MARKETING SCIENCE
Vol. 6, No. 3, Summer 1987
Printed in U.S.A

A TWO-STAGE DISAGGREGATE ATTRIBUTE CHOICE MODEL

DENNIS H. GENSCH
University of Wisconsin—Milwaukee

A two-stage disaggregate attribute choice model is proposed and empirically implemented. The first stage of the model is attribute processing to screen the number of alternatives down to a lesser number. The second stage is brand (alternative) processing which considers the attributes simultaneously while allowing for tradeoffs among the attributes.

This two-stage approach is then applied to the same real world data set as two single stage disaggregate models, logit and Maximum-Likelihood-Hierarchical (MLH) which are state of the art models representing the alternative and attribute processing approaches, respectively. The predictive accuracy of the two-stage approach compares favorably to the single stage models. In addition, it seems to offer diagnostic information that can provide managerial insights not found in the output of the single stage model.

(Choice Modeling; Hierarchical; Brand Choice; Multistage Decisions)

Disaggregate attribute choice models have been developed with a wide variety of logic structures, assumptions, and purposes. The data requirements for the various approaches vary considerably in terms of the information load placed upon respondents. Conjoint analysis requires an experimental design in obtaining respondent judgments and the difficulties in obtaining sufficient input per respondent in order to run these powerful models are well known. See Green and Srinivasan (1978). Models of the multidimensional scaling family (such as LINMAP and PREF-MAP) which use paired comparisons to map a preference space also have the practical problem that the number of paired comparisons increases geometrically rather than linearly as the size of the decision set increases. A third class of disaggregate models uses survey data, does not require an experimental design, and has perhaps the lowest information load of the three classes in situations where there are a large number of alternatives, each of which is to be evaluated on numerous attributes, and attribute values are assumed to be intervally scaled. Respondents are simply asked to rate the attributes per alternative and also, in some models, to rank the attributes in order their importance.

This article will focus only on the disaggregate models using the latter type of survey data input. There is no intention to imply that models using this third class of respondent input are superior to models in the other classes. Rather, this limitation is simply a recognition of the fact that choice models using the various classes of input are quite different. Once a certain class of data has been collected, models from the other classes are either inoperable or severely restricted in their application.

0732-2399/87/0603/0223$01.25

There is a basic dichotomy in categorizing the various models that use this third class of simple survey data. Models that postulate that the decision makers make choices by comparing brands (alternatives) on an attribute by attribute basis, generally assume there is a hierarchical sequence in which the attributes are considered. Elimination-By-Aspects, HIARC, preference tree, and Maximum-Likelihood-Hierarchical are models of this type. The second basic process assumes all of the attributes are considered in a simultaneous compensatory structure, thus assigning a total value to each alternative. After each alternative is evaluated and assigned a total value, the alternatives are compared and the alternative with the highest total value is selected. Analytic techniques using this structure to model a choice situation include regression, logit, and probit. Current applications of these attribute choice models are all single stage model applications. Normally all members of the (sample) population are assumed to process information either by attribute *or* by alternative, depending on the particular disaggregate attribute choice model being used. Recent articles provide empirical evidence that segmenting a population (sample) *a priori* based on such variables as degree of respondent's knowledge (Gensch 1987) or degree of involvement (Gensch and Javalgi 1987a) and then analyzing each segment with a *different* decision model significantly improves the prediction rate and perhaps the researcher's understanding of the underlying process over the standard practice of analyzing a population (sample) with one model. This work tends to support the simple assumption that not all members of a population (sample) are using the same decision process on a given choice situation.

This article attempts to give a more parsimonious explanation to differences in the decision processes respondents may be using at the particular point in time they are interviewed regarding a particular choice situation. It follows previous suggestions made in the literature that an individual may go through more than one stage in the decision process and that the decision rules may vary from one stage to another. Thus individuals may be using different decision processes because they are at different stages in a multistage process at the time of the interview. A two-stage disaggregate choice model is proposed in which individuals are assumed to use an attribute processing approach as they reduce the set of all feasible alternatives down to a final choice set of just a few alternatives. In the second stage of the choice process the few final alternatives are closely compared to each other and one alternative is selected. It is assumed that in this stage all the attributes are considered simultaneously and tradeoffs among the attributes are made. In this second stage, the choice process is by alternative rather than by attribute.

In the next section, support for this two-stage approach in the marketing literature is reviewed. Following this section, a specific two-stage model is postulated. The predictive and diagnostic results of the two-stage model are then compared and contrasted to single-stage models on the same real world data set. The significant improvement in predictive accuracy and change in diagnostic information lead to conclusions and implications of this research that will be discussed in the final section of this paper.

Literature Review of Two-Stage Choice Process

There are a number of good review articles on disaggregate choice models that explain the algorithms, look at the empirical applications, and discuss the assumptions underlying the particular choice models. See Forrester (1979) for a general overview, Corstjens and Gautschi (1983) for utility based models, Malhotra (1984) for marketing applications of logit, and Gensch and Javalgi (1987b) for hierarchical models and discussion of appropriateness of various models or combinations of models for various choice situations. Given these published reviews, this article will not duplicate this material by also reviewing each of the disaggregate choice models. It is sufficient to note that each of the models reviewed above is a single-stage model that compares attribute values in a hierarchical sequence across alternatives or in a simultaneous compensatory manner by alternative.

There is some evidence in the literature on individual level choice models to suggest that decision makers use different decision rules at various stages of the decision process in a manner that suggests the two-stage approach described in this article. There is no empirical evidence at the disaggregate or group level on this topic. The scanty empirical evidence that does exist comes primarily from behavioral studies using protocol analysis in which the individual decision maker verbalized his/her choice process.

Pras and Summers (1975) found that a conjunctive model performed better when all the alternatives were included and a linear rule was better when only acceptable alternatives were included in the analysis. Payne (1976) analyzed the protocols from six subjects and found that subjects tended to use conjunctive and elimination-by-aspects strategies early in the decision process to simplify the decision task by eliminating alternatives until only a few alternatives remained as choice possibilities. For the remaining choice alternatives, subjects used the additive difference model to make the final evaluation and choice. Einhorn (1970) suggested a similar explanation of decision making based upon attempts to fit various choice model structures statistically to actual choice data. Wright and Barbour (1977) used protocol analysis and found that decision makers used forms of attribute processing strategies to eliminate a number of alternatives and then used different decision rules to select one choice from their final set. The decision rules used in making the final selection appeared to be of a more compensatory nature than those used in the initial elimination stages. Lussier and Olshavsky (1979) report that subjects tended to eliminate unacceptable alternatives by using a conjunctive strategy and they then evaluated the remaining acceptable alternatives using a compensatory strategy. Lehtinen (1974) used questionnaire data on automobile choices and reports the finding of an initial screening process followed by a more compensatory analysis of the final choice set. Sheridan et al. (1975) using questionnaire data found evidence of an initial screening of job alternatives followed by a more detailed choice phase.

While the subjectivity and weaknesses associated with the protocol approach must be acknowledged (see Haines 1974), it is interesting that all of these researchers indicate a two-stage pattern in the choice process. The initial screening takes place using a hierarchical attribute screening process to reduce the feasible set of alternatives down to a final choice set and the final choice set uses a more compensatory decision rule in which alternatives are fully evaluated and then compared. This is the same approach postulated in artificial intelligence by Newell and Simon (1972). In marketing Bettman (1979, pp. 179–185) postulates this same concept of a two-stage process for consumers. Individuals will screen down or simplify the amount of information by eliminating alternatives until they have reduced the information load sufficiently to deal comprehensively with the final set of information.

Malhotra (1986) generated 25 alternative profiles and had the respondents rate these profiles, indicating which alternatives they were unwilling to pursue, and then analyzed the responses using tobit analysis and metric conjoint. Since the predictive results and estimated parameters of the two approaches were similar, Malhotra concludes that the data required from respondents could be significantly reduced, if in some way the unacceptable alternatives for each respondent could be identified. While one is very hesitant to generalize from these results to situations other than a controlled experimental setting, Malhotra's finding that it is important to identify the unacceptable alternatives tends to support the logic of the approach proposed in this article.

Thus, while there has been considerable theoretical hypothesizing about the existence of a two-stage choice approach, there has been little empirical work related to this concept. The little empirical evidence that does exist is rather subjectively interpreted results about a few individuals. Empirical evidence at the disaggregate or group level that either supports or rejects the concept of a two-stage choice process is currently nonexistent. This article will present empirical evidence at the disaggregate level which tends to support the hypothesis of a two-stage decision process.

Data Set

Various disaggregate choice models are more or less appropriate for the particular choice situation being considered. It should be noted that the author would have preferred to illustrate this two-stage approach on a data set in which respondents were knowledgeable enough to rate five or more attributes on a large number of alternatives. Unfortunately, such a data set was not available to this author so an industrial data set involving a choice among four suppliers is used. With a choice set of 15 to 20 alternatives, it is my belief that the improvement in prediction rates over single-stage disaggregate models would be even more substantial than those reported in this article.

So that readers may evaluate the appropriateness of the two-stage model proposed in this article, the choice situation and available attribute data will be described prior to specifying the model. The data set to be analyzed is derived from an industrial marketing survey conducted by a nationally known market research firm for a major supplier of the electrical equipment (transformers, generators, switchgear, etc.) purchased by electrical utility companies. Respondents are key decision makers at the various utilities. The purpose of the survey is to identify the product features and supplier services most salient in the choice of suppliers. Following a pretest ($n = 98$) in which 21 product and service attributes related to the purchase of a specific type of electrical equipment were factor analyzed and a reduced set of 9 relatively independent attributes (which explained over 79% of the variation in the original 21 space) was obtained, the actual questionnaire contained 9 attributes that were ranked and rated by the 182 respondents in this survey. Previous analysis on this data set (Gensch 1984) indicated that one of these attributes, appearance of product, was generally insignificant in the choice decisions and therefore was dropped prior to this analysis (see Table 1 for list of 8 retained attributes). The respondent provided his current perception using a poor to good scale on each of the attributes for each supplier known to him. Line continuums provided a linear scale from 1 to 50 on each of the above respondents' judgments. Respondents indicated how familiar they were with each supplier they rated and which supplier was most likely to obtain their next order for a specific type of electrical equipment. The response to the mail questionnaire was over 40%. A follow-up phone check of nonrespondents failed to detect any significant nonresponse bias.

Previous research on this data set revealed distinct segments with very different value structures (i.e. the relative importance weights on the 8 attributes were relatively homogeneous within a segment but often quite different between segments). Those readers interested in the details of the segmentation analysis and/or the factor analysis are referred to Gensch (1984). There were 182 industrial buyers who rated the same four major suppliers on the 8 selected attributes and were in segments judged to have similar value structures. All observations were retained in this data set and the 182 industrial buyers were randomly split into two sets of 91 each.

Randomly splitting the sample into two groups allows for a double crossover research design in which the choice models will be calibrated first on group one individuals while group two individuals become the hold-out sample used to estimate predictive accuracy. The process is then reversed by making group two the calibration sample and group one the hold-out sample.

Two-Stage Disaggregate Choice Model

The first stage in this disaggregate choice process will be to apply the maximum-likelihood-hierarch model (MLH) developed by Gensch and Svestka (1984), to reduce the number of alternatives in the feasible set down to a final choice set. The second stage in this model will select a single chosen alternative from the final choice set using a logit approach.

A number of attribute processing approaches could be used in the first stage of a multistage model. Conjunctive, disjunctive, EBA, HIARC, preference tree, nested logit, generalized extreme value models, and MLH are candidates. The MLH model was selected as a first-stage model over the other disaggregate attribute processing models mentioned above because it appears to have theoretical and/or computational advantages over the other models for the particular data set being analyzed. (See Appendix 2 for a brief overview of the MLH approach.)

The data set for this project clearly is more suited for the MLH model than for the other noncompensatory models. The conjunctive and disjunctive approaches were not used in the first stage because it is not clear how the critical values on each attribute would be estimated. Einhorn (1970) indicates that statistical fits to what respondents said were their critical values were very poor. He concluded individuals could not indicate accurately the critical values they used when they claimed to be using a conjunctive choice process. Even if one could obtain a reasonable estimate of an individual's critical values, how would this be aggregated? It seems unlikely that all members of a population would have the same set of critical values.

A number of considerations weighed against using the elimination-by-aspects and preference tree approaches as the first-stage model. First, the attribute data on the four suppliers was intervally scaled, not binary. Second, there was no homogeneous hierarchical sequence in which the attributes would be considered (i.e., some decision makers would consider one attribute as most important; some another). This lack of a clear structure of conditional decisions also made the generalized extreme value approach as operationalized by a nesting of logit models[1] inappropriate for this industrial choice problem.

The HIARC model (Gensch and Svestka 1979) deals with the problems of intervally scaled data, doesn't require the same sequence of attribute considerations for all members of the population, and doesn't require the assumption of universal cut-points for all members of the population. Conceptually this is appealing, but the HIARC model is deterministic. The MLH is a probabilistic version of HIARC. As such MLH is designed to deal with intervally scaled data on the attribute rating on each alternative, allows for different sequences of attributes in the screening process employed by different individuals, computes aggregate cut-points from the sample data and applies the cut-points at the individual level, thus allowing different individuals to reject alternatives on different value of the same attribute. This probabilistic model sets chance constraints on the aggregate tolerances so that in a population approximately $1 - \alpha$ will pass the estimated aggregate tolerance with their chosen alternatives. In my initial formulation of the two-stage approach the HIARC model was used as the first-stage model, but the predictive results were significantly below those produced by using MLH as the first stage. The chance constraints make the MLH model less sample specific to the calibration sample and allow for the elimination of the influence oı outliers or atypical decision makers in setting the aggregate tolerances. This is why the MLH significantly outperformed HIARC as the first-stage period.

[1] The nested logit model may be viewed as a multistage model, but the process and decision rules remain constant for all such stages.

The second stage in the two-stage approach requires a comparison of the final set of alternatives that survive the first-stage screening phase. This phase should consider the alternatives over the range of all salient attributes allowing for trade-offs among the attributes. This more intensive evaluation of the remaining alternatives calls for a disaggregate choice structure that considers the attributes of the final choice set in a simultaneous compensatory fashion. The four alternatives with eight attributes is too large a problem to be solved directly by probit.[2] Because logit is currently the "state of the art" for this class of model and because logit requires the same attribute rating data as the MLH model, thus making the two stages data compatible, logit was selected as the second stage in the two-stage disaggregate choice model.

No claim is being made that the particular two-stage disaggregate choice model proposed in this paper is *the* two-stage model that is most appropriate for all data sets and choice situations. Since there currently does not exist a recognized multistage disaggregate choice model and it is beyond the scope of this article to formulate and test the infinity of other possible n-stage model structures, one can only present the rationale behind the particular formulation postulated in this article and indicate for this *particular* problem and data set why the particular two stage model proposed in this article was selected.

The MLH model will be used to screen the initial set of alternatives (i.e. the four suppliers in our case example). For those individuals that eliminate all but one alternative, the final choice is made in the first stage of the two-stage model. For the remaining individuals the MLH model indicates more than one alternative remaining after the hierarchical evaluation on an attribute by attribute basis has been completed. The logit model then is run for those individuals with more than one alternative remaining. The logit model is capable of dealing with the fact that different individuals may have a different number of alternatives passing the first stage.

The two stages are not independent in that the logit function estimates in the second stage is dependent on the set of alternatives passed on from the first stage. These are essentially sequential steps because the alternative set on which the B_i's are estimated are a function of the first stage (T_i) parameters, but the B_i's are not functionally related to the T_i's. Thus because the two sets of parameters must be sequentially estimated, not jointly optimized, and one set is not functionally related to the other, one cannot specify a single maximum likelihood function for the entire model. The key parameter linking the two stages is α, which is set prior to estimating the aggregate tolerances (T_i's) in the MLH. As explained above, the larger the value of α, the looser the constraint on percentage of the population having to retain their actual choice; thus generating smaller T_i values. Smaller T_i values will tend to eliminate more alternatives than will larger T_i values. Therefore the number of alternatives passed on to the second stage is determined by α. In a number of test runs on this and other data sets while attempting to determine the best value of α to use, the author has made the following observations:

1. The best value of α varies from one data set to another.
2. On a particular data set a gradient search algorithm tends to converge reasonably efficiently.

Thus, while no guarantee can be given of a global optimal solution for this two-stage model, the following heuristic is offered to obtain reasonable empirical estimation of α for a given data set.

1. Select an initial α value denoted α_0.
2. Compute $\alpha_1 = \alpha_0 + \alpha_0/2$, $\alpha_2 = \alpha_0 - \alpha_0/2$.

[2] Approximations to a probit solution exist. However, selecting a particular approximation immediately raises two important research questions that are very difficult to answer. First, how good is the approximation? Second, for this problem why is one particular approximation better than other approximations?

3. Compute predictive accuracy of two-stage model using α_0, α_1, and α_2.

4. Compare predictive results. If α_0 has maximum predictive results, go to Step 5. If α_1 or α_2 has maximum predictive results, redefine this α_i value as α_0 and return to Step 2.

5. Compute $\alpha_3 = \alpha_0 + \alpha_0/4$, $\alpha_4 = \alpha_0 - \alpha_0/4$.

6. Compute predictive accuracy of two-stage model using α_0, α_3, and α_4.

7. Compare predictive results. If α_0 has maximum predictive results, then STOP. If α_3 or α_4 has maximum predictive results, then redefine this α_i variable as α_0 and return to Step 2.

This simple heuristic used with various starting points should allow the researcher to get a reasonably good empirically derived estimate of α for a given data set.

Research Design and Empirical Results

The double crossover research design provides two randomly assigned samples of 91 each. The first step is to run the single stage models of MLH and logit on the total samples. This represents the current approach to disaggregate choice modeling where all applications are single-stage models. The MLH results represent a hierarchical non-compensatory set of predictions and diagnostic information on the same data set for which the logit results represent a simultaneous compensatory set of predictions and diagnostics. Thus, in the first step, results from models representing the two basic approaches to attribute choice modeling are presented.

In the second step predictive and diagnostic results will be computed for the two-stage approach and these results compared to the single-stage model outputs.

A logit model is calibrated on the 91 individuals in group one, and then these estimated logit coefficients are used in predicting the choices of sample two members. The process is then reversed and sample two becomes the calibration sample whose estimated coefficients are used to predict the choices of sample one members. The t-values of the logit coefficients calibrated on sample one are given in the first column of Table 1. In parentheses under each of the sample one values are the t-values of the logit coefficients calibrated on sample two.

The standard statistical test of the hypothesis that two sets of logit coefficients, $\theta^1 = \theta^2$, are from the same population is: $-2\{L(\theta^P) - [L(\theta^1) + L(\theta^2)]\}$.

Here $L(\theta^P)$ represents the log-likelihood function value for the parameters of the pooled example of the two samples being tested. This test statistic is asymptotically distributed chi-square with k degrees of freedom, where k is the number of parameters in the model. For details and suggested uses of this test see Chapman and Staelin (1982) or Gensch (1985). Applying this to total samples in Table 1 we get $-2\{(-180.22) - [(-84.67) + (-90.16)]\} = 10.78$. The critical value for 8 degrees of freedom at the 0.20 significance level is 11.03. Thus we cannot reject the null hypothesis of no difference in the *sets* of logit coefficients at the 0.20 level. In a similar fashion we cannot reject the hypothesis of no differences for the second stage samples at the 0.15 level. Thus, although the significance levels on a particular attribute differ, it is reasonable to treat the sets of logit coefficients as coming from the same population. The managerial implications of how the logit coefficients for the total sample relate to the logit coefficients of the second stage of the two-stage model are discussed later in this article after the predictive accuracy of the logit, MLH, and two-stage models are compared.

Disaggregate models such as logit and MLH give predictions at the individual level (i.e., the probability that individual i will select each of the alternatives in the choice set). By averaging these individual probabilities one can get an aggregate probability for

TABLE 1

The t-Values of Logit Coefficients for Industrial Purchasers of Electrical Equipment

Attribute	Total Samples $n = 91$ ($n = 91$)	Second Stage Samples $n = 29$ ($n = 37$)
1. Invoice Price	3.45**	1.94**
	(1.81)*	(1.02)
2. Energy Losses	3.23**	2.62**
	(1.25)	(3.06)**
3. Maintenance Requirements	0.95	0.48
	(1.89)*	(1.98)*
4. Warranty Coverage	−0.61	0.34
	(0.34)	(0.88)
5. Availability of Spare Parts	0.50	−0.78
	(0.53)	(0.12)
6. Ease of Installation	−0.48	0.68
	(0.12)	(0.31)
7. Problem Solving	2.62**	0.78
	(2.95)**	(0.96)
8. Manufacturing Quality	2.00*	0.44
	(1.50)	(0.94)
Log-Likelihood Function	−84.67	−29.06
	(−90.16)	(−25.18)

* Significant at the 0.05 Level Using a One-Tailed Test.
** Significant at the 0.01 Level Using a One-Tailed Test.

each alternative. Predictive results have been reported at both the aggregate and individual level for most disaggregate models. Individual level predictions place more demands on the model and allow us to compare the strength of prediction on the chosen alternative, Krishnan (1977); thus, individual level predictions will be emphasized.

In the industrial data set each respondent considered four alternatives. A simple summation model in which the individual's self-reported importance weight on each attribute was used to weight the attribute rating on a given alternative. This produced four sums, one for each alternative, the highest sum was treated as the predicted alternative. In 55 of the 182 cases (30%) the predicted alternative matched the actual chosen alternative. This simple summation model which produces predictive results somewhat above the chance level gives us one benchmark for considering predictive accuracy on this data set.[3] All three disaggregate approaches do significantly better than the self-explicated individual level models.

Using the criteria of highest individual probability as the predicted choice, 42 of the 91 (46%) respondents in sample one have their predicted choice match their actual choice in the single-stage logit approach. The second sample contains 45 correct predictions out of 91 (49.5%), using the single-stage logit. The sample double crossover design is used for the single-stage MLH model,[4] resulting in 42 (46%) correct predictions in sample one and 41 (45%) correct predictions in sample two.

[3] One should not dismiss the self-explicated model as a "naive" or weak model on the assumption that derived utility functions will always outperform self-explicated ones. For example, Wright and Kriewall (1980) found that "the reported functions consistently beat the derived ones."

[4] The single-stage MLH used an $\alpha = 0.0375$, which is the same α value used in the first stage of the two-stage model. This was done for comparative purposes. Slightly better prediction rates (48% and 51%) can be achieved by a single-stage MLH using an $\alpha = 0.075$ on this particular data set.

The two-stage model, using MLH as a first stage, calibrated the MLH model on one sample and then used these aggregate tolerances to predict the choices of the members of the other sample. Three starting points, α =0.25, 0.10, and 0.05, were run, using the heuristic previous described for the two-stage model. The α value producing the best predictive results for the two-stage model is α = 0.0375. In sample one, 62 of the respondents used the MLH to make a unique choice. The remaining 29 had more than one alternative remaining and these 29 moved into stage two. Similarly, on the second sample, 54 individuals reached a unique choice in the first stage and 37 moved on into the second stage. In the second stage a logit model was calibrated on the 29 respondents and these coefficients were used to predict the choices of the 37 respondents in sample two. The process was then reversed and a logit model calibrated on the 37 respondents was used to predict the responses of the 29 members of sample one. The t-values associated with each of these second-stage samples are presented in Table 1 and the implications of these coefficients will be discussed later in this article, following this presentation of predictive results.

In sample one, 40 of the 62 (65%) respondents who reached a unique choice in stage one had their actual choice correctly predicted. Logit correctly predicted the choice of 19 of the 29 (66%). Thus the prediction rate for sample one was 59 out of 91 (65%). For the second sample, 34 of the 54 were correctly predicted in the first stage and 26 of 37 in the second stage for an overall prediction rate of 66%.

One way of explaining the predictive improvement of the two-stage model over the single-stage models is that the two-stage model in effect segmented the population into those whose actual decisions are approximated quite well by a disaggregate model using a noncompensatory rule and another segment whose actual decisions required decision rules in addition to the noncompensatory rule in order to obtain good predictive approximations.

A simple McNemer test is used to show the differences in correct predictions between two models on the same sample. The details of this test and actual computations are presented in Appendix 1. The Z scores comparing the two-stage model with logit are Z = 2.42 on sample one and Z = 2.79 on sample two. Using a one-tailed test, the difference in the prediction rates between the two-stage modeling approach and the logit approach are both significant at the 0.01 level. In a similar manner the improvement in predictive accuracy of the two-stage model's approach over the maximum likelihood hierarchical approach produced Z values of 2.96 on the first sample and 3.96 on the second sample. Thus, the conclusion is that the improvement in predictive accuracy over the MLH approach is significant on both samples at the 0.01 level. It should be recognized that the two-stage model estimates twice as many parameters as do either of the single-stage models and, to some degree, the better data fit (increased predictive accuracy) attained by the two-stage model is attributed to this fact.[5,6]

[5] Because I could not estimate a single maximum likelihood function for the two-stage model, due to its interactive sequential estimation process, I could not use the difference in likelihood functions to compute the expected increase in prediction from the single stage to the two-stage model structure. I therefore generated 50 data sets of 91 individuals who rank ordered the importance of 8 attributes, rated the 8 attributes over 4 alternatives, and provided the chosen alternative, using random numbers; i.e., constructed a data base with no interrelational structure. The two-stage process provided an average improvement in prediction of 1.6 individuals (1.75% improvement) over the single logit and 2.3 (2.5% improvement) over the single MLH (all α values set at 0.05). These averages provide a rough guide to how much of the predictive improvement is strictly a function of the two-stage model estimating more parameters and making two passes at the data. The improvements in the predictive accuracy of the two-stage over the single-stage models are still very significant even when the above adjustment is added to the single-stage results.

[6] Another way to gain some insight into the extent to which each model is sample specific, i.e. additional

Krishnan (1977) suggests a procedure specifically designed to compare the predictive strength of two disaggregate choice models. Let P_1^* and P_2^* denote the predicted mode choice probabilities of the logit model, and π_1^* and π_2^* denote the predicted probabilities of the two-stage model. If A_1 was the actual choice in the sample, one would expect P_1^* and π_1^* to be close to 1 for that individual; if A_2 was the choice, then P_2^* and π_2^* would be close to 1. In either case, the model for which the predicted probability is closer to 1 is a better predictor. For each individual in the sample, the decision rule is, then, as shown in Table 2. Let

$$X_s = \begin{cases} 1 & \text{if the two-stage model is deemed superior for the } s\text{th individual,} \\ 0 & \text{otherwise,} \qquad \text{and} \qquad p = P(X_s = 1) \qquad \text{for all } s. \end{cases}$$

Since the sample observations are independent, $\{X_s\}$ represents a set of n Bernoulli trails and the sum $r = \sum_{s=1}^{n} X_s$ represents the number of times the two-stage model is deemed superior.

We now wish to test the hypothesis that the logit model and the two-stage model are equally good predictors of individuals' choices, that is, $p = 0.5$; the alternative hypothesis is that the two-stage model is a better predictor, that is, $p > 0.5$. If n is large and $p = 0.50$, the distribution of r can be approximated by a normal distribution with mean $n/2$ and variance $n/4$. The criterion for accepting or rejecting the null hypothesis is

Accept the null hypothesis of $Z \leqq Z_\alpha$,

Reject the null hypothesis of $Z > Z_\alpha$,

TABLE 2

Decision Rule for Deemed Superiority

Actual Choice	$\pi_1^* < P_1^*$	$\pi_1^* \leqq P_1^*$
A_1	Two-Stage Model is Superior	Logit is Superior
A_2	Logit is Superior	Two-Stage Model is Superior

parameters or stages take advantage of the data, is to compare the first choice predictive accuracies for the calibration sample and the predictive sample for each model.

	Sample One Parameters Used	
Model	Sample One Predictions	Sample Two Predictions
Logit	62%	49%
MLH	49%	41%
Two-Stage	78%	66%

	Sample Two Parameters Used	
Model	Sample One Predictions	Sample Two Predictions
Logit	46%	61%
MLH	42%	51%
Two-Stage	65%	80%

The decrease in predictive accuracy from calibration results to hold out (prediction sample) is about the same for logit and the two-stage model and less for MLH. I was surprised that the two-stage was not more sample specific. This was possibly due to the fact that the first-stage MLH (which was the least sample specific of the three models) accounted for most of the predictions in the two-stage; also the calibration on the two-stage may be approaching a practical limit on predictive accuracy. Thus it appears that the significant improvements in prediction of two-stage over the single-stage models are not explainable by structural advantages (more parameters and more stages).

where $Z = (r - n/2)/(n/4)^{1/2}$ and Z_α is the upper α percentile of the standard normal distribution.

Table 3 provides a comparison of the actual choices versus the predicted ones by the two-stage model versus the logit model for the two samples of 91. This is followed by a comparison of the two-stage model with the MLH results.

Using the decision rule shown in Table 2, the two-stage model is superior to logit in 63 out of 90 cases on sample one and 65 out of 91 on sample two. This produces Z values of 3.75 on sample one and 4.02 on sample two. Thus, in terms of strength of prediction on the chosen alternative, the two-stage model outpredicts the logit model at the 0.001 level of significance.

Similarly, the Z scores for the two-stage versus the MLH were $Z = 2.41$ on sample one and $Z = 3.28$ on sample two. These values were not quite as high, possibly because of the numerous ties. Yet, even with the numerous ties, the two-stage model outpredicts the MLH model at the 0.01 level of significance on the first sample and the 0.001 level of significance on the second.

Some of this increase must be directly attributed to the structural advantage the two-stage model has over the single stage models in that it estimates twice the parameters and makes two passes at the data. To a degree, these advantages make a direct comparison of predictive accuracy "unfair" to the single stage models. However, if the

TABLE 3

Comparison of Two-Stage Model to Single-Stage Logit and MLH Models

Two-Stage Model versus Logit

Sample	Sample Size	Supplier Choice	Two-Stage Model Superior	Logit Superior	Ties	Total Number of Actual Choices
Sample One	91	Supplier A	14	10		24
		Supplier B	9	8		17
		Supplier C	24	6		30
		Supplier D	16	3	1	20
			63	27	1	91
Sample Two	91	Supplier A	19	11		30
		Supplier B	11	4		15
		Supplier C	19	5		24
		Supplier D	16	6		22
			65	26		91

Two-Stage Model versus MLH

			Two-Stage Model Superior	MLH Superior	Ties	
Sample One	91	Supplier A	6	10	8	24
		Supplier B	8	5	4	17
		Supplier C	16	3	11	30
		Supplier D	12	4	4	20
			42	22	27	91
Sample Two	91	Supplier A	10	2	18	30
		Supplier B	6	2	7	15
		Supplier C	9	4	11	24
		Supplier D	11	4	7	22
			36	12	43	91

data can support the more complex choice model, why not use the more complex model rather than a simpler model, with much lower predictive accuracy?

Summarizing the predictive information, the increase in predictive accuracy of the two-stage approach over the single-stage models appears to be more than a chance occurrence for both samples. The predictive accuracy was compared first in terms of frequency of correct prediction and then in terms of strength of prediction on the chosen alternative.

Diagnostic Information

Statistical models should not be evaluated solely or principally in terms of goodness-of-fit or predictive accuracy. Two considerations that are of equal or greater importance than predictive accuracy in selecting among models are evaluations of the underlying behavioral logic structures assumed by the various models and the diagnostic information generated. The rationale underlying the two-stage approach of a screening stage to reduce the set of alternatives followed by a more intensive comparison of the remaining alternatives has been discussed in this article.

However, one must be careful that the better predictive performance of the two-stage model is not interpreted as proof that individuals in this sample, much less the population, are using the above decision process. One can only say that the two-stage model better approximates the actual choices of the individuals in this particular data set.

To determine each individual's actual choice process would require some type of protocol analysis and/or tests of decision process. This is generally impractical for market surveys designed to gather choice data. Thus comparisons of assumed decision processes underlying the various disaggregate choice models is done at the theoretical level. Is it more realistic, for a given problem, to assume all members of the population are using a hierarchical decision process, all are using a simultaneous compensatory process, or the two-stage structure? The two-stage model is designed for cross-sectional survey data to fit a population in which, at the given point of time the survey is conducted, some members are using a strictly hierarchical process, some are using a strictly simultaneous compensatory process, and some are using the hierarchical first, followed by the simultaneous compensatory. I would argue this latter structure is more realistic for most practical choice problems.

It is the diagnostic information which provides managerial insights into the relative influence of various attributes or factors on the predicted choice distribution. Thus, for applied decision makers, the relative diagnostic information provided by the various modeling approaches may be of greater relevance than the relative predictive accuracy of the various modeling approaches. The diagnostic information provided by the two-stage modeling approach is very different from that provided by the single model approaches.

Table 1 indicates the t-values associated with the eight logit coefficients. The first column indicates the t-values on each of the total samples ($n = 91$). There are two entries for each attribute; the top value is the t-value associated with the logit coefficient in sample one. The second entry (in brackets) under the first entry is the t-value on the attributes logit coefficient in the second sample. The second column provides the t-values of the logit coefficients for the second-stage logit samples from the two-stage model.

The most striking differences between total samples and the second-stage samples are in terms of the salience of the problem solving attribute and, to a lesser degree, the drop in salience for the attribute manufacturing quality. Before going any further in analyzing or speculating about implications of the information presented in Table 1, consider the following diagnostic information presented in the first stage of the two-stage approach. Table 4 indicates the percentages of alternatives eliminated by each of the attributes in the first stage screening process using MLH.

Note that two attributes dominated the screening process: manufacturing quality and problem solving. What this seems to imply is that the industrial buyer for the utility begins by considering the suppliers in terms of perceived manufacturing quality and on the basis of which suppliers are capable of providing expert help in solving the problems associated with integrating new products and technology into existing systems or in providing the products and/or technology to provide solutions to problems that have developed on the current system. The supplier must pass the screening test on these two attributes before being considered in the final decision set.

Once a supplier is considered in the final decision set, then the product attributes, such as energy losses, become more salient in making the final choice. Recall that, because of the factor analysis, the attribute energy losses represents the factor of product attributes. Thus, having passed the screening stage, the comparison of product attributes becomes the key decision criteria. The first sample regards price as a second salient attribute in the final choice stage whereas the second sample is more concerned with maintenance requirements.

The diagnostic information from a two-stage disaggregate attribute model provides an industrial manager with new information and insights into the marketing process that are not provided by the diagnostic information of either single-stage model. It may well be that attributes vary in relative influence from one stage to another. An industrial manager who views his problem as not making the final bid list or being in the set of suppliers chosen for final negotiations may look to improve his performance on the screening attributes that are most discriminating. The manufacturer who is generally on the bid lists and a party to the final negotiations may wish to improve his/her performance on the attributes most salient in the second stage of the model. Furthermore, an examination of the percentage of times the manufacturer passes the first stage and the firm's win ratio in the second stage would give a manufacturer an indication of where, in the process, his firm is having problems. The diagnostic information then suggests the dimensions the firm should concentrate on to reduce these problems.

TABLE 4
Percentage of the Eliminated Alternatives by Attribute in First Stage of Two-Stage Model

Attribute	Sample One Sample Two
1. Invoice Price	2.46 (3.64)
2. Energy Losses	1.12% (0.78)
3. Maintenance Requirements	0% (0%)
4. Warranty Coverages	0% (0%)
5. Availability of Spare Parts	0% (0%)
6. Ease of Installation	0% (0%)
7. Problem Solving	26.16% (25.80%)
8. Manufacturing Quality	70.26% (69.78%)
Total	100% (100%)

A two-stage disaggregate attribute choice model that postulates a different decision process in each stage is presented. The rationale behind an attribute processing first stage that screens down the number of alternatives followed by a second-stage brand processing approach is generally supported by the existing literature. A particular two-stage model using MLH as a first stage and logit as a second stage is empirically applied to a real world data set. The two-stage approach did well in comparison of predictive accuracy when compared to two powerful current state of the art single-stage disaggregate choice models. It is pointed out that the diagnostic information generated by the two-stage model may contain managerial insights not found in the diagnostic information presented by single-stage models.

It is quite possible that as impressive as the current results are from the various single-stage choice models, we have only started to scratch the surface of the potential understanding to be gained from attribute choice models. The assumption underlying all of the current applications is that all members of the (sample) population are relatively homogeneous in terms of using the same single-stage process. Consideration of segments within a population using alternative choice processes (models) and the consideration of multistage disaggregate choice models such as the prototype introduced in this article could lead to a new generation of applications in which many of the basic approaches are mixed, matched, and integrated to better approximate the complex and generally nonhomogeneous reality of most choice situations.[7]

[7] This paper was received July 1985 and has been with the author for 2 revisions.

Appendix 1

Two models on the same data set can be compared in terms of the correct predictions. A McNemer test can be used to determine if the differences in correct predictions by individual is statistically significant for the sample. For each model the individual is coded a one if the alternative with the highest estimated probability of choice is the chosen alternative and zero otherwise. In a pairwise comparison of models there are four possible cases: (A) Both models predict correctly; (B) Model one is correct and model two is incorrect; (C) Model one is incorrect and model two is correct; (D) Both models are incorrect. If the number of respondents in cases B + C is as large as 10, then the following Z test statistic can be used to test the null hypothesis:

$H_0: p_1 = p_2$ and p_1 is the proportion of individuals in the population correctly predicted by model one and p_2 is the proportion of individuals in the population correctly predicted by model two.

The alternative hypothesis is, of course,

$H_a; p_1 \neq p_2$ (Glass and Stanley 1970, p. 327).

$$Z = \frac{C - B}{\sqrt{C + B}}.$$

The following two boxes give the data arrays for comparing logit to the two-stage model for the two samples. Based on the data arrays a Z score is computed using the above formula.

Sample One

		Two-Stage Models Correct	*Two-Stage Models* Incorrect	
LOGIT	Correct	A 26	B 16	42
LOGIT	Incorrect	C 33	D 16	49
		59	32	

$$Z = \frac{33 - 16}{\sqrt{33 + 16}} = 2.42,$$

Sample Two

		Two-Stage Models Correct	*Two-Stage Models* Incorrect	
LOGIT	Correct	A 38	B 7	45
LOGIT	Incorrect	C 22	D 24	46
		60	31	

$$Z = \frac{22 - 7}{\sqrt{22 + 7}} = 2.79.$$

The McNemer test is now applied to the same two samples comparing the predictive accuracy of the two-stage model approach to that of the maximum likelihood hierarchical (MLH) model.

Sample One

		Two-Stage Models Correct	Incorrect	
MLH	Correct	A 34	B 8	42
	Incorrect	C 25	D 24	49
		59	32	

$$Z = \frac{25 - 8}{\sqrt{25 + 8}} = 2.96$$

Sample Two

		Two-Stage Models Correct	Incorrect	
MLH	Correct	A 39	B 2	41
	Incorrect	C 21	D 29	50
		60	31	

$$Z = \frac{21 - 2}{\sqrt{21 + 2}} \approx 3.96$$

Appendix 2

The following is a brief description of the Maximum-Likelihood-Hierarchical (MLH) model for readers not familiar with this model or its prototype HIARC. The emphasis in this review is on providing readers with an understanding of the model rather than on technical rigor. For a more rigorous presentation including empirical applications, readers are referred to Gensch and Svestka (1979, 1984, 1985).

MLH is a multiattribute disaggregate lexicographic model. It is probabilistic and accommodates Tversky's observation that choice behavior is often inconsistent, hierarchical, and context dependent. MLH is distinct from the current individual lexicographic models in that it generates maximum likelihood estimators of the aggregate "cut-off" values. It also differs from most choice models in that it assumes individuals determine their attribute ratings by making relative judgments which are a function of the alternatives under consideration rather than absolute judgments which are independent of the other alternatives in the choice set.

MLH and the Decision Making Process

Consider the choice situation where a random sample of individuals $\{n|n = 1, 2, \ldots, N\}$ from the population are faced with a finite set of alternatives $\{j|j = 1, 2, \ldots, J\}$; where each is evaluated by every individual with respect to a finite and common set of attribrutes $\{i|i = 1, 2, \ldots, I\}$. Each individual provides a ranking of the attributes from first to last, where the first ranked attribute signifies the most important attribute. Thus different individuals may have different rank orders. Each individual also provides an attribute rating on each alternative on a monotonic poor to good scale, A^n_{ij}.

Then for each individual, the model calculates the [standardized] value of attribute i, alternative j conditional on what alternatives remain in the consideration set:

$$C^n_{ij} = \frac{\max\limits_{m \in J(n,k)} [A^n_{im}] - A^n_{ij}}{\max\limits_{m \in J(n,k)} [A^n_{im}]}. \tag{1}$$

Here C^n_{ij} is computed as the difference of the individual's evaluation of the alternative judged best with respect to attribute i of the alternative that is still in the feasible set $J(n, k)$, and his or her evaluation, of the alternative j on the attribute i. Further, this difference is standardized by dividing the numerator by the maximum attribute score of the alternative in the feasible set. $C^n_{ij} = 0$ denotes that alternative j has the best value on attribute i among the alternatives $J(n, k)$ still remaining.

Now the model finds the globally optimal set of parameters $(T_1, T_2, \ldots, T_I)$ called aggregate tolerances, which maximize the discrimination (number of alternatives eliminated), subject to the constraint that $1 - \alpha$ of the estimation population will retain their actual chosen alternative using the following "procedure:"

(1) Calculate $T_i - C^n_{ij} = r^n_{ij}$, where $0 \leq T_i \leq 1$.
(2) If $r^n_{ij} \leq 0$, set it to 0.
(3) Calculate $V^n_j = \pi r^n_{ij}$.
(4) Calculate $U_n = \sum V^n_j$.
(5) Let $P_n = V^n_{J^*}/U_n$ where J^* is the person's chosen alternative.

(6) Select that set of T's (i.e., $T_1^*, \ldots, T_I^*$) which maximizes $\pi(P_n)^{X_n}(1-P_n)^{1-X_n}$ where $X_n = 1$ if rule still has person retain choice.

Step (1) sets up a value for each attribute for each alternative for each individual ranked in terms of the importance of the attribute. Large values indicate the alternative is substantially above the cutoff threshold, i.e., the alternative will not be deleted because of its value on this particular attribute.

Steps (2) and (3) just say that an alternative can only remain (i.e., have some positive chance of being chosen) if $T_i > C_{ij}^n$. Otherwise $V_j^n = 0$. Large values of V_j^n indicate the alternative exceeds the cutoff thresholds on each attribute by a large amount. (Note that just one attribute level just equaling the threshold value renders a zero value for V_j^n.)

Step (4) adds up for each individual the values for all alternatives which have C_{ij}^n values smaller than the values of T_i, $i = 1, 2, \ldots, I$, i.e., have values that pass the threshold limits. The process ensures that $U_n > 0$, i.e., there remains at least one alternative per individual.

Step (5) just states (assumes) the person's probability of choosing the alternative is based on the relative values of the alternatives still left in the choice set.

Step (6) is the objective formation which maximizes discrimination (i.e., drives $p_n \to 1$) while also insuring that the alternative selected is the one the person chose (i.e., $X_n = 1$).

Acknowledgements. Research support for this article was provided by the National Science Foundation under grant SES-8023430.

References

Bettman, James (1979), *An Information Processing Theory of Consumer Choice*, Chicago: Addison-Wesley.

Chapman, Randall G. and Richard Staelin (1982), "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model," *Journal of Marketing Research*, 19 (August), 288–301.

Corstjens, Marcel A. and David A. Gautsch (1983), "Formal Choice Models in Marketing," *Marketing Science*, 2, 1 (Winter), 19–56.

Einhorn, Hillel J. (1970), "Use of Nonlinear Noncompensatory Models in Decision Making," *Psychological Bulletin*, 73, 221–230.

Forester, James F. (1979), "Mode Choice Decision Process Models: A Comparison of Compensatory and Noncompensatory Structures," *Transportation Research*, 13A (February), 17–28.

Gensch, Dennis H. (1984), "Targeting the Switchable Industrial Customer," *Marketing Science*, 3, 1 (Winter), 41–54.

——— (1985), "Empirically Testing a Disaggregate Choice Model for Segments," *Journal of Marketing Research*, 22 (November), 462–67.

——— (1987), "Empirical Evidence Supporting the Use of Multiple Choice Models in Analyzing a Population," *Journal of Marketing Research*, 24 (May), 197–207.

——— and Rajshekhar G. Javalgi (1987a), "The Influence of Involvement on Disaggregate Attribute Choice Models," *Journal of Consumer Research*, 14 (June), 171–182.

——— and ——— (1987b), "The Role of Disaggregate Hierarchical Choice Models in Understanding Choice Distributions," forthcoming in *Mathematical Modeling*.

——— and Joseph A. Svestka (1979), "An Exact Hierarchical Algorithm for Determining Aggregate Statistics from Individual Choice Data," *Management Science*, 25, 939–952.

——— and ——— (1984), "A Maximum Likelihood Hierarchical Disaggregate Model for Predicting Choices of Individuals," *Journal of Mathematical Psychology*, 25, 2 (June), 160–178.

——— and ——— (1985), "A Concave Constrained Maximum Likelihood Hierarchical Choice Model," Working Paper 85-11, School of Business, University of Wisconsin-Milwaukee.

Glass, Gene V. and Julian C. Stanley (1970), *Statistical Methods in Education and Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5, 103–123.

Haines, George H. (1974), "Information and Consumer Behavior," In Jagdish Sheth (ed.), *Models of Buyer Behavior*. New York: Harper & Row, 108–125.

Krishnan, K. S. (1977), "Incorporating Thresholds of Indifference in Probabilistic Choice Models," *Management Science*, 23, 11 (July), 1224–1233.

Lehtinen, Oulevi (1974), "A Brand Choice Model: Theoretical Framework and Empirical Results," *Journal of European Research*, 2, 51–68.

Lussier, Denis A. and Richard W. Olshavsky (1979), "Task Complexity and Contingent Processing in Brand Choice," *Journal of Consumer Research*, 6, 154–165.

Malhotra, Naresh K. (1984), "The Use of Linear Logit Models in Marketing Research," *Journal of Marketing Research*, 21 (February), 20–32.

——— (1986), "An Approach to the Measurement of Consumer Preferences Using Limited Information," *Journal of Marketing Research*, 23 (February), 33–40.

Newell, Allen and Herbert A. Simon (1972), *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Payne, John L. (1976), "Task Complexity and Contingent Processing in Decision Making," An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance*, 16 (August), 366–387.

Pras, Bernard and John Summers (1975), "A Comparison of Linear and Nonlinear Evaluation Process Models," *Journal of Marketing Research*, 12 (August), 276–281.

Sheridan, John E., Max D. Richards and John W. Slocum (1975), "Comparative Analysis of Expectancy and Heuristic Models of Decision Behavior," *Journal of Applied Psychology*, 60, 361–368.

Wright, Peter L. and Fredrick Barbour (1977), "Phased Decision Strategies: Sequels to Initial Screening," in North Holland TIMS Studies in the Management Science: *Multiple Criteria Decision Making*, Vol. 6, M. K. Starr and Milan Zeleny (Eds.), Amsterdam: North Holland, 91–109.

——— and Mary Ann Kriewall (1980), "State-of-Mind Effects on the Accuracy With Which Utility Functions Predict Marketplace Choice," *Journal of Marketing Research*, 17 (August), 277–93.