - f(\alpha_0, \beta_0)}{f(\alpha_0, \beta_0 + \Delta) - f(\alpha_0, \beta_0)}, \quad \text{and}$$

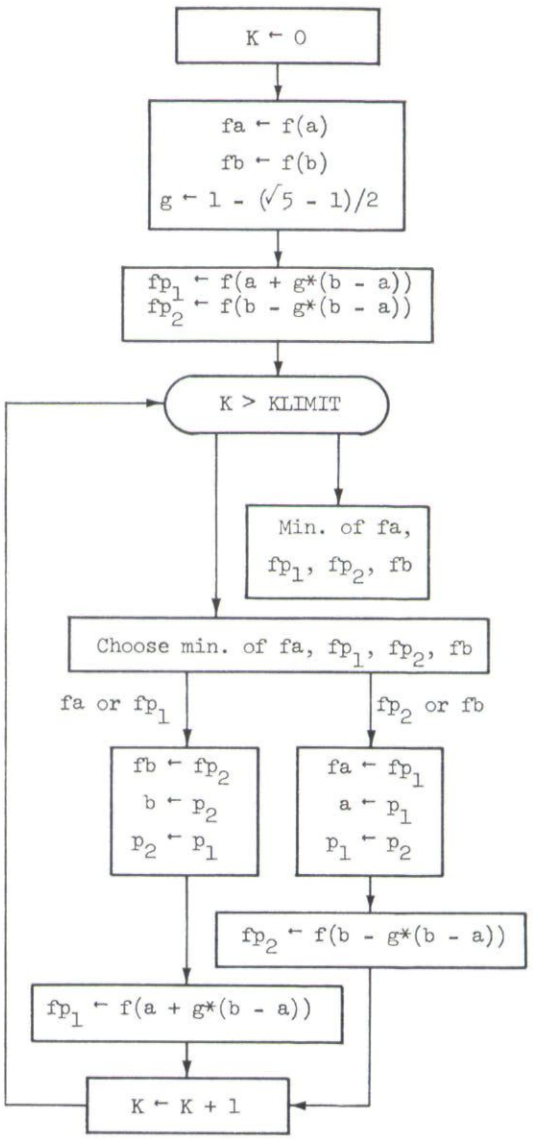
$$K_1 + K_2 = 1.$$

Using the direction $K_1\alpha + K_2\beta$, we can search along the line which descends at the steepest angle from (α_0, β_0), by golden section search. This strategy is somewhat better than univariate search. In some cases, it is spectacularly successful. Although it is not always recognized as such, linear programming is a special case of steepest ascent search. However, steepest descent search has inadequacies also. In Figure 5, if we were at the point (α_1, β_0), steepest descent search would not be able to find any better values for α and β. What is needed is an approach which can find and follow valleys or ridges. The next example illustrates such a strategy.

EXAMPLE V

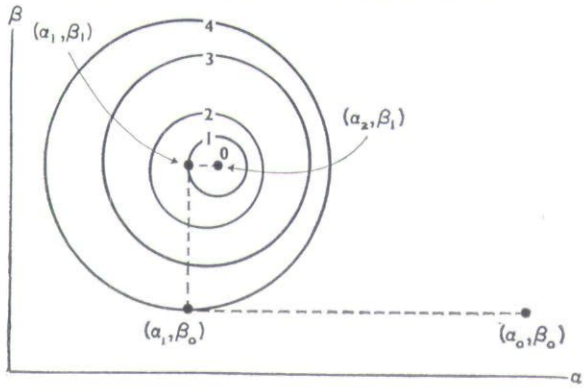
Looking for the minimum of $f(\alpha, \beta)$ in this case, we employ *pattern search* [1]. This method employs two types of trial evaluations of the function. The method makes a series of exploratory moves during which it increments the arguments of the function in a serial manner, noting the effect of this action on the value being minimized. The exploratory moves produce what is essentially an estimate of the slope of the response surface in the neighborhood of the point. After this limited preliminary exploration, the program makes a pattern move, whose direction depends completely on discoveries about the response surface during the exploratory moves. The distinctive feature of the pattern move is that it involves all of the arguments and is repeated and amplified, if successful, using exploratory moves to adjust its direction. If failure occurs, the method establishes a new trial solution and begins a new series of exploratory moves to establish a new pattern. The working hypothesis of the technique is that any set of adjustments to the independent variables which produces an improvement in the critical value can be exploited to produce greater improvement on subsequent trials. In effect, the method is conditioned by past successes and/or failures. Operationally, this conditioning results in larger and larger moves as a successful pattern is developed. When the pattern fails, shorter step sizes (or increments to the arguments) are implemented as the method attempts to establish a new pattern, or direction.

Figure 3
FLOW CHART: GOLDEN SECTION SEARCH



Wilde's [7] experience with this method shows it to be particularly well suited for nonlinear curve-fitting problems. In addition, Hooke and Jeeves [1] report that the search time necessary for a solution increases only as the first power of the number of parameters being estimated. In general, computational time grows as the cube of the number of parameters. Wilde suggests that this efficiency results from the method's excellent ability to track the value of the function along steep valleys (or ridges) in the response surface.

Figure 4
WEAK INTERACTION BETWEEN VARIABLES



PATTERN SEARCH

Local Exploration

The first procedure in pattern search is local exploration. Given an arbitrary point (α, β) and the associated value of the function, we evaluate nearby points as follows: if the computed value of $f(\alpha + \Delta\alpha, \beta)$ is smaller than the given value of f, we record it. If not, and if the computed value of $f(\alpha - \Delta\alpha, \beta)$ is smaller than the given value of f, we record it. If not, we record the best associated given value of f. We have now recorded either $\alpha + \Delta\alpha$, $\alpha - \Delta\alpha$, or α and the value of the function. We call this point α_1 and the best function value f_1. Proceeding from the recorded (or remembered) point, we make the same computations with β. If the computed value of the function $f(\alpha_1, \beta + \Delta\beta)$ is smaller than f_1, we record it. If not, and if the computed value of $f(\alpha_1, \beta - \Delta\beta)$ is smaller than f_1, we record it. If not, we record β. We have now recorded either $\beta + \Delta\beta$, $\beta - \Delta\beta$, or β, calling the point β_1. We also have the value of the function at $f(\alpha_1, \beta_1)$, unless $\alpha_1 = \alpha$ and $\beta_1 = \beta$, where we have the original given value of f. In this discussion, we explored around a point by adding or subtracting some fraction (Δ) of the point's value from the point. This fraction Δ need not be the same for both directions—we could have used one Δ for exploring in the α direction and different Δ to explore in the β direction. For convenience, we assume Δ to be equal for both directions.

Notice also, that if the original value of α or β is zero, we have not found a new point to evaluate. For this reason, some investigators favor exploratory moves of $\alpha + \Delta$ and $\alpha - \Delta$ where Δ is the amount of distance to move, rather than a fraction of the α or β value. The superiority of one approach over the other is highly problem-dependent.

Pattern Moves

The other major procedure used by pattern search is the pattern move. Starting with an arbitrary point (α_0, β_0), we perform local exploration, yielding (α_1, β_1). Assuming that the new point is not identical with the original point, we take a step. Since (α_1, β_1) was better than (α_0, β_0) we compute (α_2, β_2) as $\alpha_2 = \alpha_1 + 2(\alpha_1 - \alpha_0)$; $\beta_2 = \beta_1 + 2(\beta_1 - \beta_0)$. The choice of using the constant 2 is arbitrary and results in a doubling of step size. Any value greater than unity will work, and some investigators prefer using a step size value of 1.5.

We now have a new point (α_2, β_2) which may result in an improved value for our function. However, we do not immediately compute $f(\alpha_2, \beta_2)$, because this would simply be proceeding in a straight line, without noticing any improvements obtained by modifying our course. Instead, we perform local exploration around (α_2, β_2), using $f(\alpha_1, \beta_1)$ as the value of the function to compare for improvement. This results in (α_3, β_3) if an improvement is found. If none is found, we can then evaluate $f(\alpha_2, \beta_2)$ to see if this direction should be maintained. If, however, $f(\alpha_3, \beta_3)$ is better than $f(\alpha_1, \beta_1)$, we now compute (α_4, β_4) as follows:

$$\alpha_4 = \alpha_3 + 2(\alpha_3 - \alpha_1); \beta_4 = \beta_3 + 2(\beta_3 - \beta_1).$$

In general:

$$\alpha_i = \alpha_{i-1} + 2(\alpha_{i-1} - \alpha_{i-3})$$
$$\beta_i = \beta_{i-1} + 2(\beta_{i-1} - \beta_{i-3})$$

for even $i > 2$.

The coordinates (α_i, β_i) for odd i are found by local exploration. The exception to this rule is our first step, where $\alpha_2 = 3\alpha_1 - 2\alpha_0$ and $\beta_2 = 3\beta_1 - 2\beta_0$. At some point, our step will result in a point where local exploration around point (α_i, β_i) does not yield an improvement and point (α_i, β_i) itself yields no improvement. In this case, we take point $i - 1$ (our best yet) as our arbitrary starting point and conduct local exploration again. If it yields an improvement, we begin another set of pattern moves.

Eventually, local exploration will yield no improvement, and we must reduce Δ by an arbitrary amount (many investigators favor dividing Δ by 2). Then we begin again attempting local exploration about our best point to date. If successful, we begin the pattern move process again. If not, we reduce Δ again. Eventually Δ falls below some arbitrary limit, the criterion of termination. Figures 6 and 7 are flowcharts for pattern and exploratory moves.

AN APPLICATION OF PATTERN SEARCH

The application discussed here resulted from an interest in estimating parameters for a complex model of demand behavior for a frequently purchased, low-cost, branded consumer product. In particular, the investigation was concerned with determinants of competitive market share. The nonlinear structure whose parameters were estimated was of interest because it stems directly from recent empirical findings on the brand-shifting characteristics of individual consumers [2]. Various forms of the nonlinear demand model were fitted and compared with several linear

forms of the relationship commonly found in econometric research: for a complete description of all the models analyzed and of the data[1] used, see [6].

The most interesting form of the nonlinear relationships investigated was:

$$S_{B,t} = rS_{B,t-1} + (1 - r) \cdot \left[b \frac{Q_B P_{B,t}{}^{\epsilon_P}}{\sum_B Q_B P_B{}^{\epsilon_P}} + (1 - b) \frac{Q_B P_{B,t}{}^{\epsilon_P} A_{B,t}{}^{\epsilon_A}}{\sum_B Q_B P_{B,t}{}^{\epsilon_P} A_{B,t}{}^{\epsilon_A}} \right] + \gamma_{B,t}$$

where:

$S_{B,t}$ is share of market of Brand B in period t.

r is a parameter which reflects consumers' habitual purchasing behavior. It may be interpreted as the probability of routinely purchasing a previously purchased brand if there are no sharp increases in its retail price and availability. (This parameter is assumed constant over time and brands, although such an assumption is not logically necessary.)

b is the probability that a consumer is unaffected by advertising or, alternatively, the expected portion of consumers who are unaffected by advertising.

$P_{B,t}$ is the price of Brand B in period t.

ϵ_P is the parameter reflecting the sensitivity of market shares to relative prices of brands and is closely related to economists' concept of price elasticity of demand: see [3].

Q_B is a set of dummy parameters to represent two interacting variables for which actual data were not available at the time of this analysis:

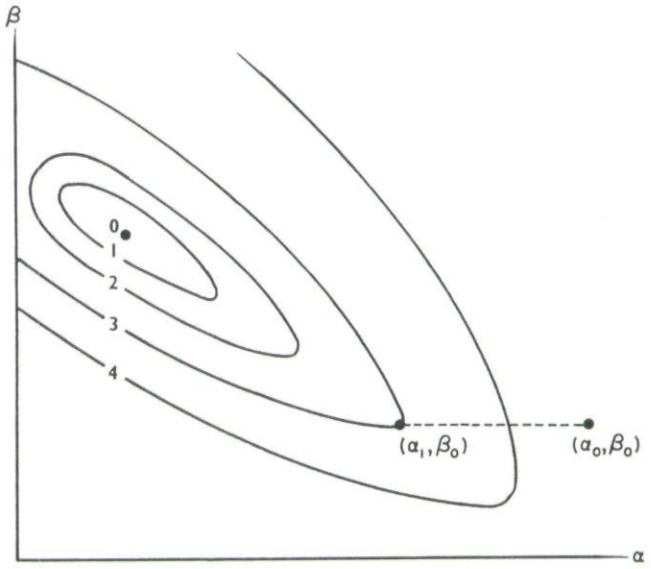
1. *product quality:* a measure of perceived desirability of a brand, reflecting intrinsic differences as determined in blind product tests and psychic differences stemming from positioning of the brand due to packaging and advertising. This term should be multidimensional to reflect segmentation in the market (see [4]) but is treated in the current research as a scalar in a homogeneous market.
2. *effective distribution:* retail availability, including quantity and quality of shelf space and special displays. This measure should also be treated as multidimensional for some product classes to reflect differences in availability by type of retail outlet; this is not necessary, however, for the particular product being analyzed.

$A_{B,t}$ is advertising expenditures for Brand B in period t.

ϵ_A is a measure of the sensitivity of market shares to the relative levels of advertising expenditures of the competing brands. Its counterpart with respect to price, ϵ_P, is defined above.

$\gamma_{B,t}$ is the stochastic term.

Figure 5
STRONG INTERACTION BETWEEN VARIABLES



Two assumptions were made about the model before it was fitted to the data: (1) the parameter describing the habitual repurchase behavior of consumers was assumed to be zero, and (2) the effects of lags on the effectiveness of advertising expenditures were ignored. Rationale for these assumptions is presented in detail in [6].

The sum of squared residuals minimized was

$$\sum_{t=1}^{24} \sum_{B=1}^{3} (S_{B,t} - S_{B,t})^2.$$

RESULTS

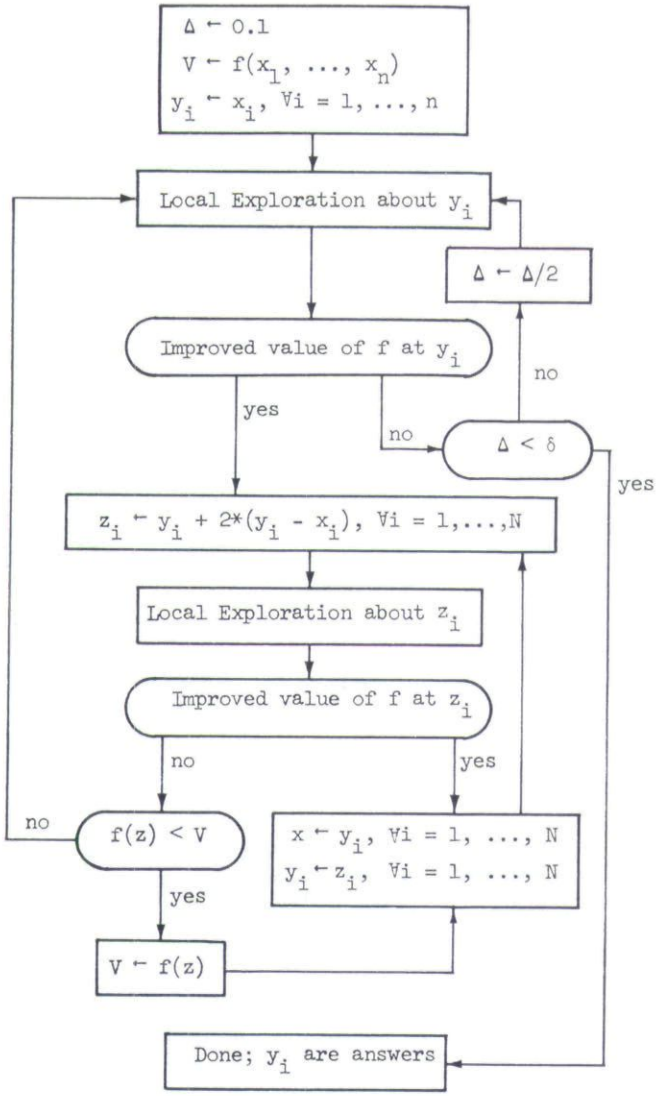
The search procedure produced the following estimates of the nonlinear model's parameters: $\epsilon_P = -3.19$, $\epsilon_A = 3.93$, $b_P = .957$, $Q_1 = 3.66$, $Q_2 = .821$, and $Q_3 = .724$. The implied R^2 was .937 and the termination tolerance[2] was .005.

There are both theoretical and practical limitations associated with all direct search procedures. The most critical of these is that a solution is not guaranteed in a finite time and, any solutions discovered by the procedure are not guaranteed to be the minimum minimorum (or maximum maximorum). The nature of the search procedure is such that it may arrive and stop at a local minimum (or maximum) point.

[1] In brief, the data were 48 bimonthly observations of market share, price, and advertising expenditures for three national brands marketed in the Chicago metropolitan market. The data represent the purchasing behavior of a sample of 899 families collected by the *Chicago Tribune's* Family Survey Bureau.

[2] The termination tolerance defines the conditions under which pattern search will stop attempting to improve its solution. Pattern search will stop work when it believes it has values for the parameters that are within $\Delta \cdot a_i$ (for all i) of the true optimum parameters (a_i is the ith parameter of the functional and Δ is the error tolerance established by the user).

Figure 6
FLOW CHART: PATTERN MOVE



Where: N – number of parameters
δ – limiting "tolerance"
$x_1, \ldots, x_n$ – starting parameter values

Some indication of the nature of a solution discovered by the program can be obtained by starting the procedure at different points and observing whether or not it returns to the original solution point. This technique was applied to test the above solution. For this test, the program was started at values of: $\epsilon_p = 2.80$, $\epsilon_A = 3.00$, $b_p = .49$, $Q_1 = .42$, $Q_2 = .03$, and $Q_3 = .03$. It rediscovered the original optimum parameters in approximately 15 minutes.[3] Another test was conducted in which the values of the optimum parameters were interchanged and altered slightly: $\epsilon_p = 3.19$, $\epsilon_A = -3.90$, $b_p = .05$, $Q_1 = .75$, $Q_2 = .821$, and $Q_3 = 3.67$. The program again returned to the optimum values after 9.5 minutes of calculation.

In order to provide an additional benchmark for evaluating the powers and limitations of the technique, the parameters for another nonlinear model of the demand relationship were estimated. This particular structure was also transformed into a linear form and its parameters estimated by regression analysis. The nonlinear form is:

$$S_{B,t} = a_0 (P_{B,t}/\bar{P}_t)^{a_1} (A_{B,t}/\bar{A}_t)^{a_2} \exp (a_3 Q_1 + a_4 Q_2),$$

and the model transformed for regression analysis is:

$$ln (S_{B,t}) = ln (a_0) + a_1 \cdot ln (P_{B,t}/\bar{P}_t) + a_2 \cdot ln (A_{B,t}/\bar{A}_t) + a_3 Q_1 + a_4 Q_2,$$

where: $S_{B,t}$ is market share for Brand B at time t, $P_{B,t}$ is price (dollars per ounce) for Brand B at time t, $A_{B,t}$ is advertising expenditures (thousands of dollars) for Brand B at time t, $\bar{P}_t$ is average price (weighted by volume) for all three brands for period t, and:

$$\bar{A}_t = \tfrac{1}{3} \sum_B A_{B,t}.$$

The parameters $a_0, a_1, \cdots, a_4$ were estimated twice by SEARCH from the same set of initial conditions but with different termination tolerances. The results of these efforts and the results of the regression analysis are presented in the table.

PATTERN SEARCH AND REGRESSION RESULTS

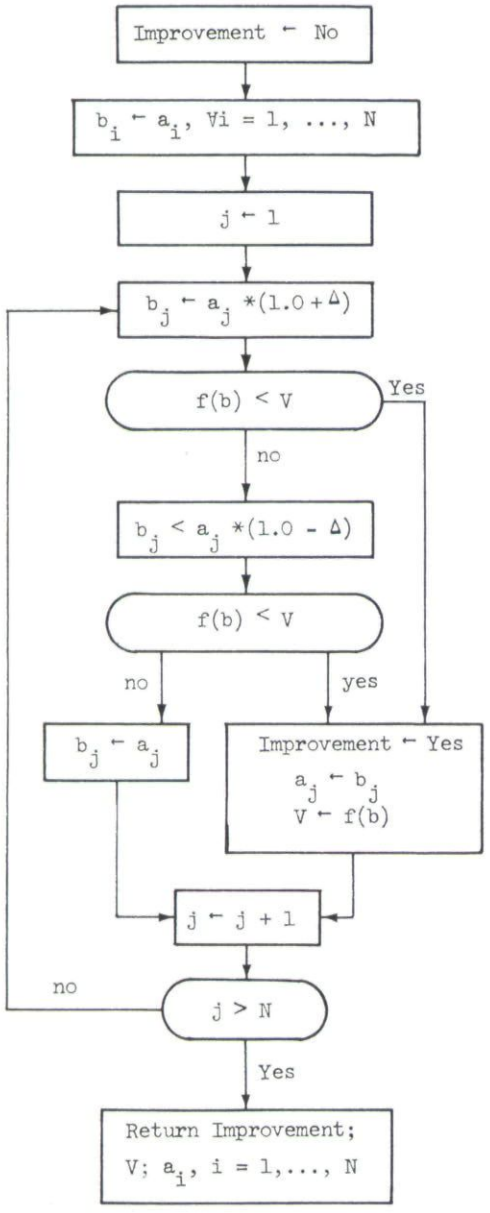
	Pattern search			*Regression*
	Starting conditions	*Results under termination tolerance*		*Results*
		.001	*.0005*	
a_0	1.008	.1604	.1593	.1549
a_1	−0.895	−2.6620	−2.7080	−2.5660
a_2	.0154	.0707	.0615	.0723
a_3	.00878	1.420	1.4400	1.4410
a_4	.5007	.0068	.0045	.0524
Error sum of squares		.1264	.1245	.1196

The table shows the minimum value for the error sum of squares calculated in approximately eleven seconds by the regression analysis[4] was .1196. Pattern search,

[3] All times cited in this article reflect actual computing time on an IBM 7090.

[4] Antilogs of $ln (S_{B,t})$ and $(\hat{S}_{B,t})$ where taken before this value was computed.

Figure 7
FLOW CHART: EXPLORATORY MOVE



Given: N – number of parameters
Δ – step size factor
V – best yet value of function
$a_1 \ldots a_n$ – point to explore about

working with a tolerance of .001, arrived at an estimate of .1264 for the error sum of squares after 128 seconds, while it considered 963 different values of the function. With a tolerance of .0005, pattern search needed an additional 5,000 evaluations and approximately 10 additional minutes to refine this estimate of the error sum of squares from .1264 to .1245. Such small improvement in the critical value indicates that the function is likely quite flat, and thereby insensitive to changes in its parameters' values in the neighborhood of the minimum.

Moreover, since there is some finite amount of error inherent in the matrix inversion routine utilized by the regression program, the minima of the two models are not likely to be identical. For example, pattern search was given the regression estimates as starting values and began work with a tolerance of .00000010. After calculating for 4.9 minutes, the program arrived at a minimum value of .1225 for the error sum of squares and the following estimates of the parameters: $a_0 = .15394$, $a_1 = -2.7550$, $a_2 = .04725$, $a_3 = 1.4918$, and $a_4 = .072607$. Although slightly different from those of the regression program, the results produced by pattern search are sufficiently close to be entirely usable.

Long program running times, as those previously mentioned, suggest another disadvantage of such optimization techniques. The solutions may not be obtainable at a reasonable cost in time and/or money. The times cited above were for the IBM 7090 which rented (at the time of the research) commercially for $450 per hour. Although there are many computers which now enjoy economic advantages over the 7090, direct search procedures still can be very costly.

SUMMARY

This article has discussed the attributes of direct search as a methodology for solving problems in least squares analysis too complex to be analyzed by linear regression techniques. The major costs associated with direct search techniques are their relatively long computer running times. Even for the application involving five parameters and 48 observations, the computer times involved were significantly greater than those associated with similar problems solved with regression analysis. With the advent of the faster computation speeds associated with third generation data processing equipment, however, direct search holds great promise as a methodology for testing complex simulation models.

REFERENCES

1. Robert Hooke and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems," *Journal of the Association for Computing Machinery*, 8 (April 1961), 212–28.
2. Alfred A. Kuehn, "Consumer Brand Choice as a Learning Process," *Journal of Advertising Research*, 2 (December 1962), 10–7.
3. ——— and Timothy W. McGuire, "A Model of Consumer Behavior and its Implications for a Theory of Imperfect Competition," paper presented at the Winter Meetings of the Econometric Society, December 1962.

4. Alfred A. Kuehn and Doyle L. Weiss, "The Marketing Analysis Training Excercise," *Behavioral Science,* 10 (January 1965), 51–67.
5. Kenneth Levenberg, "A Method for Solution of Certain Non-Linear Problems in Least Squares," *Quarterly of Applied Mathematics,* 2 (1944–45), 164–8.
6. Doyle L. Weiss, "An Analysis of Market Share Behavior for a Branded Consumer Product," unpublished doctoral dissertation, Graduate School of Industrial Administration, Carnegie-Mellon University, 1966.
7. Douglass J. Wilde, *Optimum Seeking Methods,* Englewood Cliffs, N.J.: Prentice-Hall, 1964.
8. ——— and C. S. Beightler, *Foundations of Optimization,* Englewood Cliffs, N.J.: Prentice-Hall, 1966.
9. E. T. Whittaker and G. Robinson, *The Calculus of Observations,* London: Blackie and Son, 1937.