THE ROLE OF PROCESSING-BY-ATTRIBUTE NONCOMPENSATORY CHOICE MODELS

Rajshekhar G. Javalgi
College of Business Administration, Marquette University, Milwaukee, WI 53233

Dennis H. Gensch
School of Business Administration, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

Abstract. Research work on modeling consumer multiple choice problems using logit, regression, and probit is gaining more attention. However, in their work, Russ (1971), Tversky (1972), Newell and Simon (1972), Tversky and Sattath (1979), and Gensch and Svestka (1979; 1984) indicate that for many problems, choice behavior appears to be context dependent and hierarchical. With this specific issue in mind, this paper discusses a mathematical model which estimates threshold tolerances, eliminates nonchosen alternatives, provides choice probabilities and finally offers diagnostic information regarding the key attributes that are responsible for making a final decision. The use of other individual specific models such as: lexicographic, conjunctive, etc., have been briefly explicated.

Keywords. choice; disaggregate; threshold tolerance; hierarchical; noncompensatory.

INTRODUCTION

Research focused on the development of disaggregate choice models (e.g., logit, regression, and probit), and derived from the principle of utility maximization, is receiving more attention. Other multiattribute models, built upon the notion of threshold concepts which are noncompensatory in structure, need to be examined further. The existence of the thresholds has been acknowledged for many years (Georgescu-Roegen, 1936).

The simultaneous compensatory choice models (e.g., logit, regression, and probit) are common to many disciples. These disaggregate models use individual observations to estimate parameters of the population. Almost all empirical work on multiattribute choice models uses a simultaneous compensatory structure. The reason may be that there are efficient algorithms that are commonly available, which allows the research to empirically analyze data in order to predict choices, even though the algorithm does not purport the underlying behavioral process.

However, Tversky (1972), Newell and Simon (1972), and others, argue that for many problems, choice behavior seems to be both hierarchically structured and context dependent. A well known noncompensatory model proposed by Tversky (1972), which is probabilistic and content dependent, is the Elimination-by-Aspect (EBA) model. The EBA model is designed for those situations in which the salient attributes are unique to single alternatives, or a specific subset of the alternatives in the choice set. Another model developed by Gensch and Svestka (1984), which is like the EBA (context dependent and probabilistic), is the Maximum Likelihood Hierarchical (MLH) Choice model. This MLH model is designed for the situation in which salient attributes are common to all alternatives. The remainder of this paper will attempt to explain the concepts and potential uses of the MLH model. The role of other individual specific models such as: lexicographic, conjunctive, and disjunctive models, are discussed.

We first briefly explicate the data requirement and underlying behavioral decision process of the MLH model.

Data Requirement to Model the Choice Process Using the MLH

Unlike the individual-based lexicographic model, the disaggregate MLH model operates in two distinct modes: (1) calibration (generating aggregate estimates of the threshold tolerances), and (2) prediction (employing given estimates to predict individual responses). In the calibration mode, MLH generates the aggregate estimates of the threshold tolerances from information provided by a sample of individuals. In order to generate calibrated coefficients, information obtained from each individual should include: (i) rank order of the attributes, indicating the sequence in which they are considered, (ii) a set of self-perceived values of the given alternatives, with respect to each attribute, and (iii) their information on actual choice. Here items (i) and (ii) are imperative in the prediction mode, but item (iii) is needed if the predictions are to be compared to the actual selected alternatives.

A Focus on How MLH Eliminates Nonchosen Alternatives

The concept of individual threshold tolerances, which is fundamental to the MLH model, is defined in the context of individual behavior. The threshold tolerance is assumed to be a relative value, related to the attribute values of the alternatives under consideration, rather than some absolute value an individual constantly uses independent of his set of alternatives. The MLH mathematical model, which generates aggregate estimates of these threshold tolerances, uses sequential processing of ranked attributes, and eliminates the nonchosen alternatives at each stage of the process. MLH is a disaggregate hierarchical model, and is distinct from current lexicographic models in that it does not require the analyst to know a priori individual threshold tolerances (cutoff values).

To understand how elimination works, let us clarify some notations. Denote the i-th attribute associated with the r-th importance ranking of decision maker n as i(r). Furthermore, define A^n_{kj} as the perceived value (rating) of alternative j, with respect to attribute i(r), given by the n-th individual. Let $J(n,k)$ be the set of all alternatives still under consideration by individual n, when the j(r)-th attribute is considered. Further, associated with any given attribute i(r), there is a critical tolerance T_i^n between the decision maker n's evaluation of any alternative on the attribute i(r), and an acceptable standard. This quantity T_i^n will be considered distributed over the population Π. These critical tolerances, denoted by T_i, are central tendency parameters of these distributions with certain special properties. Since the model is formulated (see: Gensch and Svestka 1984) as a concave programming problem, whose solutions are globally optimal, these solutions are precise aggregate estimators of T_i. If aggregation is to be affected, the information processed by two or more individuals must be compatible. Hence, the standardized individual is defined as a real number C_{ij}^n; where j ranges over all alternatives which have not been eliminated, and I ranges over the set of attributes which are arranged in the order of importance. To explain more clearly, at the beginning of the choice process, an individual considers the full set of alternatives denoted by $J(n,0)$ $\{j|j=1,2,\ldots,J\}$; where n denotes the n-th individual in the sample, and zero indicates that the alternative has been evaluated with respect to no attributes. After an individual implies his cut point for the first ranked attribute to the alternatives, the set of alternatives (which may or may not be reduced) is denoted by $J(n,1)$. In general, after the application of the first K ranked attributes, the set is $J(n,K)$. Thus the individual standardized values, which are a function of those alternatives still under consideration, are defined as:

$$C_{ij}^n = \frac{\max\limits_{m\varepsilon J(n,k)} [A_{im}^n] - A_{ij}^n}{\max\limits_{m\varepsilon J(n,k)} [A_{im}^n]} \quad . \qquad (1)$$

It is clear from the above formation that C_{ij}^n lies in the interval [0,1], and that data from two or more individuals is compatible. It may also be noted that once the set $J(n,k)$ is reduced to a single alternative, the values C_{ij}^n remain fixed.

Consider now the aspect of the individual's set of threshold tolerances (cut points). Without a loss of generality, these cut points may be standardized in the manner of individual values. The standardized individual cut points, which also lie in the interval [0,1], are denoted τ_i^n and are called individual tolerances.

The aggregate threshold tolerances, denoted τ_i, are central tendency parameters of these distributions τ_i^n, with certain special properties. MLH generates the estimates of these parameters which are called aggregate tolerances.

Initially, an individual n evaluates the set of alternatives with respect to his first ranked attribute i. An alternative j will be eliminated if the individual's tolerance (standard cut point) τ_i^n is less than the individual's (standardized) value for the alternative, with respect to that attribute. That is, alternative j will be eliminated if:

$$\tau_i^n < C_{ij}^n \qquad (2)$$

and not eliminated if

$$\tau_i^n \geq C_{ij}^n \qquad (3)$$

Consider now the definition of these sets (2), where the individual tolerances τ_i^n are replaced by estimates of the distribution parameters τ_i. The application of the parameter estimates to the individual data will simulate the evaluation process for a set of individuals, and the definition of the resulting sets of alternatives remaining. For the mathematical formulation and generation of the choice probabilities, refer to Gensch and Svestka (1984).

Data Description

In the Agriculture data set developed for the U.S. Department of Agriculture, a sample of 800 Iowa farmers identified their favorite cooperative and independent retail outlets. They rated each outlet in terms of 24 attributes, and provided a ranking of the attribute relative to how important the particular attribute is in determining at which retail outlet they would purchase their fertilizer. Farmers indicated at which retail outlet they made their major fertilizer purchase during the given year. The attributes were factor analyzed so as to reduce the multicollinearity in the data set. By observing the attribute loadings on the independent factors and the average importance ranking of the attribute, eight relatively independent and important attributes were selected as the independent variables for choice modeling (see Gensch, 1983), for details and how the attribute space was reduced.

The other sections of the questionnaire included demographics about the respondents and their spouses' interest in farming. Respondents also indicated the number of courses they have taken related to the agricultural and farming area.

Discrimination and Prediction of MLH -- Some Empirical Results

The aspect of discrimination refers to the number of alternatives which are eliminated for a set of individuals. Another aspect, called predictive accuracy, refers to perfect prediction or the number of individuals in a hold-out sample who, after the application of a set of calibration tolerances (which were previously generated by an independent set of individuals), retain exactly one alternative, and the alternative retained is the chosen alternative.

Table 1 below contains the discriminant and predictive performance of the MLH model, when applied to an agricultural data set described below.

TABLE 1. Performance of the MLH Model on 8 Attributes and 2 Alternatives

Holdout Number	Sample Size	% Discrimination (MLH)	% Prediction (MLH)
1	150	97.33	88.00
2	150	98.00	83.33
3	100	100.00	86.15

As discussed previously, discrimination refers to the number of alternatives which are eliminated for a set of individuals. Consider, for example, if 10 individuals each have two alternatives in their choice set, and the model eliminates 10 of the 20 total alternatives leaving exactly one alternative for each individual, 100% discrimination would have occurred. Now the MLH will attain perfect discrimination in a calibration run, but the application of the tolerances generated from a calibration sample, and applied to a totally independent holdout sample generally results in less than perfect performance.

In Table 1, for the holdout sample number 1 containing 150 respondents, the discriminating ability of the MLH model is 97.33% and the predictive performance is 0.88. This means that 88% of the individuals who were processed through the lexicographic-type MLH model retained their chosen alternative. The remaining individuals either did not retain their chosen alternative or may have retained more than one alternative. Since MLH approximates the choice process of consumers who process by attribute, it could be possible that one or two individuals may retain all alternatives in the process. The information that is presented in Table 1 can be useful to marketing practitioners, in the sense that one can compute the aggregate market share of each of the alternatives that are used in the decision process by individuals in the sample (population).

In addition to discrimination and prediction information, MLH also generates a report stating the number of alternatives correctly and incorrectly eliminated by the attributes. Table 2 presents this information.

TABLE 2. Attribute Discrimination

Attribute	No. of Alternatives Eliminated	No. Correctly Eliminated	% Total Eliminated
1	36	36	35.64
2	22	20	33.78
3	0	0	0.00
4	11	7	10.89
5	0	0	0.00
6	0	0	0.00
7	11	8	10.89
8	21	17	8.79

The Table 2 information is very helpful in knowing the effective number of attributes necessary to reduce the set of alternatives to the selected choice. Furthermore, it also indicates key attributes which eliminated most of the alternatives at each stage of the decision process. In Table 2, attributes 1 and 2 eliminated close to 70% of the alternatives. This indicates that attributes 1 and 2 are key attributes that are responsible for eliminating most of the alternatives. This information is especially important for practitioners in modifying or developing policy-oriented decisions regarding the product (alternative) attributes. In their policy decisions, they may strongly emphasize these salient attributes to the appropriate market segment.

In the next section, the role of sequential processing models as a function of task-complexity, and various other factors influencing the use of sequential processing models are discussed.

THE ROLE OF SEQUENTIAL PROCESSING MODELS

The consumer literature supports the view and in many cases presents empirical evidence, that consumers simplify their choice problems by using attribute evaluation in a sequential manner to eliminate a number of alternatives (size of the choice set) from consideration (Payne, 1976; Russ, 1971). While there is general agreement in the literature that, as the number of alternatives increases, decision makers tend to apply noncompensatory models to whittle down the domain of alternatives to manageable size, and thereby make a selection from the remaining alternatives, using a compensatory evaluation approach (Payne, 1976; Olshavsky, 1980).

Many of the researchers (Payne, 1976) strongly believe that decision making is a multistage process with a noncompensatory hierarchical attribute stage, used to reduce the alternatives to a feasible set, followed by a simultaneous compensatory approach used to select the chosen alternative from the feasible set. In their study, Russo and Johnson (1980) and Van Raaij (1976) provide support for sequential screening by attribute in the early phase if the process to eliminate alternatives, and a brand processing (compensatory evaluation) strategy in later phases of the process. The results of these studies are analyzed from the individualized modeling process.

In another study, which uses the disaggregate mathematical modeling approach, Gensch (1985) showed that, for more than three alternatives, the two-stage MLH-MNL model predicted better than the single phase MNL (or MLH) model. Here it is interesting to note that both MLH and MNL are disaggregate probabilistic models; where MLH is derived from the threshold-based concept and MNL is a utility-based principle model. This approach, which whittles down the alternatives using MLH and gives final choice probabilities using the MNL model, not only gives the information on market share of each alternative, but also indicates which attributes are responsible to reduce the size of the choice set. Thus these findings indicate that, as the number of alternatives increases, the utilization of the hierarchical attribute algorithm which is used to narrow the alternatives to a more feasible size, has gained more credibility.

Lastly, incomplete data about available alternatives, and added extraneous data about each alternative, affects the use of decision rules. Consumers often face decisions where some of the relevant information on one or more of the alternatives is not available. In this situation, incomplete information creates more difficulty for compensatory models and less for noncompensatory screening-by-attribute models (Wright and Barbour, 1977). Slovic and MacPhillamy (1974) provide evidence to report that a "Common Dimension Strategy," a strategy in which a decision maker chooses the best option on the dimension where the data is complete, is preferred when the data is incomplete.

In addition to incomplete information, noncomparable scalings are common in real world choice problems. Wright (1977) reported that as incomplete scalings and extraneous data were added, the use of compensatory strategies decreased and lexicographic rules increased. Finally, a situation in which a linear compensatory strategy led to ties in evaluations, even without incomparable scalings of extraneous data, with a use of lexicographic strategy, was examined by Wright (1977). Thus, from the above discussion, it seems obvious that the use of the various types of choice models are not only affected by the individual differences, but also by other factors such as size of the choice set, and incomplete data information. Recognition that segments within the population may be using different decision processes on the same choice and how to a priori identify these segments appears to be a research area that could offer significant breakthroughs in our understanding of choice modeling. We may have to switch from the current practice of assuming one process (model) is used by the entire population.

CONCLUDING REMARKS

An important part of understanding consumer behavior is the construction of formal representations of choice processes. Formalisms such as decision nets, conjoint analysis, and discrete mathematical models all attempt to model the relationship between characteristics of a product and observed choice. Most applied mathematical models used in the consumer behavior literature represent choice processes in terms of some form of algebraic model. However these models do not represent the way consumers make their decisions. The present paper clearly delineates that hierarchical models tend to represent choice behavior of individuals better than compensatory models.

The hierarchical disaggregate model, for example, the MLH model, and the use of other sequential processing models are briefly explained. We hope that this paper stimulates the readers to further apply such models to their problems. More work in this direction will enable us to better understand the underlying behavioral process of decision makers.

REFERENCES

Gensch, D. H. (1983). Iowa cooperative fertilizer retail outlets: Farmers' attitudes and perceptions (Agricultural Cooperative Service Research Report 29, Washington, DC: U.S. Department of Agriculture).

Gensch, D. H. (1983). Multistage disaggregate choice model in marketing, Working Paper, University of Wisconsin-Milwaukee.

Gensch, D. H. and J. A. Svestka (1984). A maximum likelihood hierarchical disaggregate model for predicting choices of individuals. Journal of Mathematical Psychology, 160-178.

Georgescu-Roegin, N. (1958). Threshold in choice and theory of demand. Econometrica, 26, 157-168.

Newell, A. and Simon, H. A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Olshavsky, R. W. (1979). Task complexity and contingent processing in decision making: A replication and extension. Organizational Behavior and Human Performance, 24, 300-316.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

Russ, F. A. (1971). Consumer evaluation of alternative product models. Unpublished Ph.D. Dissertation, Carnegie-Mellon University.

Slovic, P. and D. MacPhillamy (1974). Dimensional commensurability and a utilization in comparative judgement. Organizational Behavior and Human Performance, 11, 172-194.

Tversky, A. (1972). Choice by elimination. Journal of Mathematical Psychology, 9, 341-367.

Tversky, A. and Sattath, S. (1979). Preference trees. Psychological Review, 86, 542-573.

Van Raaij, W. F. (1976). A contingency approach to consumer information processing. Unpublished Paper, Tilburg University.

Wright, P. and B. Fedrick (1977). Phased decision strategies: Sequels to an initial screening. In Martin K. Starr and Milan Zeleny (eds.), North Holland/TIMS Studies in the Management Sciences, Vol. 6: Multiple Criteria Decision Making. Amsterdam: North-Holland, 91-109.