

A SIMULATED ANNEALING METHODOLOGY FOR CLUSTERWISE LINEAR REGRESSION

WAYNE S. DESARBO

MARKETING AND STATISTICS DEPARTMENTS
GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF MICHIGAN

RICHARD L. OLIVER AND ARVIND RANGASWAMY

MARKETING DEPARTMENT
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA

In many regression applications, users are often faced with difficulties due to nonlinear relationships, heterogeneous subjects, or time series which are best represented by splines. In such applications, two or more regression functions are often necessary to best summarize the underlying structure of the data. Unfortunately, in most cases, it is not known a priori which subset of observations should be approximated with which specific regression function. This paper presents a methodology which simultaneously clusters observations into a preset number of groups and estimates the corresponding regression functions' coefficients, all to optimize a common objective function. We describe the problem and discuss related procedures. A new simulated annealing-based methodology is described as well as program options to accommodate overlapping or nonoverlapping clustering, replications per subject, univariate or multivariate dependent variables, and constraints imposed on cluster membership. Extensive Monte Carlo analyses are reported which investigate the overall performance of the methodology. A consumer psychology application is provided concerning a conjoint analysis investigation of consumer satisfaction determinants. Finally, other applications and extensions of the methodology are discussed.

Key words: cluster analysis, combinatorial optimization, regression analysis, simulated annealing, consumer psychology.

1. Introduction

Least-squares multiple regression is one of the most frequently utilized techniques to examine the effect of independent variables on a specified dependent variable. Let:

- $i = 1, \dots, I$ observations/subjects;
- $j = 1, \dots, J$ independent variables;
- y_i = the i -th observation/subject's dependent variable value;
- x_{ij} = the i -th observation/subject's value on the j -th independent variable;
- e_i = error;
- b_j = the j -th regression coefficient (to be estimated).

Then, the traditional linear model can be written as:

The authors wish to thank the editor, associate editor, and three anonymous reviewers for their insightful comments and thorough review of this manuscript.

Requests for reprints should be sent to Wayne S. DeSarbo, S. S. Kresge Professor of Marketing and Statistics, Marketing and Statistics Departments, Graduate School of Business, University of Michigan, Ann Arbor, MI 48109-1234.

$$y_i = \sum_{j=1}^J x_{ij}b_j + e_i \quad (1)$$

or

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad (2)$$

where $\mathbf{y} = ((y_i))$, $\mathbf{X} = ((x_{ij}))$, $\mathbf{b} = ((b_j))$, and $\mathbf{e} = ((e_i))$. Given a sample of observations for \mathbf{y} and \mathbf{X} , one is typically interested in solving for the regression coefficients b_j in order to minimize the following error sums of squares:

$$\begin{aligned} \text{Min } H &= \sum_{i=1}^I \left[y_i - \sum_{j=1}^J x_{ij}b_j \right]^2 \\ &= \sum_{i=1}^I e_i^2. \end{aligned} \quad (3)$$

The closed form expression for estimating \mathbf{b} that minimizes (3) is

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (4)$$

Maddala (1976) and others have reviewed the numerous statistical properties of such an estimator.

There are several occasions where the estimation of a single set of regression coefficients via (4) may be problematic and potentially misleading. For example, suppose one were interested in utilizing linear regression to approximate a nonlinear relationship as shown in Figure 1. Figure 1 presents an adaptation of Locke's (1982) example of nonlinear performance response to increasingly difficult goal levels, a phenomenon which has been empirically demonstrated in the organizational psychology literature (Locke, 1967; Locke, Shaw, Saari, & Latham, 1981). One estimated linear regression function (approximated by the straight longer line in Figure 1) would give the impression that higher goal levels (even in the impossible range) would result in equivalently high performance. This one regression function would also result in a nonrandom pattern of residuals. A more accurate representation here would be to estimate two regression functions (i.e., splines or switching regressions, Quandt, 1972) indicated by the two connected lines. Here, the observations can be arranged by goal levels after which those from the easy to difficult range can be used to estimate the first regression line with highly positive slope, and the observations from the difficult through impossible range can be used to estimate the second line with near zero slope. Often times, plotting such data can help determine such points of inflection.

In other applications, this plotting is nearly impossible to accomplish. For example, Figure 2 presents consumer derived utility for a hypothetical product as a function of price. Such a representation is quite common as a result of conjoint analysis studies (Green & Rao, 1971) which decompose overall preference or intention to buy judgments into part worth utilities for the specified attributes and their levels of the stimulus. Figure 2 depicts a common result where, for one group of consumers, higher utility corresponds to lower prices while the opposite is true for the other group (as shown by the two dashed lines). The former reflects traditional microeconomic views of household utility and negative price elasticity (Henderson & Quandt, 1982) while the latter group may react to price as an indicator of brand quality (Gabor & Granger, 1966;

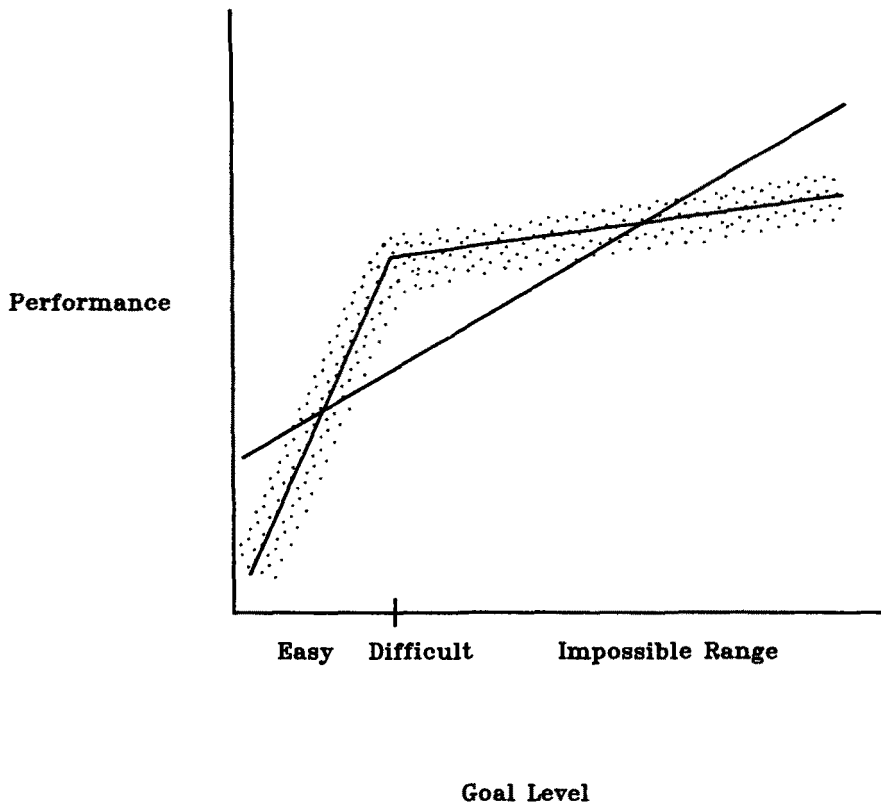


FIGURE 1.
Nonlinear performance response.

Sowter, Gabor, & Granger, 1971). The solid horizontal line indicates the one regression line estimated for the total sample, indicating the need for two regression lines here. In such analyses of preference or demand, it may be impossible to a priori group the consumers into these two groups who react quite differently to price.

There are several other examples (e.g., splines in time series work, market segmentation studies, etc.) where such natural groupings of the observations arise which have substantially different regression functions. In many cases, it is extremely difficult or impossible to a priori sort out these observations/subjects into groups and clusters, especially in those situations involving several independent variables. This paper presents a methodology that simultaneously estimates cluster membership and corresponding regression functions for a sample of observations/subjects. We first review related procedures. Next, the new methodology is presented with its simulated annealing based algorithm. An extensive set of Monte Carlo analyses are presented which examine the performance of the algorithm. A consumer psychology application concerning consumer satisfaction determinants is presented next. Lastly, a discussion of future research areas is provided.

2. Related Methods

If there are sufficient replications per subject which would allow the estimation of the b 's per subject, one could attempt to solve this clustering-regression problem via a traditional three-step procedure. In step one, individual level b 's are estimated by subject. In step two, these b coefficients are submitted to a cluster analysis to form

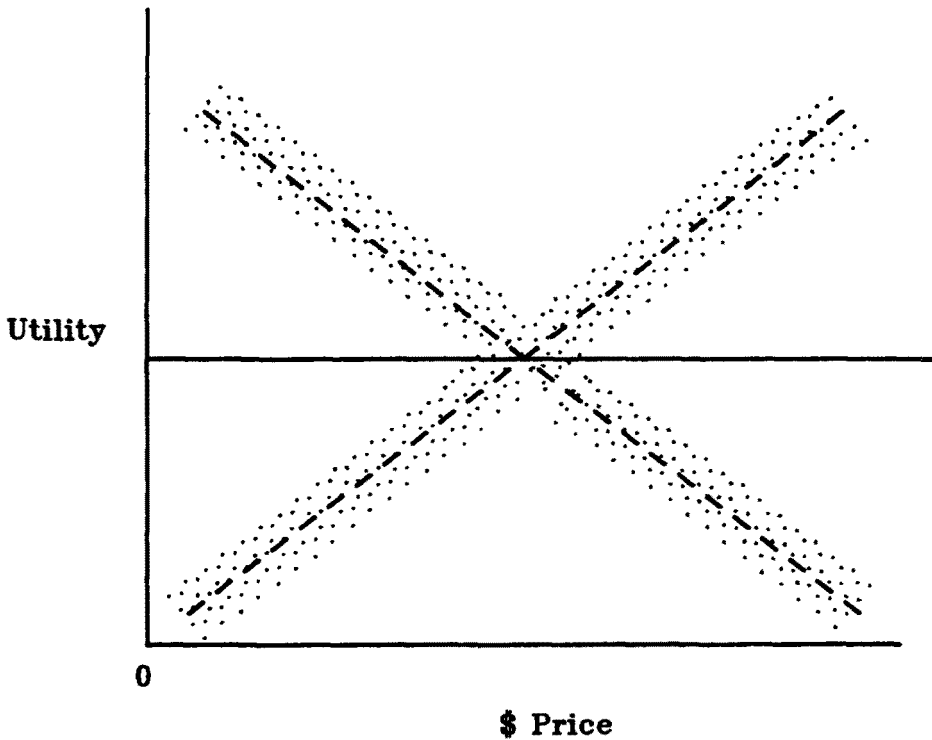


FIGURE 2.
Utility functions and price.

homogeneous groups. In step three, according to the resulting partitions derived from the cluster analysis from the previous step, one can then estimate a separate regression function for each cluster or group. There are, of course, several problems with this approach. One is that the resulting partitions would likely differ depending upon the specific clustering algorithm utilized. A second major drawback is that the regression and clustering procedures typically optimize different criteria and may not be congruent. Finally, most studies do not collect replications by subject, making such individual level estimation impossible.

DeSarbo, Carroll, Clark, and Green (1984) developed the SYNCLUS methodology that simultaneously solves for a non-overlapping clustering and a set of rescaling constants for the variables, all to optimize one common objective function. DeSarbo and Mahajan (1984) later extended this SYNCLUS methodology to accommodate constraints, different types of clustering schemes, and a general linear transformation of the variables. De Soete, DeSarbo, and Carroll (1985) have derived an optimal variable weighting scheme for hierarchical clustering where both variable weights and ultrametric tree are simultaneously estimated. Note, however, that none of these approaches are appropriate for a regression context with dependent and independent variables.

Haggerty (1985) recently devised a methodology to improve the predictive accuracy of regression based preference or conjoint analysis by grouping respondents with similar preferences. Assume that I subjects each evaluate L stimuli (e.g., brands or products), where these L stimuli are defined by a design matrix X . Let Y be the $L \times I$ matrix of responses (dependent variable) where each column of Y is the data for an individual respondent. Haggerty assumes that X and Y are column standardized to zero mean and unit standard deviation. To compute individual-level regressions on Y , the least squares estimates are still $B = (X'X)^{-1}X'Y$. However, Haggerty attempts to

“improve” on these columns of \mathbf{B} by weighting similar subjects together in some fashion. He attempts to find a transformation matrix \mathbf{A} which transforms the responses via:

$$\mathbf{Z} = \mathbf{Y}\mathbf{A} = \mathbf{Y}\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}', \quad (5)$$

where \mathbf{S} is an $I \times K$ matrix (K = the number of factors or clusters), in order to maximize predictive accuracy which is defined as the average (over all subjects and validation trials) expected mean squared error (EMSE) between prediction and validation trial (i.e., from a holdout sample):

$$\begin{aligned} Q_s &= E \left[\frac{1}{I} \sum_{i=1}^I \frac{1}{M} \sum_{m=1}^M (V_{im} - \hat{Y}_{im})^2 \right] \\ &= (1 - P^2) + \frac{(1 - P^2)J \operatorname{tr}(\mathbf{A})}{MI} + \frac{\operatorname{tr}[(\mathbf{I} - \mathbf{A})\mathbf{Y}'_T\mathbf{Y}_T]}{MI}, \end{aligned} \quad (6)$$

where:

V_{im} = subject i 's m -th validation trial response,

\hat{Y}_{im} = subject i 's predicted value for the response on the m -th validation trial,

P^2 = the true proportion of variance accounted for by the regression in the population,

$\mathbf{Y}'_T\mathbf{Y}_T/M$ = the matrix of true correlation between respondents over the M validation trials.

Assuming $\mathbf{S}'\mathbf{S} = \mathbf{I}$, Haggerty (1986) estimates \mathbf{S}^* to minimize Q_s as the first C^* eigenvectors of the correlation matrix between subjects, where C^* is the number of eigenvalues greater than $1/I(1 - P^2)J/M$. Hence, Haggerty's optimal weighting scheme reduces to a principal axis factor analysis of the correlation matrix of respondents. As such, one would rarely end up with an overlapping or nonoverlapping clustering scheme in \mathbf{S}^* unless explicit constraints were placed on the elements in \mathbf{S}^* which would result in a more complicated algorithm. In addition, there are the typical interpretive problems associated with nonunique solutions given the well-known rotational indeterminacy of such factor solutions. Finally, the procedure has been tailored for conjoint analysis with replications per subject which are not typically collected in most regression problems. Similar adaptations to conjoint analysis include Ogawa's (1987) logit approach and Kamakura's (1988) hierarchical clustering least-squares procedure.

One of the two most successful treatments of the problem to date is the “clusterwise regression” methodology due to Späth (1979, 1981, 1982, 1985). He devised an exchange algorithm to form a partition of length K and corresponding sets of parameters \mathbf{b}_k such that the sum of the error sums of squares over all clusters is minimized:

$$\operatorname{Min} F = \sum_{k=1}^K \|\mathbf{X}^k \mathbf{b}_k - \mathbf{y}^k\|^2. \quad (7)$$

where

\mathbf{X}^k = the observations of the vector of independent variables for the objects classified in cluster k ;

\mathbf{y}^k = the observations of the dependent variable for the objects classified in cluster k .

Note, to guarantee the existence of a solution \mathbf{b}_k , it is required that the rank $\mathbf{X}^k = J$. A necessary condition for this is $I_k \geq J \Rightarrow I \geq KJ$. This procedure simultaneously solves for the optimal feasible partition $P(K, I)$ and regression weights per cluster \mathbf{b}_k that minimize (7). For the L_2 norm (e.g., the error sums of squares) in (7), Späth (1982, 1985) has developed efficient up- and down-dating formulas for the solution of these regression problems when an observation is added or removed utilizing QR-decompositions. This stepwise-optimal method works sequentially on the observations and is conceptually similar to K -means (MacQueen, 1967). The procedure can be summarized as follows:

1. Choose some initial partition P_1, \dots, P_K such that $|P_k| \geq J$, and some starting observation;

2. Set $t = t + 1$ and reset $t = 1$ if $t > I$. For $i \in P_j$ and $|P_j| > I_k^* (I_k^* > J)$, examine whether there are clusters P_k with $k \neq j$ such that shifting observation i from P_j to P_k reduces the objective function, (7). If so, then choose P_r such that the reduction becomes maximal and redefine $P_j: P_j - \{i\}$, $P_r = P_r \cup \{i\}$. Otherwise return to Step 2.

3. Repeat Step 2 as long as you get any reduction in the objective function, else stop.

According to Späth (1982), the final result depends on the initial partition, on the starting observation, and on the choice of I_k^* , a minimum number of observations in each cluster. Because of severe problems with locally optimal solutions, Späth recommends running multiple analyses for a prespecified K altering the initial starting partition and several values for the starting observation and I_k^* . (Späth (1987) has since updated his clusterwise linear regression procedure given a "slight error" in the previous version.) While Späth's procedure renders a non-overlapping partitioning of the observations and simultaneous estimates of \mathbf{b}_k , it is somewhat limited in its use. In its present form, it cannot readily accommodate multivariate specifications in \mathbf{Y} , that is, the collection and use of multiple dependent variables. It is unable to render overlapping clusters. It cannot handle replications per subject such as those found in conjoint analyses of preference. In its present form, it cannot accommodate user-specified constraints on cluster membership such as those discussed in DeSarbo and Mahajan (1984).

The other successful treatment of this problem is the conditional mixture, maximum likelihood methodology recently proposed by DeSarbo and Cron (1988). The authors model the dependent variable, y_i , as a finite sum or mixture of conditional univariate normal densities:

$$y_i \sim \sum_{k=1}^K \lambda_k f_{ik}(y_i | x_{ij}, \sigma_k^2, b_{jk}) \quad (8)$$

$$= \sum_{k=1}^K \lambda_k (2\pi\sigma_k^2)^{-1/2} \exp \left[\frac{-(y_i - \sum_{j=1}^J x_{ij} b_{jk})^2}{2\sigma_k^2} \right], \quad (9)$$

where

b_{jk} = the value of the j -th regression coefficient for the k -th cluster;

σ_k^2 = the variance term for the k -th cluster; and

λ_k = the (unknown) mixture proportion where $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^K \lambda_k = 1$.

Given a sample of I independent subjects or observations, K , y , and X , DeSarbo and Cron (1988) estimate λ_k , σ_k^2 , and b_{jk} that maximize the following log-likelihood expression:

$$\ln L = \sum_{i=1}^I \ln \left[\sum_{k=1}^K \lambda_k (2\pi\sigma_k^2)^{-1/2} \exp \left[\frac{-(y_i - \sum_{j=1}^J x_{ij}b_{jk})^2}{2\sigma_k^2} \right] \right], \quad (10)$$

using a modified EM-algorithm. This methodology produces a “fuzzy” clustering of the subjects/observations via application of Bayes’ rule. While this approach displays considerable merit, especially given its performance in Monte Carlo experiments, it is limited in terms of being unable to accommodate multiple dependent measures, overlapping clusters, user-imposed membership constraints, or multiple replications per subject.

The goals of this research are to provide a flexible methodology for clusterwise regression which can simultaneously estimate regression parameters \mathbf{b}_k and a partitioning $P(I, K)$ which accommodates:

1. univariate or multivariate \mathbf{Y} ,
2. overlapping or nonoverlapping clusters,
3. replications per subject,
4. user-specified constraints on cluster membership.

3. Methodology

The Model

Let:

- $s = 1, \dots, S$ subjects,
- $r = 1, \dots, R$ replications per subject,
- $i = 1, \dots, I$ total observations = SR ,
- $j = 1, \dots, J$ independent variables,
- $k = 1, \dots, K$ prespecified number of clusters,
- $l = 1, \dots, L$ dependent variables.

The full general clusterwise-regression model can be written as:

$$y_{il} = \sum_{k=1}^K \sum_{j=1}^J a_{ik} x_{ij} b_{jl}^k + e_{il}, \quad (11)$$

where:

- y_{il} = the i -th observation of the l -th dependent variable;
- x_{ij} = the i -th observation of the j -th independent variable;
- b_{jl}^k = the value of the j -th regression coefficient for predicting the l -th dependent variable in the k -th cluster; 0
- $a_{ik} = \begin{cases} 1 & \text{if observation } i \text{ is assigned to cluster } k, \\ 0 & \text{else} \end{cases}$
- e_{il} = error.

Given $\mathbf{Y} = ((y_{il}))$, $\mathbf{X} = ((x_{ij}))$, and K , the objective is to estimate $\mathbf{A} = ((a_{ik}))$ and $\mathbf{B} = ((b_{jl}^k))$ in order to minimize:

$$\Phi = \sum_{i=1}^I \sum_{l=1}^L \left[y_{il} - \sum_{k=1}^K \sum_{j=1}^J a_{ik} x_{ij} b_{jl}^k \right]^2, \quad (12)$$

subject to $I_k > J$ for all k , where I_k is the number of observations assigned to cluster k , and subject to any user specified constraints concerning cluster membership (DeSarbo & Mahajan, 1984). Given the model in (11) and the objective function in (12), one can calculate a variance accounted for statistic for the total data set, as well as one for each dependent variable. (Given the form of (12), for multiple dependent variables, it is assumed that the dependent measures are comparably scaled.) Note, for $L = 1$, $R = 1$, nonoverlapping clusters, and no constraints, (12) reduces to (7) in the sense of handling Späth's (1985) model as a special case.

The Algorithm

A generalized simulated annealing procedure is utilized in combination with multiple regression to estimate a_{ik} and b_{jl}^k in (11) in order to either maximize total variance accounted for ($\sum_l G_l^2$), or minimize Φ in (12) or a bounded minimand (given possible scale differences in the measurement of multiple dependent variables, \mathbf{Y}) $Z = L - \sum_l G_l^2$, where G_l^2 is defined as an r -square measure:

$$G_l^2 = \left(1 - \frac{\sum_i (y_{il} - \hat{y}_{il})^2}{\sum_i (y_{il} - \bar{y}_{.l})^2} \right) \quad (13)$$

where $\hat{y}_{il} = \sum_k \sum_j a_{ik} x_{ij} b_{jl}^k$ and $\bar{y}_{.l}$ is the mean of the l -th dependent variable. The simulated annealing procedure historically derives its name from the "annealing" or cooling process of heated metals in which many final crystalline configurations which correspond to different energy states are possible depending upon the rate of the associated cooling process. According to Kirkpatrick, Gelatt, and Vecchi (1983), this procedure can be traced to Metropolis who originally attempted to simulate the behavior of an ensemble of atoms in equilibrium at a given temperature. Metropolis constructed a mathematical model of the behavior of such a system that contained a method for minimizing the total energy of the system.

It is a fact that the atoms of a molten metal when cooled to a freezing temperature will tend to assume relative positions in a lattice in such a way as to minimize the potential energy of their mutual forces. Because of the huge number of atoms and resulting possible lattice arrangements (a combinatorial problem), the final derived state will typically correspond to only a local optimum and not a global one. The resulting solidified metal can be reheated and cooled again slowly with the hope that it will then move to a lower state of energy. This metallurgic process is called annealing and the mathematical modeling of this process is called simulated annealing.

Computationally, simulated annealing has been devised as a general optimization methodology to find the extremum of a function that may have several local extrema and may not be regular (continuous or smooth). The technique is based on a controlled random search that samples the objective function in a feasible region of the parameter space. In our clusterwise-regression methodology, simulated annealing is utilized to estimate a_{ik} . At each accepted step, it is followed by multiple regression to estimate b_{jl}^k . According to Kirkpatrick, Gelatt, and Vecchi (1983), Levy and Montalvo (1985),

Lundy (1986), Bohachevsky, Johnson, and Stein (1986), and van Laarhoven and Aarts (1987), the traditional simulated annealing procedure can be summarized (for a minimization problem) in general as follows (see Bohachevsky, et al.).

Starting with an initial feasible solution selected randomly, the corresponding value of the objective function, Z_0 , is calculated. Next, a random point is selected on the surface of the unit m -dimensional hypersphere, where m is the dimensionality of the optimization problem. This specifies a random direction in which a step of size Δ is taken. Now, the new value of the objective function, Z_1 , is calculated and this new solution is accepted with probability p where:

$$p = \begin{cases} 1 & \text{if } \Delta Z = Z_1 - Z_0 \leq 0 \\ \exp(-\alpha \Delta Z) & \text{if } \Delta Z > 0, \end{cases} \quad (14)$$

where $\alpha > 0$. This differentiates simulated annealing from most other combinatorial optimization procedures. Here, a detrimental step may be accepted according to some probability distribution. A random number β is generated from a uniform distribution, $\cup(0, 1)$, and compared to $\exp(-\alpha \Delta Z)$. If $\beta < \exp(-\alpha \Delta Z)$, then the step is accepted; else the step is rejected and a move in a different random direction is attempted. Thus, the probability of accepting a worse solution is inversely related to the amount of the increase in the objective function. Note that the expression for the conditional acceptance probability p is motivated by the physical process modeled by simulated annealing. In statistical mechanics, the probability that the system will move from energy state E_1 to energy state E_2 with $E_2 - E_1 = \Delta E > 0$ is $\exp(-\Delta E/\gamma T)$, where γ is the Boltzmann constant and T is the absolute temperature. Thus, $\alpha = 1/\gamma T$, and the lower the temperature, the smaller the probability of movement to a higher energy state. This nonzero probability of moving away from potential locally optimal solutions differentiates the simulated annealing procedure from most other combinatorial optimization methods and has been cited as the primary reason for the "success" in the overall performance of this procedure.

Bohachevsky, Johnson, and Stein (1986) have recently examined the performance of this simulated annealing algorithm in the solution of continuous statistical-related optimization problems. They found that the random solution path wanders excessively in and out of various extremes and does not provide a logical termination criterion. The authors state that because of this undesirable property, the method tends to require an excessively large number of steps to find a satisfactory solution. To overcome these numerical problems, Bohachevsky, Johnson, and Stein have extended simulated annealing by making the probability of accepting a detrimental step tend to zero as the random walk approaches the extremum. They found that this modification leads to a solution path that more often terminates at the global extremum and requires fewer steps than the standard annealing method. Our experimentation with a related modification and the traditional approach with clusterwise regression also bears this out, and it is our related modification described below which we adopt for use.

The specific modified combinatorial simulated annealing and OLS procedure utilized for our generalized clusterwise regression problem is summarized as follows:

1. K , Y , X , S , R , constraints, and type of clustering is specified as input (desired constraints are embedded in a subroutine).

2. Depending upon K the type of clustering selected, and type of constraints, a matrix of feasible binary row vectors is created in V which render feasible cluster memberships for any subject. Thus, V contains the complete set of feasible binary row vectors, indicating cluster membership, over which the simulated annealing procedure

will search. V will contain D rows and K columns where D will depend on K , the type of clustering, and the constraints desired.

3. An initial solution A^0 is randomly generated from V and the associated b_k^0 are estimated via OLS. The objective function Z_0 is calculated. Set $t = 0$.

4. Now, r rows (subjects) in A^0 are randomly altered to create A^1 . A^1 is again checked for feasibility in terms of satisfying any set of user specified constraints. If not feasible, a different A^1 is generated until feasible. If feasible, b_k^1 is calculated via OLS and so is Z_1 .

5. If $Z_1 \leq Z_0$, and $A^0 = A^1$ and $Z_0 = Z_1$, return to Step 4.

6. If $Z_1 > Z_0$, calculate $p = \exp(-\alpha\Delta Z)$ where $\alpha = 1 + (\gamma \times t)$, with $\gamma = 0.95$ (here, γ controls the speed or "cooling schedule" at which the annealing process approaches the final solution). Generate a uniform 0 - 1 variate μ ; if $\mu \geq p$, set $t = t + 1$; if $t < \text{maxit}$, go to Step 4; else stop. If $\mu < p$, set $A^0 = A^1$, $Z_0 = Z_1$, and go to Step 4.

Program Options

A number of options exist for preprocessing data, selecting different model specifications, and specifying values for the algorithm's operation. Concerning data preprocessing, program options currently exist to center, normalize, or standardize Y and/or X (rows or columns). In addition, an option exists to initially orthogonalize X via principal components and use the derived principal component scores as independent variables. With respect to model specification, the user can fit either overlapping or nonoverlapping clusters, with or without (prespecified) constraints (DeSarbo & Mahajan, 1984). In addition, more than one dependent variable can be used, as well as having replications per subject/observation. Finally, concerning algorithm control parameters, the user can specify maxit , γ , the number of rows to be perturbed in A , and the number of "cooking iterations"—the maximum number of times the procedure will accept an inferior solution. An option for combinatorial polishing is provided where singleton changes are allowed to attempt to improve the loss function using the Lin and Kernighan (1973) procedure.

While theoretical results for simulated annealing have been devised concerning convergence to global versus local optima given γ and maxit (Gidas, 1985; Mitra, Romeo, and Sangiovanni-Vincentelli, in press), the computational effort necessary to guarantee these results tends to be considerable from a practical standpoint. This is particularly relevant given the recent experience of De Soete, Hubert, and Arabie (1988a, 1988b) with simulated annealing in solving combinatorial data analysis problems in unidimensional seriation and Dubes and Klein (1987) regarding its use in data analysis. (Also, see Aarts & Korst, 1989; Davis, 1987; Goldberg, 1989; Johnson, 1989; and van Laarhoven & Aarts, 1987). There is, therefore, a considerable need to investigate these properties on real and/or synthetic data.

4. Monte Carlo Analyses

Simulated Annealing Methodology Tests

In order to rigorously examine the performance of this simulated annealing based methodology, a Monte Carlo analysis was designed to analyze a number of synthetically constructed data sets as a variety of data, error, and model factors were experimentally manipulated. DeSarbo (1982), DeSarbo and Carroll (1985), and De Soete, DeSarbo, Carroll, and Furnas (1984) have utilized such an approach in testing the performance of various clustering and multidimensional scaling algorithms. Table 1

presents a list of some twelve factors and the associated levels tested. As shown, these twelve factors span the areas of data size, type of clustering, single or multiple dependent variables, replications, error in the data, as well as a number of algorithm control parameters which may impact the performance of the simulated annealing methodology. As in the previous psychometric and classification literature cited above where such studies have been historically performed, an asymmetric fractional factorial design (Addelman, 1962) is devised for main effects estimation. Table 2 presents the specific $3^5 2^7$ design and the specific trials used to create the synthetic data and algorithm parameter settings in this Monte Carlo analysis. The four dependent measures collected were:

1. the root mean square (RMS) between y_{ij} and $\hat{y}_{ij} = \sum_k \sum_j a_{ik} x_{ij} b_{jk}$;
2. the RMS between \mathbf{B} and $\hat{\mathbf{B}}$ after appropriate permutation;
3. the simple matching coefficient between \mathbf{A} and $\hat{\mathbf{A}}$ after appropriate permutation;
4. elapsed CPU time measured in minutes on a Vax 8550.

These measures deal with the major areas of data recovery, parameter recovery, and computational effort.

Tables 3 through 6 present the four ANOVAs performed for the various dependent measures. Table 3 displays the table for the first dependent variable: the root mean square (RMS) between \mathbf{Y} and $\hat{\mathbf{Y}}$. As can be seen, no main effect is significant in predicting this RMS dependent variable, indicating somewhat consistent fitting over different data types and model specifications. Of particular surprise here is the fact that the error factor (X_7) is not significant. While inspection of the cell means indicates higher RMS values for higher error levels, the differences are not significant.

Table 4 presents the ANOVA results on the \mathbf{A} recovery measure. Here, the number of replications (X_5), combinatorial polishing (X_{12}), number of rows (X_1), and value of γ (X_8) are significant factors. Upon inspection of the cell means for these factors, recovery of the true \mathbf{A} matrix of cluster membership *deteriorates* as: there are fewer replications per subject, there is no combinatorial polishing requested, there are more subjects or rows in the data, and smaller values of γ are specified in the simulated annealing process. (Note, similar results were found using logit and arcsin transformations on this dependent measure.)

Table 5 presents the ANOVA results for the RMS dependent measure between \mathbf{B} and $\hat{\mathbf{B}}$. The only significant factor present is the number of clusters (X_3). Upon inspection of the cell means, it appears that \mathbf{B} recovery is somewhat more difficult for larger cluster specifications, although this result is significant at only $p \leq .10$.

Finally, Table 6 presents the ANOVA results concerning the CPU time dependent measure. Here, the number of clusters (X_3), number of replications (X_5), γ values (X_8), and the number of rows altered in the simulated annealing process (X_9) are significant factors. From a cursory examination of the cell means, it appears that CPU time significantly increases when there are larger cluster specifications, larger number of replications per subject are present, γ is set at .95, and only one row in \mathbf{A} is altered at a time.

While these Monte Carlo results look promising, a number of legitimate concerns must be raised. Given the complexity of the design with so many different factors, a full factorial design with replications, while more rigorous, was computationally impossible. There may be significant interactions present that we cannot measure with a main effect design. The next Monte Carlo analysis examines a smaller full factorial design in comparing the Späth (1985, 1987) procedure with our simulated annealing methodology for a restricted class of models.

TABLE 1

Experimental Design Factors for First Monte Carlo Analysis

<u>Factor</u>	<u>Levels</u>	<u>Codes</u>
X ₁ . No. of Rows	I = 40	0
	I = 65	1
X ₂ . No. of Independent Variables	J = 3	0
	J = 6	1
X ₃ . No. of Clusters	K = 2	0
	K = 4	1
X ₄ . Clustering Type	Nonoverlapping	0
	Overlapping	1
X ₅ . No. of Replications	R = 1	0
	R = 3	1
X ₆ . No. of Dependent Variables	L = 1	0
	L = 2	1
X ₇ . Error in Data (v = the variance of y _{i1})	None	0
	N(0, .3v)	1
	N(0, .6v)	2
X ₈ . γ	.35	0
	.65	1
	.95	2
X ₉ . No. of Rows Altered	1	0
	2	1
	3	2
X ₁₀ . No. of Cooking Iterations	100	0
	200	1
	300	2
X ₁₁ . No. of Major Iterations	100	0
	200	1
	300	2
X ₁₂ . Combinatorial Polishing	No	0
	Yes	1

TABLE 2

 3^{527} Asymmetric Fractional Factorial Design

<u>Trial</u>	<u>Independent Factor:</u>											
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	2	1	2	1	2	0
3	0	0	0	0	0	0	1	2	1	2	1	1
4	0	1	1	0	0	0	1	1	1	2	2	0
5	0	1	1	0	1	1	0	2	0	0	1	0
6	0	1	1	0	0	0	2	0	2	1	0	1
7	0	0	0	1	0	0	2	2	2	1	1	0
8	0	0	0	1	1	1	1	0	1	2	0	0
9	0	0	0	1	0	0	0	1	0	0	2	1
10	1	0	1	1	0	1	0	1	1	1	1	1
11	1	0	1	1	1	0	2	2	0	2	0	0
12	1	0	1	1	0	0	1	0	2	0	2	0
13	1	1	0	0	0	1	1	2	2	0	0	1
14	1	1	0	0	1	0	0	0	1	1	2	0
15	1	1	0	0	0	0	2	1	0	2	1	0
16	1	0	0	0	0	1	2	0	0	2	2	1
17	1	0	0	0	1	0	1	1	2	0	1	0
18	1	0	0	0	0	0	0	2	1	1	0	0
19	0	0	0	0	0	0	0	2	2	2	2	0
20	0	0	0	0	1	0	2	0	1	0	1	1
21	0	0	0	0	0	1	1	1	0	1	0	0
22	0	1	0	1	0	0	1	0	0	1	1	0
23	0	1	0	1	1	0	0	1	2	2	0	1
24	0	1	0	1	0	1	2	2	1	0	2	0
25	0	0	1	0	0	0	2	1	1	0	0	0
26	0	0	1	0	1	0	1	2	0	1	2	1
27	0	0	1	0	0	1	0	0	2	2	1	0

Comparison Tests with Späth's Methodology

We performed comparison tests with the single dependent variable, only one replication, no constraints, and the nonoverlapping cluster option of our methodology with that of Späth's (1987). Table 7 presents the list of factors used in the full factorial experiment with two replications per trial (i.e., 162 runs were performed with each methodology). As before, the specific trial designated the size of the synthetic data set, as well as the amount of error to be introduced. Each of the 162 data sets was then

TABLE 3

ANOVA Results on \bar{Y} Recovery

<u>Source</u>	<u>SS</u>	<u>DF</u>	<u>MS</u>	<u>F</u>
X ₁	.048	1	.048	1.41
X ₂	.040	1	.040	1.18
X ₃	.010	1	.010	0.29
X ₄	.005	1	.005	0.15
X ₅	.009	1	.009	0.26
X ₆	.070	1	.070	2.06
X ₇	.096	2	.048	1.18
X ₈	.115	2	.057	1.68
X ₉	.145	2	.072	2.12
X ₁₀	.020	2	.010	0.29
X ₁₁	.001	2	.000	0.02
X ₁₂	.002	1	.002	0.06
Explained:	.571	17	.034	1.00
Unexplained:	.306	9	.034	
Total:	.877	26		

r-square = 0.652

* p ≤ .10
 ** p ≤ .05
 *** p ≤ .01

analyzed with the two methodologies. The dependent measures collected were the same four dependent measures collected in the first Monte Carlo analysis previously reported: RMS (Y , \hat{Y}), RMS (B , \hat{B}), simple matching coefficient of A and \hat{A} , and CPU time. Table 8 presents the means and standard deviations for these four dependent measures across all 162 trials. As can be seen, the performance of both procedures are nearly identical for the first three dependent measures involving data and parameter recovery. The real significant difference occurs in CPU time where the Späth procedure clearly dominates the simulated annealing approach. Thus, it appears that the simulated annealing methodology will perform nearly the same as Späth's for the specialized case of a single dependent variable with nonoverlapping clusters, no constraints, and no replications with respect to data and parameter recovery, but with much higher computational effort (a finding very consistent with the work of De Soete, et al., 1987; Dubes & Klein (1987); and van Laarhoven & Aarts, 1987).

Tables 9 through 12 present the ANOVAs for each dependent measure on the difference (simulated annealing—Späth, 1987, procedure) of the results. Table 9 displays the full factorial ANOVA results for the RMS (Y , \hat{Y}) differences. As seen, there are a number of significant higher order interactions which account for differences in

TABLE 4
ANOVA Results on \hat{A} Recovery

<u>Source</u>	<u>SS</u>	<u>DF</u>	<u>MS</u>	<u>F</u>
X ₁	.038	1	.038	4.07*
X ₂	.000	1	.000	0.00
X ₃	.019	1	.019	2.04
X ₄	.014	1	.014	1.50
X ₅	.094	1	.094	10.07***
X ₆	.000	1	.000	0.00
X ₇	.015	2	.008	0.86
X ₈	.065	2	.032	3.43*
X ₉	.054	2	.027	2.89
X ₁₀	.048	2	.024	2.57
X ₁₁	.008	2	.004	0.43
X ₁₂	.104	1	.104	11.14***
Explained:	.464	17	.027	2.99**
Unexplained:	.084	9	.009	
Total:	.548	26		

r-square = 0.847

* $p \leq .10$
 ** $p \leq .05$
 *** $p \leq .01$

the performance of these two methodologies. Based upon an inspection of the cross classified cell means, one observes that the simulated annealing methodology outperforms Späth's procedure (mean = -0.49) where there is no error ($X_4 = 1$), $I = 35$ ($X_1 = 1$), and $J = 6$ ($X_2 = 3$), whereas the reverse holds true (mean = 1.40) for no error ($X_4 = 1$), $I = 100$ ($X_1 = 3$), and $J = 6$ ($X_2 = 3$). Similarly, the simulated annealing procedure outperforms Späth's procedure in the no error case ($X_4 = 1$) with $K = 2$ clusters ($X_3 = 1$), and $J = 4$ or 6 ($X_2 = 2, 3$) (means = -.39, -.45). The reverse is true for $X_4 = 1$, $X_2 = 1$, $X_3 = 2$ or 3 (i.e., $K = 3$ or 4 clusters) (means = .77, .89).

Table 10 displays the ANOVA results on the differences in the simple matching coefficient for (A, \hat{A}) recovery. While no interactions appear to be significant at $p \leq .05$, there are two significant main effects involving I and K . Upon inspection of the cell means, it appears that the simulated annealing methodology outperforms the Späth procedure here when $I = 35$ ($X_1 = 1$) (mean = 0.04) and when $K = 2$ ($X_3 = 1$, mean = 0.04), where the reverse is true for $I = 100$ ($X_1 = 3$, mean = -0.02) and $K = 4$ ($X_3 = 3$, mean = -0.02), although these differences are indeed small.

Table 11 presents the ANOVA results on the RMS (B, \hat{B}) difference. Here, there is one significant main effect, two-way interaction, three-way interaction, and four-way

TABLE 5

ANOVA Results on \tilde{B} Recovery

<u>Source</u>	<u>SS</u>	<u>DF</u>	<u>MS</u>	<u>F</u>
X ₁	12.498	1	12.498	0.96
X ₂	37.603	1	37.603	2.88
X ₃	53.309	1	53.309	4.08*
X ₄	4.937	1	4.937	0.38
X ₅	16.843	1	16.843	1.29
X ₆	8.402	1	8.402	0.64
X ₇	42.854	2	21.427	1.64
X ₈	28.469	2	14.234	1.09
X ₉	21.880	2	10.940	0.84
X ₁₀	33.320	2	16.660	1.28
X ₁₁	33.948	2	16.974	1.30
X ₁₂	21.829	1	21.829	1.66
Explained:	317.04	17	18.65	1.428
Unexplained:	117.53	9	13.06	
Total:	424.57	26		

r-square = 0.731

* $p \leq .10$
 ** $p \leq .05$
 *** $p \leq .01$

interaction, all of which involve K (X_3). Upon inspection of the associated cell means, the simulated annealing procedure outperforms the Späth methodology when:

1. $I = 35$ ($X_1 = 1$), $J = 6$ ($X_2 = 3$), $K = 3$ ($X_3 = 2$), and no error ($X_4 = 1$), mean = -2.25);
2. $I = 65$ ($X_1 = 2$), $J = 6$ ($X_2 = 3$), $K = 4$ ($X_3 = 2$), and the low error condition ($X_4 = 2$), mean = -2.19);
3. $I = 100$ ($X_1 = 3$), $J = 6$ ($X_2 = 3$), $K = 4$ ($X_3 = 3$), and the high error condition ($X_4 = 3$, mean = -2.80).

The Späth procedure significantly outperforms the simulated annealing methodology when:

1. $I = 35$ ($X_1 = 1$), $J = 6$ ($X_2 = 3$), $K = 4$ ($X_3 = 3$), and no error ($X_4 = 1$, mean = 3.65);
2. $I = 35$ ($X_1 = 1$), $J = 4$ or 6 ($X_2 = 2, 3$), $K = 4$ ($X_3 = 4$), and the high error ($X_4 = 3$) condition (means = 2.31 and 3.05).

TABLE 6

ANOVA Results on CPU Time

<u>Source</u>	<u>SS</u>	<u>DF</u>	<u>MS</u>	<u>F</u>
X ₁	1987.203	1	1987.203	1.65
X ₂	3183.155	1	3183.155	2.64
X ₃	15128.179	1	15128.179	12.53***
X ₄	1372.749	1	1372.749	1.14
X ₅	10415.667	1	10415.667	8.63***
X ₆	25.000	1	25.000	0.02
X ₇	1663.531	2	831.765	0.69
X ₈	9427.208	2	4713.604	3.91*
X ₉	11479.952	2	5739.976	4.75*
X ₁₀	4015.263	2	2007.632	1.66
X ₁₁	2077.243	2	1038.622	0.86
X ₁₂	653.941	1	653.941	0.54
Explained:	63094.700	17	3711.45	3.074**
Unexplained:	10866.223	9	1207.36	
Total:	73960.923	26		

r-square = 0.853

* p ≤ .10
 ** p ≤ .05
 *** p ≤ .01

Finally, Table 12 presents the ANOVA results on the CPU difference measure. Note, how all but two interactions are significant. Upon inspecting the complete set of cross classified cell means, in no cell did the simulated annealing procedure run quicker than the Späth methodology, although the difference was smaller for smaller data sets.

In conclusion, while there are some scenarios which appear to favor the use of one of the two procedures, there do not appear to be systematic differences found with respect to the first three performance measures. The real, clear distinction occurs with respect to computational efficiency where the Späth procedure clearly outperforms this simulated annealing approach. However, as previously discussed, the simulated annealing methodology can accommodate more general clusterwise linear regression formulations that the present Späth (1987) methodology cannot handle involving:

1. multiple dependent variables,
2. constraints on cluster membership,
3. replications by subject (as in conjoint analyses),
4. overlapping and nonoverlapping clusters.

The next section presents an application involving many of these commonplace specifications.

TABLE 7

Second Monte Carlo Design Factors

		<u>Independent Factors</u>		<u>Code</u>
X ₁	No. of Subjects	I = 35		X ₁ = 1
		I = 65		X ₂ = 2
		I = 100		X ₃ = 3
X ₂	No. of Independent Variables	J = 2		X ₂ = 1
		J = 4		X ₂ = 2
		J = 6		X ₂ = 3
X ₃	No. of Clusters	K = 2		X ₃ = 1
		K = 3		X ₃ = 2
		K = 4		X ₃ = 3
X ₄	Error	None		X ₄ = 1
		N(0, .3v)		X ₄ = 2
		N(0, .6v)		X ₄ = 3

5. Application

Oliver and DeSarbo (1988) and DeSarbo, Oliver, and De Soete (1986) have recently performed consumer psychology-related experimental studies to quantify the impact of five hypothesized determinants of consumer satisfaction:

1. *Expectations*—prepurchase beliefs about how the product/service will perform;
2. *Performance*—how the purchased product/service actually performs;
3. *Disconfirmation*—whether the purchased product/service performed better than expected (positive disconfirmation), the same as expected (zero disconfirmation), or worse than expected (negative disconfirmation; see Oliver, 1980);

TABLE 8

Univariate Results of Second Monte Carlo Analysis

<u>Dependent Measure:</u>	<u>Mean</u>		<u>S.D.</u>	
	<u>Sim.An.</u>	<u>Späth</u>	<u>Sim.An.</u>	<u>Späth</u>
1. RMS (\bar{Y} , \hat{Y})	1.327	1.301	1.200	1.097
2. Match (\hat{A} , \hat{A})	0.798	0.766	0.147	0.141
3. RMS (\hat{B} , \hat{B})	0.882	0.863	1.465	1.209
4. CPU time	1.048	0.053	1.163	0.039

TABLE 9

ANOVA Results on RMS (\bar{Y} , \hat{Y}) Difference

Source of Variation	Sum of Squares	DF	Mean Square	F
Main Effects	2.358	8	.295	1.280
X_1	.899	2	.449	1.951
X_2	.818	2	.409	1.776
X_3	.308	2	.154	.669
X_4	.334	2	.167	.725
2-way Interactions	12.792	24	.533	2.315***
X_1 X_2	2.558	4	.639	2.778**
X_1 X_3	1.906	4	.476	2.069*
X_1 X_4	1.469	4	.367	1.595
X_2 X_3	.648	4	.162	.704
X_2 X_4	3.047	4	.762	3.309**
X_3 X_4	3.164	4	.791	3.436**
3-way Interactions	17.106	32	.535	2.322***
X_1 X_2 X_3	2.693	8	.337	1.462
X_1 X_2 X_4	7.614	8	.952	4.134***
X_1 X_3 X_4	2.238	8	.280	1.215
X_2 X_3 X_4	4.561	8	.570	2.476**
4-way Interactions	2.991	16	.187	.812
X_1 X_2 X_3 X_4	2.991	16	.187	.812
Explained	35.247	80	.441	1.914***
Residual	18.648	81	.230	
Total	53.895	161	.335	

* $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$

4. *Attribution*—whether the outcome of the purchase, construed as either a success or failure, is attributed to oneself or to some external agent;

5. *Inequity*—how one party's outcomes in an exchange compare to those received by the other party.

In the former study, performance was varied while, in the latter, it was held constant. We will examine an unpublished third study where performance was varied as an independent factor and disconfirmation was tested at all three levels. In the first study, disconfirmation was examined at only two levels (positive vs. negative).

In these studies, the authors show that higher levels of consumer satisfaction are typically associated with high expectations, high performance, positive disconfirmation (product performs better than expected), internal (self) attribution, and "favorable inequity" (lower outcome ratio for the other party). In order to quantify these effects,

TABLE 10

ANOVA Results on RMS (\hat{A} , \hat{A}) Difference

Source of Variation	Sum of Squares	DF	Mean Square	F
Main Effects	.246	8	.031	2.101**
X_1	.102	2	.051	3.491**
X_2	.025	2	.013	.866
X_3	.089	2	.044	3.025*
X_4	.030	2	.015	1.023
2-way Interactions	.530	24	.022	1.509*
X_1 X_2	.109	4	.027	1.855
X_1 X_3	.096	4	.024	1.631
X_1 X_4	.091	4	.023	1.547
X_2 X_3	.084	4	.021	1.431
X_2 X_4	.073	4	.018	1.251
X_3 X_4	.078	4	.020	1.338
3-way Interactions	.617	32	.019	1.317
X_1 X_2 X_3	.121	8	.015	1.036
X_1 X_2 X_4	.087	8	.011	.742
X_1 X_3 X_4	.192	8	.024	1.638
X_2 X_3 X_4	.217	8	.027	1.853*
4-way Interactions	.039	16	.002	.166
X_1 X_2 X_3 X_4	.039	16	.002	.166
Explained	1.433	80	.018	1.223
Residual	1.187	81	.015	
Total	2.620	161	.016	

* $p \leq .10$
 ** $p \leq .05$
 *** $p \leq .01$

the authors devised simulated stock market trading scenarios and successfully embodied these interrelated constructs via an experimental design. A stock market transaction was selected because it contained naturally all of the five factors listed above. These were expectations for the stock's performance, the ability to make the investment decision personally (internal attribution) or to rely on one's broker (external attribution), a performance outcome easily compared to expectations (disconfirmation), and a comparison of the investor's outcome (gain) to that of the broker (the commission).

The present example replicates the Oliver and DeSarbo (1988) and DeSarbo, Oliver and De Soete (1986) studies with modifications as noted. The basic design, shown in Table 13, consists of eight simulated stock market scenarios presented to subjects in paired comparison format using main effects only estimation. The main effect design was used to reduce the number of pairings and, hence, respondent fatigue. Further justification is provided in Oliver and DeSarbo where a full factorial interaction design

TABLE 11

ANOVA Results on RMS (\hat{B} , \hat{B}) Difference

<u>Source of Variation</u>	<u>Sum of Squares</u>	<u>DF</u>	<u>Mean Square</u>	<u>F</u>
Main Effects	11.306	8	1.413	1.393
X_1	3.764	2	1.882	1.855
X_2	.797	2	.399	.393
X_3	6.610	2	3.305	3.258**
X_4	.135	2	.068	.067
2-way Interactions	37.185	24	1.549	1.527*
X_1 X_2	4.011	4	1.003	.989
X_1 X_3	18.012	4	4.503	4.439***
X_1 X_4	3.854	4	.963	.950
X_2 X_3	6.638	4	1.660	1.636
X_2 X_4	4.364	4	1.091	1.076
X_3 X_4	.306	4	.076	.075
3-way Interactions	41.493	32	1.297	1.278
X_1 X_2 X_3	19.373	8	2.422	2.387**
X_1 X_2 X_4	4.430	8	.554	.546
X_1 X_3 X_4	8.687	8	1.086	1.070
X_2 X_3 X_4	9.003	8	1.125	1.109
4-way Interactions	31.195	16	1.950	1.922**
X_1 X_2 X_3 X_4	31.195	16	1.950	1.922**
Explained	121.179	80	1.515	1.493**
Residual	82.164	81	1.014	
Total	203.343	161	1.263	

* $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$

was tested. No significant disordinal (crossover) interactions were found suggesting additive processing of the experimental factors. The orthogonal array used for the main effects presentation is from Addelman (1962).

Details of the design scenarios are as follows. The attribution dimension was manipulated by suggesting to the subject that the decision to buy the stock was either his/her decision (internal attribution) or was that of a broker (external attribution). For the expectation treatment, the stock was predicted either to exceed the Standard and Poor's 500 index by 5% in six months ("high" expectations) or would just match the overall market in this time period ("low" expectations). Performance was manipulated by describing the stock as having risen 6% (regardless of the market's performance) in six months ("high" performance) or as having risen only half that amount ("low" performance). Disconfirmation was manipulated relative to the expectation treatment. For positive disconfirmation, the stock was described as exceeding the expectation treatment level by 5% (10% above the market for "high" expectations, 5% above for

TABLE 12

ANOVA Results on CPU Difference

<u>Source of Variation</u>	<u>Sum of Squares</u>	<u>DF</u>	<u>Mean Square</u>	<u>F</u>
Main Effects	114.735	8	14.342	88.947***
X ₁	55.478	2	27.739	172.033***
X ₂	25.226	2	12.613	78.224***
X ₃	32.486	2	16.243	100.738***
X ₄	1.545	2	.773	4.792**
2-way Interactions	48.220	24	2.009	12.461***
X ₁ X ₂	14.736	4	3.684	22.848***
X ₁ X ₃	21.277	4	5.319	32.989***
X ₁ X ₄	3.154	4	.789	4.891***
X ₂ X ₃	5.419	4	1.355	8.402***
X ₂ X ₄	3.395	4	.849	5.264***
X ₃ X ₄	.240	4	.060	.371
3-way Interactions	21.507	32	.672	4.168***
X ₁ X ₂ X ₃	8.421	8	1.053	6.528***
X ₁ X ₂ X ₄	5.350	8	.669	4.148***
X ₁ X ₃ X ₄	1.530	8	.191	1.186
X ₂ X ₃ X ₄	6.206	8	.776	4.811***
4-way Interactions	10.113	16	.632	3.920***
X ₁ X ₂ X ₃ X ₄	10.113	16	.632	3.920***
Explained	194.575	80	2.432	15.084***
Residual	13.061	81	.161	
Total	207.636	161	1.290	

* p ≤ .10

** p ≤ .05

*** p ≤ .01

“low” expectations); for negative disconfirmation, the stock was described as falling short of the expectation treatment by 5%; and for zero disconfirmation, the stock was described as matching expectations whether they were “high” or “low.” Finally, for favorable inequity, the investor’s commission structure was described so that the actual monetary outcome net of commissions was 20% above the broker’s two-way (purchase, sale) commission while, for unfavorable inequity, the broker’s commission exceeded the investor’s gain by 20%.

Business school students at a large northeastern United States university with stock market experience were recruited to participate in a study of responses to market transaction outcomes. Thirty subjects, including 20 males and 10 females, were presented with the 28 pairs of the eight scenarios in random order. After reading and comparing the two scenarios, subjects indicated which one was “most satisfying to you.”

Utilizing the procedure discussed by Carroll (1972), the resulting 30 × 8 × 8 array

TABLE 13

Consumer Satisfaction Study - The Data

<u>Design Matrix</u>					
<u>Construct:</u>					
Trial	Attribution	Expectation	Disconfirmation	Performance	Inequity
1	External	Low	None	Low	Unfavorable
2	Internal	High	None	High	Favorable
3	External	Low	Positive	High	Favorable
4	Internal	High	Positive	Low	Unfavorable
5	External	High	Negative	Low	Favorable
6	Internal	Low	Negative	High	Unfavorable
7	External	High	Positive	High	Unfavorable
8	Internal	Low	Positive	Low	Favorable

Satisfaction Dominance Scores

		<u>Trial:</u>							
		1	2	3	4	5	6	7	8
Subject:	1	-1	7	1	-7	-5	-3	5	3
	2	-3	7	1	-7	-5	-1	5	3
	3	-1	7	1	-5	-7	-3	5	3
	4	-3	3	7	-1	-3	-1	5	-7
	5	-3	7	1	5	-3	-7	-3	3
	6	-5	7	-3	3	-7	1	1	3
	7	-5	3	1	7	-5	-5	1	3
	8	-5	5	-1	1	-1	-3	1	3
	9	-1	5	1	5	-5	-5	-3	3
	10	-3	7	3	-3	-7	3	1	-1
	11	-7	5	-1	-1	-3	1	5	1
	12	-3	7	1	5	-7	1	-5	1
	13	-3	7	1	3	-1	-7	-5	5
	14	-5	1	3	1	-7	3	7	-3
	15	-3	3	-1	3	-7	1	7	-3
	16	-3	1	1	-1	-5	-1	5	3
	17	-5	3	-1	3	-7	1	-1	7
	18	-5	7	3	1	-7	1	1	-1
	19	-3	3	5	-1	-7	1	3	-1
	20	-7	3	3	7	-5	-1	-1	1
	21	-5	5	3	1	-7	3	3	-3
	22	-5	7	3	-3	3	-7	1	1
	23	-5	7	1	-1	-5	1	5	-3
	24	-5	5	5	-5	-3	1	3	-1
	25	-5	7	-3	5	-7	3	-1	1
	26	-7	3	-1	5	-5	1	1	3
	27	-5	5	1	3	-7	-1	3	1
	28	-3	5	5	-3	-3	1	5	-7
	29	-3	7	5	-3	-5	-5	3	1
	30	-3	7	3	-1	-5	-7	3	3

of paired comparison judgments was converted into a 30 × 8 matrix of dominance judgments which denote the number of times a scenario was selected over others (positive) or others were selected over it (negative). The matrix of integer dominance counts is shown at the bottom of Table 13. Treating the experimental design (converted to dummy variables) as the design matrix of independent variables and these dominance counts as the dependent variable, each having eight replications/observations per sub-

TABLE 14

Total Sample Regression Results for the Consumer Satisfaction Study

Intercept	-4.000	
X ₁	2.667***	Self-attribution
X ₂	1.400***	High expectation
X ₃	0.617	Positive disconfirmation
X ₄	-3.633***	Negative disconfirmation
X ₅	3.833***	High performance
X ₆	1.300***	Favorable Inequity
S.E.	2.860	
r-square	0.545	
adj r-square	0.533	
F	46.441***	
Z	0.455	

* $p \leq .10$
 ** $p \leq .05$
 *** $p \leq .01$

ject, a main-effect total sample regression analysis was performed and the results are shown in Table 14. Indeed, as hypothesized, higher satisfaction is related to self-attribution, high expectations, high performance, the absence of negative disconfirmation, and favorable inequity. The positive disconfirmation effect was not significantly different from zero disconfirmation. If we treat the entire sample as members of one cluster, we obtain $Z = 0.455$.

Given the recent findings of Oliver and DeSarbo (1988) which demonstrate that different subjects can use a variety of different response styles in forming their satisfaction judgments, overlapping clusters were obtained for $K = 1, 2, 3$, and 4 clusters. (Note, neither Späth's, 1985, 1987; nor the DeSarbo & Cron, 1988, procedure can be applied here given the replications from the design and the specification of overlapping clusters.) Table 15 presents the results for the $K = 3$ solution which was selected on the basis of goodness of fit and interpretation. Note that because the clusters are overlapping, one cannot simply perform a multiple regression analysis on each cluster independently since the general clusterwise-regression model in (8) is additive between overlapping clusters. Thus, in order to obtain approximate standard errors for \mathbf{B} , we approximated the Hessian matrix of second derivatives using the first derivatives

TABLE 15

Overlapping Cluster Solution for the Consumer Satisfaction Study

				Z = 0.226	r-square = 0.774		
Subject	A			B			
	I	II	III	Cluster:			
					I	II	III
1	1	1	1				
2	1	1	1				
3	1	1	1				
4	0	1	0				
5	1	0	1	Intercept	-4.605***	-0.272	1.778***
6	1	0	0				
7	1	0	0	X ₁	6.025***	-3.442***	-2.289***
8	1	0	1				
9	1	0	1	X ₂	1.302***	-0.014	0.161
10	1	1	0				
11	1	1	0	X ₃	1.836***	-0.522	-2.186***
12	1	0	0				
13	1	0	1	X ₄	-2.364***	-0.381	-2.939***
14	1	1	0				
15	1	1	0	X ₅	1.080***	5.164***	-1.639***
16	1	1	0				
17	1	0	0	X ₆	0.148	-0.452	3.867***
18	1	1	0				
19	1	1	0				
20	1	0	0				
21	1	1	0				
22	1	1	1				
23	1	1	0				
24	1	1	1				
25	1	0	0				
26	1	0	0				
27	1	1	0				
28	0	1	0				
29	1	1	1				
30	1	1	1				

* $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$

(Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985, p. 180) and calculated the asymptotic standard errors for **B** from the inverse of this approximated Hessian matrix.

The first cluster depicts a group of investors mostly forming their satisfaction judgments on the basis of self-attribution, high expectations, positive disconfirmation, an absence of negative disconfirmation, and favorable inequity. Note that this group (consisting of 28 of the 30 subjects) most resembles the response pattern of the total sample as reported in Table 14 earlier. It is the response pattern that would be predicted from available consumer satisfaction theory (Oliver, 1980) and from the findings reported in Oliver and DeSarbo (1988).

The second cluster, comprised of 19 subjects, has a quite different profile. Here, only attribution and performance are important constructs in their satisfaction judgments. In particular, *external* attribution to the broker and high performance lead to higher satisfaction derived from these stock market transactions. Note that the performance coefficient is highly significant showing that this group is "bottom-line" oriented

and concerned with how their simulated stock actually performed. The highly significant external attribution finding strongly suggests that these subjects were more satisfied to the extent that their broker made the decision for them, perhaps because they demand that the broker "earn" the sales commission. This contrasts with cluster one where the opposite attribution tendency was displayed.

Finally, the third cluster of eleven subjects displays a dominant inequity effect reflective of individuals who see the client-broker relationship as one of proportionate exchange. Because the broker's funds are not at risk whereas the investor's are, these subjects may feel more satisfied when their gain exceeds that of the broker's fee. This is consistent with other research showing that satisfaction in sales encounters increases as the buyer's outcomes relative to those of the salesperson increase (see Oliver & Swan, in press). This is also consistent with the negative coefficient for attribution which indicates a desire for the broker to make the trade and with the Oliver and Swan results which show that satisfaction is associated with perceptions of high seller inputs to the transaction. The third cluster is apparently more sensitive to these interpersonal equity influences than are the other clusters.

Of additional interest here are the disconfirmation and performance findings for this third group. While these members prefer to avoid negative disconfirmation as do the other subjects, they appear to tolerate well the absence of positive disconfirmation and to accept simple confirmation of their expectancies. This finding was facilitated by the fact that both performance conditions in the study were gains. Thus, the confirmation (no disconfirmation) treatment level still provided a gain to the subjects. Supportive findings are also reported in the Oliver and Swan (in press) study where equity concerns were weighted more heavily than was disconfirmation. The negative performance condition is likewise understandable in that it suggests that this group appears to accept a smaller gain. Although the study design was orthogonal, the subjects may have consciously preferred the smaller of the two gains so as not to overbalance the reward ratio inordinately in their favor. In effect, they may have been attempting to strike the correct balance between what they received and what the broker received. Possibly the smaller gain appeared more "equitable" vis-a-vis the broker's commission.

The multiple membership cluster findings are suggestive of multiple response styles on the part of those subjects exhibiting membership in more than one cluster. Whereas 28 of the 30 subjects expressed sensitivity to five of the six experimental factors, two-cluster memberships were apparent. Of these two-cluster subjects, ten were more oriented to external attribution and performance influences, and four to equity concerns than the sample as a whole. Additionally, seven subjects from the 28 share memberships in all clusters, displaying greater influence from external attribution, performance, and equity. Apparently, these seven exhibit the common profile, but are more oriented than most to these satisfaction determinants. Finally, two subjects appear to demonstrate sensitivity to external attribution and performance exclusively.

The stability of this solution is shown in Table 16 which presents the R^2 values (i.e., goodness of fit statistics) for some fifty computer analyses performed for $K = 3$ with overlapping clusters for this application, each utilizing a different random starting initial solution. This was performed in order to examine the issue of locally optimal solutions and their degree of severity with real data. As shown, $R^2 = 0.774$ solution is recovered in 31 out of the 50 trials. The mean R^2 for these 50 analyses is 0.770 with a standard deviation of only 0.008. Thus, while this simulated annealing based methodology can only guarantee locally optimal solutions, the overall performance on repeated analyses for this data set is indeed encouraging. However, as with Späth's (1985, 1987) methodology, the user is encouraged to obtain multiple solutions with such applications.

TABLE 16

R-squares For Fifty Three-Cluster Solutions

<u>Run</u>	<u>R²</u>
1	0.774
2	0.774
3	0.771
4	0.771
5	0.765
6	0.774
7	0.774
8	0.774
9	0.774
10	0.771
11	0.771
12	0.774
13	0.774
14	0.774
15	0.774
16	0.774
17	0.765
18	0.774
19	0.774
20	0.774
21	0.765
22	0.771
23	0.771
24	0.765
25	0.774
26	0.765
27	0.774
28	0.774
29	0.771
30	0.732
31	0.774
32	0.774
33	0.771
34	0.771
35	0.774
36	0.774
37	0.774
38	0.774
39	0.732
40	0.774
41	0.774
42	0.774
43	0.774
44	0.774
45	0.771
46	0.765
47	0.774
48	0.765
49	0.774
50	0.774

Mean: .770

Standard Deviation: .008

6. Discussion

We have presented an alternative methodology for clusterwise linear regression, extending the work of Späth (1985, 1987) and DeSarbo and Cron (1988) to accommodate replications, multiple dependent variables, constraints, and overlapping clusters. Monte Carlo analyses have been provided which demonstrate the somewhat consistent fitting of the simulated annealing procedure for various size data sets and different model specifications. In addition, it was found that the Späth (1987) procedure and annealing-based methodology perform approximately equally well in terms of data and

parameter recovery for the special case of $L = R = 1$, no constraints, and nonoverlapping clusters, although Späth's (1987) clusterwise linear regression procedure is computationally more efficient. An application in consumer psychology was provided investigating the various response styles present in the determination of consumer satisfaction judgments. A number of various research avenues exist as a result of this work.

Potential Applications

The clusterwise regression methodology presented here can be utilized in a large number of different applications. For example, in time series analysis, this procedure could be used to estimate splines for a given number of knots or connecting points. Constraints would have to be implemented to guarantee the same cluster membership for a continuing number of observations. In consumer psychology, this clusterwise linear regression methodology could aid in market segmentation studies where different groups of consumers or market segments could be defined directly in terms of demand characteristics. Another useful application here is the grouping of sales territories according to their response to marketing effort. Identifying territories with different response behavior can help management in the assignment of territories to salespeople as well as in the evaluation of sales performance. In psychological testing, this procedure can be gainfully utilized in formulating groups of subjects with similar personality profiles. Alternatively, in clinical practice, behavior disorders such as schizophrenia or others involving personality dysfunction may be more readily diagnosed. In the financial markets, investors with different risk preferences may be identified based on their response to market movements and economic news. And, in education, students with differences in response to learning stimuli may be more easily identified. Clearly, other potential applications for such a methodology exist.

Future Research

Further research is required in a number of areas related to this research effort. Areas involving the generalization of the methodology to nonlinear regression, L_p norm based objective functions (e.g., Späth, 1986a, 1986b; and Meier, 1987 has generalized his procedure to clusterwise linear least absolute deviations regression), fuzzy clusterings, etc., are potential avenues of further research. Further, more extensive Monte Carlo analyses are also required to investigate more fully the performance of the algorithm as a number of data, error, user options, and algorithm parameters are altered. Finally, work in the area of developing faster and more efficient combinatorial algorithms is required given the potential local optimum problems associated with both annealing and Späth's (1985, 1987) clusterwise regression procedure.

References

- Aarts, E., & Korst, J. (1989). *Simulated annealing and Boltzmann machines*. New York: Wiley.
- Addelman, S. (1962). Orthogonal main effects plans for asymmetrical factorial experiments. *Technometrics*, 4, 21-46.
- Bohachevsky, I. O., Johnson, M. E., & Stein, M. L. (1986). Generalized simulated annealing for function optimization. *Technometrics*, 28, 209-217.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. I, pp. 105-155). New York: Seminar Press.
- Davis, L. (1987). *Genetic algorithms and simulated annealing*. London: Pitman.
- DeSarbo, W. S. (1982). GENCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, 47, 446-469.

- DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via weighted least squares. *Psychometrika*, *50*, 275–300.
- DeSarbo, W. S., Carroll, J. D., Clark, L. A., & Green, P. E. (1984). Synthesized clustering: A method for amalgamating alternative clustering bases with differential weighting of variables. *Psychometrika*, *49*, 57–78.
- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*, 249–282.
- DeSarbo, W. S., & Mahajan, V. (1984). Constrained classification: The use of a priori information in cluster analysis. *Psychometrika*, *49*, 187–215.
- DeSarbo, W. S., Oliver, R. L., & De Soete, G. (1986). A probabilistic multidimensional scaling vector model. *Applied Psychological Measurement*, *10*, 79–98.
- De Soete, G., DeSarbo, W. S., & Carroll, J. D. (1985). Optimal variable weighting for hierarchical clustering: An alternating least squares algorithm. *Journal of Classification*, *2/3*, 173–192.
- De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, J. D. (1984). The presentation of nonsymmetric rectangular proximity data by ultrametric and path length tree structures. *Psychometrika*, *49*, 289–310.
- De Soete, G., Hubert, L., & Arabie, P. (1988a). On the use of simulated annealing for combinatorial data analysis. In W. Gaul & M. Schader (Eds.), *Data, expert knowledge, and decisions* (pp. 328–340). Berlin: Springer Verlag.
- De Soete, G., Hubert, L., & Arabie, P. (1988b). The comparative performance of simulated annealing on two problems of combinatorial data analysis. In E. Diday (Ed.), *Data analysis and informatics*, *V* (pp. 489–496). Amsterdam: North-Holland.
- Dubes, R. C., & Klein, R. (1987, June). *Simulated annealing in data analysis*. Handout at a talk given at the 1987 Annual Meeting of the Psychometric Society, Montreal, Canada.
- Gabor, A., & Granger, D. W. (1966). On the price consciousness of consumers. *Applied Statistics*, *10*, 170–181.
- Gidas, B. (1985). Nonstationary Markov chains and convergence of the annealing algorithm. *Journal of Statistical Physics*, *39*, 73–131.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading: Addison-Wesley.
- Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, *8*, 355–363.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5*, 103–123.
- Haggerty, M. R. (1985). Improving the predictive power of conjoint analysis: The use of factor analysis and cluster analysis. *Journal of Marketing Research*, *22*, 168–184.
- Henderson, J. M., & Quandt, R. E. (1982). *Microeconomic theory: A mathematical approach* (3rd ed.). New York: McGraw Hill.
- Johnson, M. E. (1988). *Simulated annealing and optimization*. Syracuse: American Science Press.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lükepohl, H., & Lee, T. (1985). *The theory and practice of econometrics*. New York: Wiley.
- Kamakura, A. W. (1988). A least-squares procedure for benefit segmentation with conjoint experiments. *Journal of Marketing Research*, *25*, 157–167.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*, 671–680.
- Levy, A. V., & Montalvo, A. (1985). The tunneling algorithm for the global minimization of functions. *SIAM Journal of Scientific and Statistical Computing*, *6*, 15–29.
- Lin, S., & Kernighan, B. (1973). An effective heuristic algorithm for the traveling salesman problem. *Operations Research*, *21*, 498–516.
- Locke, E. A. (1967). Relationship of goal level to performance level. *Psychological Reports*, *20*, 1068.
- Locke, E. A. (1982). Relation of goal level to performance with a short work period and multiple goal levels. *Journal of Applied Psychology*, *67*, 512–514.
- Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, *90*, 125–152.
- Lundy, M. (1986). Applications of the annealing algorithm to combinatorial problems in statistics. *Biometrika*, *72*, 191–198.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *5th Berkeley Symposium on Mathematics, Statistics and Probability*, *1*, 281–298.
- Maddala, G. S. (1976). *Econometrics*. New York: McGraw-Hill.
- Meier, J. (1987). A fast algorithm for clusterwise linear absolute deviations regression. *OR Spektrum*, *9*, 187–189.

- Mitra, D., Romeo, F., & Sangiovanni-Vincentelli, A. (in press). Convergence and finite-time behavior of simulated annealing. *Advances in Applied Probability*.
- Ogawa, K. (1987). An approach to simultaneous estimation and segmentation in conjoint analysis. *Marketing Science*, 6, 66–81.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17, 460–469.
- Oliver, R. L., & DeSarbo, W. S. (1988). Response determinants in satisfaction judgments. *Journal of Consumer Research*, 14, 495–507.
- Oliver, R. L., & Swan, J. E. (in press). Consumer perceptions of interpersonal equity and satisfaction in transactions: A field survey approach. *Journal of Marketing*.
- Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67, 306–310.
- Sowter, A. P., Gabor, A., & Granger, G. W. (1971). The effect of price on choice. *Applied Economics*, 3, 167–181.
- Späth, H. (1979). Algorithm 39: Clusterwise linear regression. *Computing*, 22, 367–373.
- Späth, H. (1981). Correction to Algorithm 39: Clusterwise linear regression. *Computing*, 26, 275.
- Späth, H. (1982). Algorithm 48: A fast algorithm for clusterwise linear regression. *Computing*, 29, 175–181.
- Späth, H. (1985). *Cluster dissection and analysis*. New York: Wiley.
- Späth, H. (1986a). Clusterwise linear least absolute deviations regression. *Computing*, 37, 371–378.
- Späth, H. (1986b). Clusterwise linear least squares versus least absolute deviations regression: A numerical comparison for a case study. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 413–422), New York: North Holland.
- Späth, H. (1987). *Mathematische software zur linearen regression* [Mathematical software for linear regression]. Munich: R. Oldenbourg.
- van Laarhoven, P. J. M., & Aarts, E. H. L. (1987). *Simulated annealing: Theory and applications*. Boston: D. Reidel.

Manuscript received 7/8/87

Final version received 9/12/88