MARKETING SCIENCE
Vol. 14, No. 3, Part 1 of 2, 1995
Printed in U.S.A.

THE SPATIAL REPRESENTATION OF HETEROGENEOUS CONSIDERATION SETS

WAYNE S. DESARBO AND KAMEL JEDIDI
The University of Michigan
Columbia University

Consideration sets have been the recent focus of a large volume of research in marketing. The primary orientation of this stream of research has been toward consideration set composition, measurement, and the theoretical formation process itself. This paper proposes a new multidimensional scaling methodology (MDS) devised to spatially represent preference intensity collected over consumers' consideration sets. Predictions concerning the probability of consideration set membership, as well as the degree of preference intensity of these brands within a consideration set, are possible from such a model. In addition, consumer heterogeneity is accommodated vis á vis latent market segment level estimation. The technical details of the proposed MDS methodology are presented. Two actual commercial applications of the procedure are provided in the modeling of consideration sets and respective preference intensity for "intenders" for mid-size and luxury automobiles. Finally, limitations and directions for future research in this area are discussed.
(Buyer Behavior; Choice Models; Scaling Methods; Segmentation Research)

1. Introduction

The brand choice process has been viewed as a dynamic consumer phenomenon of narrowing choice alternatives from many to few. Shocker, Ben-Akiva, Boccara, and Nedungadi (1991) recently characterized brand choice as a sequential decision-making process based upon hierarchical or nested sets of brand alternatives. Here, there are a number of brands or services in any specified product/service class available for purchase in any given time period or market. This "universal set" contains all the brands/services in the class. The "awareness" set consists of the subset of brands/services in the universal set of which a consumer is "aware." Typically, this awareness set contains many fewer brands/services than in the universal set. The "consideration set," which is the focus of this paper, evolves from the awareness set. According to Shocker et al. (1991), a consideration set consists of goal-satisfying brands/services that are salient or accessible at a particular time. Roberts (1989), Hauser and Wernerfelt (1990), and Roberts and Lattin (1991) view consideration set formation as a tradeoff between utility and costs—a view adopted here. According to these authors, it comprises those brands or services that a consumer would seriously consider purchasing on a particular occasion. Again, it typically has fewer items in it than in the awareness set. Finally, the "choice set" contains those brands/services seriously considered just prior to actual purchase.

A large body of research on consideration sets has recently evolved within the past decade. Major areas of investigation involve: behavioral frameworks for partitioning

0732-2399/95/1403/0326$01.25

brands/services into these sets (e.g., Narayana and Markin 1975; Brisoux and Laroche 1980; Nedungadi 1987, 1990; Gensch 1987), characteristics of consideration sets (e.g., Brown and Wildt 1992, Brisoux and Cheron 1990, Crowley and Williams 1991), characteristics of brands in the sets (e.g., Klenosky and Rethans 1989, Heath and Chatterjee 1991), processes (models) leading to the formation of consideration sets (e.g., Roberts 1989, Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Ratneshwar and Shocker 1991), and consideration set measurement (e.g., Brown and Wildt 1992).

A number of quantitative models of consideration sets have appeared in the marketing literature pertaining to various aspects of consideration set formation. Hauser and Wernerfelt (1990) propose a cost-benefit approach for representing the process of consideration set formation. Conceptually, the choice process is viewed as a decision on whether to evaluate a brand, and a subsequent decision to include it in the consumer's consideration set after evaluation. The primary motivation for the Hauser and Wernerfelt (1990) model was derived from economic notions of search and information search costs (Stigler 1961). Their conceptualization is that a rational consumer should include brands in their consideration set if their expected incremental utility is higher than the decision costs to retain them in the consideration set. Similarly, a brand may be eliminated from the consideration set if it does not add a greater expected utility than the decision cost to retain it in the consideration set. Hauser and Wernerfelt (1990) examine the aggregate, normative implications of their model for pricing and advertising. Their formulation does not accommodate individual or segment level empirical analyses.

Roberts (1989) and Roberts and Lattin (1991) provide an alternative cost-benefit model of consideration set composition that can be calibrated at the individual level using a logit model formulation (see also Nedungadi and Kanetkar (1992) for a simpler logit approach). Roberts and Lattin (1991) model consideration as a compensatory process given its apparent robustness (see Johnson and Meyer 1984) in representing a variety of decision processes. Here too, a brand will enter into the consumer's consideration set if the increase in expected utility due to choosing from the enlarged set offsets the associated costs of adding the brand. Utilizing these cost thresholds, Roberts and Lattin (1991) develop a greedy algorithm for consideration set composition based on maximum utility principles. Given data on the attributes of the available brands and the self-reported consideration, they maximize a logit log-likelihood function to estimate the attribute importance and the threshold value (assumed constant across consumers). A major strength of this work is the empirical study presented involving $N = 121$ Australian households, where their proposed model predicts consideration set composition reasonably well with respect to breakfast cereals.

Given this consideration set modeling tradition, we wish to provide an alternative paramorphic representation of the consideration set composition problem. As with the two major approaches above, we also assume a random (net) utility framework with thresholds in a compensatory process. However, unlike the Hauser and Wernerfelt (1990) normative formulation of the aggregate market, we wish to provide a more descriptive, empirical approach that can be implemented at the market segment level. Unlike Roberts and Lattin (1991), we propose to uncover what the underlying attributes or dimensions of consideration set formation (rather than assume they are given and known). The goal of this research is to provide a means of segmenting consumers on the basis of their consideration sets in deriving a graphical representation of the interrelationship of brands and (unknown) market segments, so as to easily infer the consideration sets by market segment. In addition, the composition of the market segments is simultaneously estimated and determined by the patterns of evoked consideration values collected empirically.

The next section of the paper provides a description of MDS models and their applicability to such data. A technical description of the proposed spatial model is provided, as well as the estimation procedure. Two commercial applications concerning automobiles are given in the following section. Comparisons with more traditional spatial MDS procedures are discussed. Finally, some limitations and directions for future research are given.

2. A Multidimensional Scaling Approach for Representing Consideration Sets

Multidimensional Scaling (MDS) models for the analysis of preference/dominance data typically provide joint space representations of brands and consumers (Carroll 1980). On these maps, brands are depicted as points and consumers as either vectors or ideal points. In a vector or scalar products representation, consumers are assumed to value the higher levels of the product attributes (dimensions). The orthogonal projection of a particular brand onto a consumer's vector provides information about the utility of the brand to the consumer. In an ideal point representation, consumer utility decreases as the brand distance from the consumer's ideal point increases. MDS preference models have been widely utilized in marketing for concept testing (Urban and Hauser 1993), positioning new products (DeSarbo and Rao 1984, 1986), and market segmentation (Cooper 1983). Note that other spatial models including correspondence analysis (see Hoffman and Franke 1986) and optimal scaling procedures (e.g., PRINCALS) have also gained in popularity among marketing practitioners of late. Unfortunately, the lack of a meaningful interpretation of distance or scalar products between row and column points limits their use, especially for market segmentation. In this paper, we will consider the vector model, although our approach can be generalized to the ideal point model as well.

As indicated by DeSarbo, Manrai, and Manrai (1994), traditional MDS preference models are limited in their capacity of parsimoniously portraying the structure underlying empirical preference data. It is not uncommon for a market research supplier to collect data from thousands of consumers regarding their preferences for various existing brands in a certain product category. Analyzing such data with traditional MDS preference models would result in a saturated space with excessive brand points and consumer vectors, rendering interpretation very difficult. In face of this problem, researchers/users have typically resorted to cluster analyzing the obtained consumer coordinates and summarized their maps by plotting clusters in lieu of individual consumers. Ling (1971), Baker and Hubert (1976), and Chang (1983) have all cautioned against using this two-step procedure. They demonstrate that by clustering based on components (dimensions) with larger eigenvalues, one may risk losing valuable information concerning distances usually contained in the components with small eigenvalues. DeSarbo, Howard, and Jedidi (1991) have addressed this problem by developing a latent class vector MDS model called MULTICLUS that simultaneously performs MDS and cluster analysis. Their approach is parsimonious in that they explicitly model heterogeneity at the segment level. See DeSarbo, Manrai, and Manrai (1994) for a recent review of other such latent class MDS models.

Another limitation of MDS preference models lies in their incapacity to accommodate censored preference data.[1] This type of data is typically collected in market research studies where consumers are asked to render their judgment only for those brands that

[1] We utilize the econometrics usage of the term "censored" as "when a dependent variable is zero for a significant fraction of the observations" and "when values in a certain range are all transformed to (or reported as) a single value" (Greene 1993, p. 691). As Maddala (1992) states, censoring at zero with a tobit model still assumes that the latent random variable can assume negative values (i.e., strength of nonconsideration), which is assumed in the present case given the use of a scalar product, bilinear function which can be negative or positive. Strictly speaking, treating the censored data as missing, or attempting to "estimate" such missing/censored observations, are inappropriate since one knows that the preference intensity measure necessarily would be smaller for such censored observations than each respondent's minimum noncensored observation. Such order "restrictions" would not typically be enforceable via mean substitution, regression, or SVD missing data replacement methods. By treating such data as missing, one loses such order information. In addition, the vast majority of joint space MDS software can not accommodate these types of missing values.

fall within their consideration sets (Howard and Sheth 1969, Wright and Barbour 1977). For example, in ASSESSOR (Silk and Urban 1978), respondents are bound to evaluate the brands they evoke. This approach of collecting data is efficient, realistic, and is likely to produce more reliable data (cf. Green, Tull, and Albaum 1988). First, it minimizes the cognitive effort required from the respondent and hence may reduce boredom, fatigue, and attrition (Malhorta 1986). Second, it filters out the problem of brand unfamiliarity (Alba and Hutchinson 1987, Chatterjee and DeSarbo 1992) in a natural way since unfamiliar brands are typically not part of the consideration set. Third, and most importantly, this approach of collecting data fits nicely with theory in consumer behavior that characterizes human decision-making in complex situations (e.g., decisions involving a large number of alternatives) by a phased decision rule (see Bettman 1979, p. 215). In such a framework, consumers are conceptualized to undertake a two-stage process. They first filter the available alternatives and then undertake a detailed analysis on the reduced (consideration) set (Wright and Barbour 1977; Bettman 1979; Bettman, Johnson, and Payne 1991).

Analogous to the problem of using OLS in the case of censored dependent variables (see Amemiya 1985 p. 367), the application of current MDS models to the analysis of such censored preference data would produce biased and inconsistent estimates. It is the purpose of this paper to generalize the MULTICLUS model to accommodate the analysis of censored preference data as collected in the investigation of consideration sets. More specifically, we develop a latent structure vector MDS model that explicitly portrays heterogeneous consumers' considerations sets. Like Hauser and Wernerfelt (1990) and Roberts and Lattin (1991), we model consideration as a compensatory process, and we assume that a consumer will include a brand in his or her consideration set if its utility exceeds a certain threshold. This threshold reflects the trade-off between the utility gained by having the brand in the consideration set and the implicit cost incurred from its inclusion such as mental storage, processing costs, and information search costs (Roberts and Lattin 1991). As in MULTICLUS, we capture consumers' heterogeneity at the segment level using a latent structure approach. Assuming mixtures of conditional normal distributions, we derive an *E-M* algorithm to derive a joint space representation of brand location and segment vectors from such censored preference data collected for consideration sets. "Market segment," in this context, is utilized to denote a group of consumers who possess similar consideration sets.

A. *The Proposed MDS Model*

Let g index membership in a market segment ($g = 1, \ldots, G$), i index consumers ($i = 1, \ldots, I$), t index dimensions ($t = 1, \ldots, T$), and j index brands ($j = 1, \ldots, J$). We model consumer i's utility for brand j, conditional on his or her belonging to market segment g, using an additive utility function (scalar product). Thus:

$$\Delta_{ij}^* \mid g = \sum_{t=1}^{T} x_{jt} y_t^g + u_{ij}^g = \mathbf{x}_j \mathbf{y}'^g + u_{ij}^g, \tag{1}$$

where x_{jt} is the amount of dimension (attribute) t possessed by brand j, y_t^g is the weight attached (vector) by segment g to dimension t, $\mathbf{y}^g = (\mathbf{y}_t^g)$, and u_{ij}^g is an error term. Here, $\mathbf{x} = ((x_{jt}))$ can be either measured directly (external analysis) or estimated from the data (internal analysis) and is assumed to be invariant across segments; $\mathbf{x}_j$ is the jth row of $\mathbf{x}$. Note, u_{ij}^g reflects random fluctuations in the measurement of utility due to uncertainty about product dimensions, variety seeking, use occasion, and other contextual factors. We assume $E(u_{ij}^g) = 0$ and $\text{Var}(u_{ij}^g) = \sigma_g^2$ for all i, j, g.

Brand j will enter the consideration set of consumer i in segment g if its utility exceeds a certain segment specific threshold δ_g^*. In other words:

$$\lambda_{ij} = \begin{cases} 1 & \text{iff} \quad \Delta^*_{ij} \mid g > \delta^*_g, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where λ_{ij} is a consideration set indicator. Modeling brand consideration as a trade-off between utility and mental storage and processing costs, Roberts and Lattin (1991) derived a similar condition. In their formulation, δ^*_i (subject specific) measures the minimum utility requirement any brand needs to satisfy in order to enter the consideration set. This requirement is due to the costs associated with adding a new brand to the set. Typical costs of consideration include information search costs for consumer and industrial durables, and mental maintenance and processing costs for frequently purchased goods (Roberts and Lattin 1991). Note that δ^*_g can be set to zero (as a program option) to improve interpretation (the origin of the space becomes the threshold) without much loss of generality. This formulation can be seen as a spatial version of the standard tobit model (Tobin 1958, Amemiya 1985).

Suppose $\Delta^*_{ij} \mid g$ has a conditional normal distribution, $f(\cdot)$. Then, the unconditional distribution of Δ^*_{ij} is a finite mixture of such distributions. That is:

$$\Delta^*_{ij} \sim \sum_{g=1}^{G} w_g f(\Delta^*_{ij} \mid \mathbf{y}^g, \mathbf{x}, \sigma^2_g) = \sum_{g=1}^{G} \frac{w_g}{\sigma_g} \phi\left(\frac{\Delta^*_{ij} - \mathbf{x}_j \mathbf{y}'^g}{\sigma_g}\right), \tag{3}$$

where $\mathbf{w} = (w_1, \ldots, w_G)'$ is the vector of the G mixing proportions such that $w_g > 0$ and $\sum_{g=1}^{G} w_g = 1$; $\phi(\cdot)$ is the standard normal density function.

In practice, the utility variable Δ^*_{ij} is observed only for the brands that fall within consumer i's consideration set and thus may be viewed as a partially latent variable. Define the observable counterpart of Δ^*_{ij} as:

$$\Delta_{ij} = \begin{cases} \Delta^*_{ij} & \text{if} \quad \Delta^*_{ij} > 0, \\ 0 & \text{if} \quad \Delta^*_{ij} \leq 0; \end{cases} \tag{4}$$

that is, Δ_{ij} is the observed censored preference utility value of brand j for consumer i. Its distribution is:

$$h(\Delta_{ij} \mid \mathbf{y}, \mathbf{x}, \mathbf{\Sigma}, \mathbf{w}) = \sum_{g=1}^{G} w_g \left[\left(1 - \Phi\left(\frac{\mathbf{x}_j \mathbf{y}'^g}{\sigma_g}\right)\right)\right]^{1-\lambda_{ij}} \left[\frac{1}{\sigma_g} \Phi\left(\frac{\Delta_{ij} - \mathbf{x}_j \mathbf{y}'^g}{\sigma_g}\right)\right]^{\lambda_{ij}}, \tag{5}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal, $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^G)$, and $\mathbf{\Sigma} = (\sigma^2_1, \ldots, \sigma^2_G)$.

The likelihood function for a sample of I randomly drawn consumers $(\Delta_1, \ldots, \Delta_I)$ from the above mixture (assuming independence over brands) is then:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} h(\Delta_{ij} \mid \mathbf{y}, \mathbf{x}, \mathbf{\Sigma}, \mathbf{w}). \tag{6}$$

The problem here is to maximize L, or $\ln L$, with respect to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{\Sigma}$, and $\mathbf{w}$, given the sample data $(\Delta_1, \ldots, \Delta_I)$, prespecified values of T and G, and taking into account the constraints imposed on $\mathbf{w}$ and $\sigma^2_g > 0$, for all g. The latter condition is necessary since the likelihood function L is unbounded when $\sigma^2_g = 0$, and hence consistent estimators are not possible.

Within each iterate, current maximum likelihood estimates $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$, $\hat{\mathbf{\Sigma}}$, and $\hat{\mathbf{w}}$ are utilized to estimate the posterior probabilities of membership (using Bayes' rule):

$$\hat{P}_{ig} = \frac{\hat{w}_g f(\Delta^*_{ij} \mid \hat{\mathbf{y}}^g, \hat{\mathbf{x}}, \hat{\sigma}^2_g)}{\sum_{k=1}^{G} \hat{w}_k f(\Delta^*_{ij} \mid \hat{\mathbf{y}}^k, \hat{\mathbf{x}}, \hat{\sigma}^2_k)}. \tag{7}$$

These probabilities represent a fuzzy classification (clustering) of the I consumers into G latent classes or market segments. Thus, the proposed latent structure vector MDS methodology estimates the joint space and market segments (including membership) all simultaneously.

B. *Model Identification*

Before discussing the *E-M* algorithm developed for parameter estimation, it is necessary to examine the identifiability of the model. There are two issues: (1) identification of the vector model parameters in an MDS context, and (2) identification of the model parameters in a finite mixture model context. Consider first the former issue. It is well known in the marketing and psychometric literature that vector model solutions are characterized by rotation and scale indeterminacies. That is, we could postmultiply $\mathbf{x}$ in (1) by an orthogonal matrix $\mathbf{A}$ *and* compensate for this by premultiplying $\tilde{\mathbf{y}}'^g$ by $\mathbf{A}'$ without altering their scalar product. Also, we could postmultiply $\mathbf{x}$ by a diagonal matrix $\mathbf{D}$ and compensate by premultiplying $\mathbf{y}'^g$ by $\mathbf{D}^{-1}$ without affecting the value of their scalar product. These indeterminacies can be circumvented by, for example, orthogonally rotating $\mathbf{x}$ to a simple structure and normalizing the transformed latent structure vectors to unit length. Concerning the second issue regarding the identifiability of the finite mixture, it has been shown that the parameters of finite mixture of univariate normals are identified (see Yakowitz and Spragins 1968 and Titterington, Smith, and Makov 1985, p. 38). However, one potential problem that may occur is solution degeneracy if any of the mixing proportions are equal to zero. This is likely to occur if an excessive number of market segments is extracted. However, this problem can be circumvented simply by reducing the number of segments if null mixing proportions are estimated.

C. *Model Estimation*

The *E-M* algorithm is a general iterative method for obtaining maximum likelihood estimates (Dempster, Laird, and Rubin 1977). This method is suited for models that deal with unobserved/missing data (e.g., censored regression models). The maximization of L in (6) can be achieved via an *E-M* algorithm since Δ is censored and segment membership is unknown. In addition, the *E-M* method provides two important advantages. First, it provides monotone increasing values of the log-likelihood function. Hence, convergence to at least a locally optimum solution is guaranteed (see Titterington, Smith, and Makov 1985). Second, the *E-M* algorithm seems to converge somewhat rapidly in the case of (aggregate) censored regression models as revealed from the simulation study performed by Schmee and Hahn (1979).

In order to formulate the *E-M* algorithm, we begin by defining a segment indicator variable z_{ig} as:

$$z_{ig} = \begin{cases} 1 & \text{iff consumer } i \text{ belongs to market segment } g, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We assume that, for a particular consumer i, the nonobserved vector $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, z_{iG})'$ is iid multinomially distributed with probabilities $\mathbf{w}$. That is:

$$(\mathbf{z}_i \mid \mathbf{w}) \sim \prod_{g=1}^{G} w_g^{z_{ig}}. \tag{9}$$

The conditional distribution of Δ_i^* given $\mathbf{z}_i$ is therefore:

$$\Delta_i^* \mid \mathbf{z}_i \sim \prod_{g=1}^{G} \left[\prod_{j=1}^{J} f(\Delta_{ij}^* \mid \mathbf{y}^g, \mathbf{x}, \sigma_g^2) \right]^{z_{ig}}. \tag{10}$$

With $\mathbf{z} = ((\mathbf{z}_{ig}))$ considered as missing data and $\Delta^* = ((\Delta^*_{ij}))$ as the data matrix with some unobserved values, the complete data log likelihood function to be maximized is given by:

$$\ln L_c(\mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \mathbf{w} \mid \Delta^*, \mathbf{z}) = \sum_{i=1}^{I} \sum_{g=1}^{G} \sum_{j=1}^{J} z_{ig} \ln f(\Delta^*_{ij} \mid \mathbf{y}^g, \mathbf{x}, \sigma_g^2) + \sum_{i=1}^{I} \sum_{g=1}^{G} z_{ig} \ln (w_g)$$

$$= -\frac{1}{2} \sum_{i=1}^{I} \sum_{g=1}^{G} \sum_{j=1}^{J} z_{ig} \ln (2\pi)$$

$$- \frac{1}{2} \sum_{i=1}^{I} \sum_{g=1}^{G} \sum_{j=1}^{J} z_{ig} \ln (\sigma_g^2) + \sum_{i=1}^{I} \sum_{g=1}^{G} z_{ig} \ln (w_g)$$

$$- \frac{1}{2} \sum_{i=1}^{I} \sum_{g=1}^{G} \sum_{j=1}^{J} \frac{z_{ig}}{\sigma_g^2} (\Delta^*_{ij} - \mathbf{x}_j \mathbf{y}'^g)^2. \tag{11}$$

The *E-M* algorithm maximizes (11) using two major steps. In the major *E*-Step, we compute the expected value of $\mathbf{z}_i$ given $\hat{\Delta}^*_i$ and provisional estimates for $\mathbf{y}$, $\mathbf{x}$, $\boldsymbol{\Sigma}$, and $\mathbf{w}$. $\hat{\Delta}^*_i$ is Δ^*_i with the unobserved values replaced by their expected values given $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{\mathbf{w}}$. In the major *M*-Step, we maximize (11) conditional on the newly estimated values of $\mathbf{z}$ in two minor phases: a minor *E*-Step in which we replace the nonpositive values of Δ^*_{ij} by their expected values given $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{\mathbf{z}}$ (i.e., we update $\hat{\Delta}^*_{ij}$) and a minor *M*-Step where we maximize (11) with respect to $\mathbf{y}$, $\mathbf{x}$, $\boldsymbol{\Sigma}$, and $\mathbf{w}$ conditional on the newly estimated values of $\mathbf{z}$ and Δ^*. The new estimates $\hat{\Delta}^*$, $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{\mathbf{w}}$ serve as provisional estimates for the next major *E*-Step. We keep iterating between the major *E*- and major *M*-steps until no further improvement in the likelihood function in (11) is possible. Jedidi, DeSarbo, and Ramaswamy (1993) have extended this framework to regression analyses (nonspatial analyses) for general types of censored data (although rather different stationary equations result as a function of such different models and data types). The appendix describes the particular multi-step estimation routine devised for the spatial MDS vector model, as well as criteria for model selection.[2]

3. Two Applications: Automobile Consideration Sets

A. *Mid-size Automobiles*

A major U.S. automobile manufacturer conducted personal interviews with $N = 289$ consumers who stated that they were intending to purchase a mid-size automobile within the next six months. The study was conducted in a number of automobile clinics occurring at different geographical locations in the U.S. One section of the questionnaire asked the respondent to check off from a list of ten mid-size cars, specified by this manufacturer and thought to compete in the same market segment at that time (based on prior research), which brands s/he would consider purchasing as a replacement vehicle after recalling their perceptions of expected benefits and costs of each brand. Afterwards, the respondent was asked to use a ten-point scale to indicate the intensity of their purchase consideration for the vehicles initially checked as in their respective consideration sets. The ten name-

[2] We created synthetic censored data from a known structure (a two-dimensional lattice) with error. We then performed MDPREF, MULTICLUS, and Correspondence analysis on this data, as well as the proposed methodology. Only the proposed methodology was able to recover appropriately the lattice structure in two dimensions, as well as the appropriate market segmentation. In addition, a complete Monte Carlo simulation analysis was performed with the proposed methodology to test experimentally the performance of the estimation procedure in recovering known parameters as the numbers of subjects, segments, brands, and dimensions were varied, as well as the censoring rate, amount, and type of error. Consistent fitting was obtained across the 27 full factorial design in terms of parameter recovery, data fitting, and segment membership.

plates tested were: Buick Century, Ford Taurus, Oldsmobile Cutlass Supreme, Ford Thunderbird, Chevrolet Celebrity, Honda Accord, Pontiac Grand Am, Chevrolet Corsica, Ford Tempo, and Toyota Camry. Note, based upon an inspection of the frequency distribution for the number of mid-size automobiles contained in the consideration sets for these $N = 289$ intenders, the vast majority of the respondents elicited consideration sets in the range of 2–6 automobiles from this list of ten, displaying congruence with the results from Hauser and Wernerfelt (1990). Here the Celebrity, Camry, Accord, Thunderbird, and Cutlass appear in the consideration sets of over one-half of the respondents, while the Corsica and Tempo are the clear "losers" for this particular sample.

These censored data were analyzed via the proposed latent structure tobit MDS vector model. Table 1 presents the various goodness-of-fit statistics for $G > T = 1, \ldots. 4$. As shown, the $G = 2$ segment, $T = 2$ dimension solution is selected giving rise to the minimum CAIC and BIC statistics. Figure 1 displays this solution graphically. The first (horizontal) dimension clearly distinguishes the two foreign nameplates (Accord and Camry) from the remaining eight domestic nameplates. The second (vertical) dimension relates to the cost and size of the automobile where the Thunderbird, Cutlass, Taurus, and Celebrity tend to be larger, more powerful, and more expensive automobiles (in 1988). Here, the two market segments are represented by the two vectors in the figure. The threshold values δ_g^* were estimated and are also shown in Figure 1. Nameplates that orthogonally project beyond the segment's δ_g^* value in the direction of increasing utility are predicted to be included in that segment's consideration set. Members of market segment 1 ($w_1 = .514$) consider primarily the two foreign nameplates Accord and Camry, with the remaining eight nameplates projecting near or beyond δ_1^* in the negative utility direction. Members of the second market segment ($w_2 = 0.486$) appear to consider primarily the Taurus, Cutlass, Thunderbird, and Celebrity, in addition to the Camry (not the Accord). All other nameplates project near or beyond δ_2^* in a negative utility direction.

This can be more formally examined vis á vis Table 2, which presents the model predicted probability of consideration and latent utility scores for each of the ten nameplates by market segment. As shown, the Accord and Camry have the highest positive latent utility scores for market segment 1, with correspondingly higher predicted probabilities of consideration (0.946 and 0.853). The Cutlass and Thunderbird have the highest positive latent utility scores for market segment 2, with correspondingly higher predicted probabilities of consideration (0.674 and 0.706), although the Taurus, Celebrity, and Camry also possess relatively high consideration probabilities as well (0.637, 0.619, and 0.625). The table demonstrates this consideration set composition heterogeneity by

TABLE 1

Latent Structure Censored Vector MDS Results for Mid-size Automobiles

G	T	DF	$-\ln L$	AIC	CAIC	BIC	CORR	MATCH	ENTROPY
1	1	12	2400.7	4825.5	4909.1	4897.1	0.101	0.624	–
2	1	16	2392.3	4816.7	4928.2	4912.2	0.169	0.604	0.560
2	2	25	2319.2	4688.2	4862.7*	4837.7*	0.298	0.703	0.802
3	1	20	2392.3	4824.6	4964.0	4944.0	0.168	0.604	0.620
3	2	30	2315.9	4691.9	4901.0	4871.0	0.316	0.704	0.791
3	3	38	2291.3	4658.7	4923.5	4885.5	0.342	0.727	0.870
4	1	24	2392.2	4832.5	4999.7	4975.7	0.169	0.604	0.604
4	2	35	2315.2	4700.5	4944.4	4909.4	0.322	0.707	0.800
4	3	44	2281.9	4651.8	4958.4	4914.4	0.321	0.732	0.900
4	4	51	2241.1	4584.2	4939.6	4888.6	0.355	0.754	0.962

* Denotes minimum CAIC (BIC).

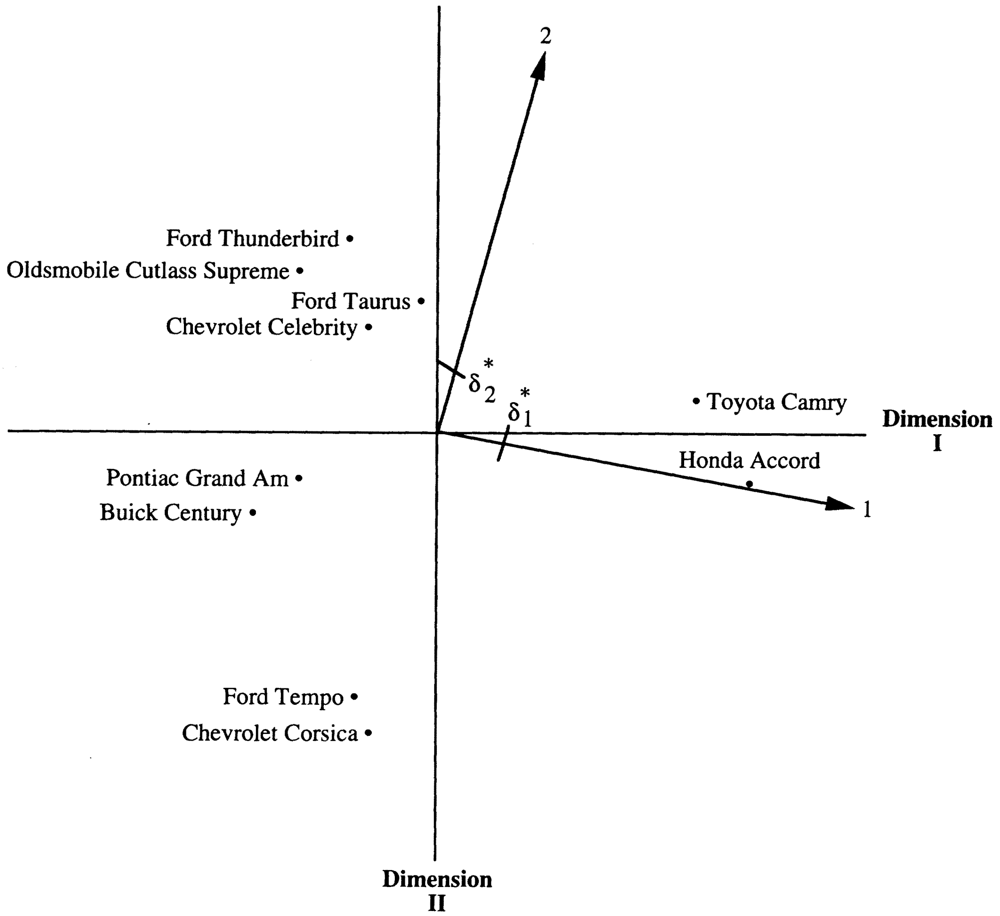

FIGURE 1. Tobit Vector MDS Solution for Midsize Cars.

market segment. Unfortunately, customer demographic or psychographic information was not available to help explain the nature of the segment differences via relating the estimated $\hat{\mathbf{P}}$ to such individual difference data.

B. *Luxury Automobiles*

This same U.S. automobile manufacturer conducted a different study in alternative geographical locations via personal interviews with $N = 240$ consumers who stated that they were intending to purchase a luxury automobile within the next six months. The use of automobile clinics to collect the data was also utilized as in the previous study. One section of the questionnaire asked the respondent to check off from a list of ten luxury cars, specified by the manufacturer and thought to compete in the same market segment at that time (based on prior research), which brands s/he would consider purchasing as a replacement vehicle after recalling their perception of expected benefits and costs of each brand. Afterwards, the respondent was asked to use the same ten-point scale to indicate the intensity of their purchase consideration for the vehicles initially checked as in their respective consideration sets. The ten nameplates tested were: Lincoln Continental, Cadillac Seville, Buick Riviera, Oldsmobile Ninety-Eight, Lincoln Town Car, Mercedes 300E, BMW 325i, Volvo 740, Jaguar XJ6, and Acura Legend. Here too, the vast majority of respondents elicited consideration sets in the range of 2–6 automobiles from the list of ten. Only the Lincoln Town Car appeared in the consideration sets of

TABLE 2

Latent Utility and Consideration Set Inclusion Probabilities by Derived Market Segment—Midsize Cars

Nameplate	Latent Utility Scores Segment 1	Segment 2
Century	−0.361	−0.224
Taurus	−0.147	0.274*
Cutlass	−0.473	0.353*
Thunderbird	−0.401	0.423*
Celebrity	−0.277	0.236*
Accord	1.098*	0.131
Grand AM	−0.293	−0.185
Corsica	0.053	−0.811
Tempo	0.045	−0.686
Camry	0.715*	0.250*

Nameplate	Consideration Probabilities Segment 1	Segment 2
Century	0.298	0.387
Taurus	0.415	0.637*
Cutlass	0.244	0.674*
Thunderbird	0.278	0.706*
Celebrity	0.342	0.619*
Accord	0.946*	0.567
Grand AM	0.333	0.407
Corsica	0.489	0.150
Tempo	0.426	0.190
Camry	0.853*	0.625*

* Designates nameplates predicted to be in market segment's consideration set.

over one-half of the respondents, followed in popularity by the Acura Legend and Lincoln Continental. The BMW 325i was the lowest considered luxury car from the list (considered by only 44 of the 240 respondents).

Table 3 presents the statistical summary of the results of applying the proposed tobit vector MDS procedure to this censored data. Here too, the minimum BIC and CAIC occur at $G = T = 2$. Figure 2 presents this solution graphically. Here, $w_1 = 0.373$ and

TABLE 3

Latent Structure Censored Vector MDS Results for Luxury Automobiles

G	T	DF	$\ln L$	AIC	CAIC	BIC	CORR	MATCH	ENTROPY
1	1	12	−4086.2	8196.5	8277.9	8265.9	0.010	0.675	–
2	1	16	−4086.2	8204.5	8313.0	8297.0	0.010	0.657	0.000
2	2	25	−3992.6	8035.3	8204.9*	8179.9*	0.117	0.699	0.985
3	1	20	−4086.2	8212.5	8348.0	8328.0	0.010	0.657	0.006
3	2	30	−3983.6	8027.4	8230.9	8200.9	0.083	0.709	0.962
3	3	38	−3960.3	7996.7	8254.5	8216.5	0.152	0.728	0.992
4	1	24	−4086.2	8220.5	8383.3	8359.3	0.010	0.657	0.024
4	2	35	−3982.6	8035.3	8272.7	8237.7	0.118	0.715	0.697
4	3	44	−3951.0	7990.0	8288.4	8244.4	0.149	0.734	0.996
4	4	51	−3940.4	7982.8	8328.7	8277.7	0.154	0.742	0.999

* Denotes minimum CAIC (BIC).

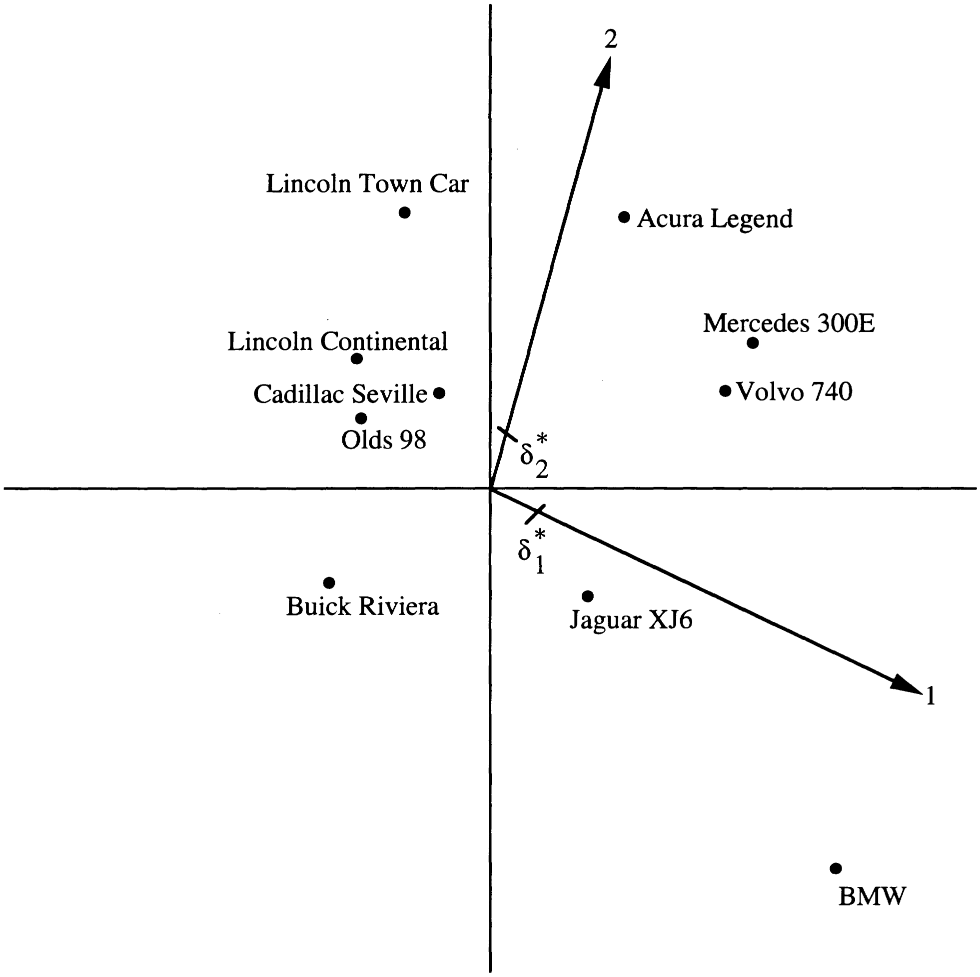

FIGURE 2. Tobit Vector MDS Solution for Luxury Cars.

$w_2 = 0.627$. The horizontal axis separates the domestic cars from the imports. The vertical axis describes a traditional vs. sporty dimension. Here, the smaller first market segment (37.3%) primarily considers the BMW, Mercedes, Volvo, and Jaguar—an import oriented segment. The larger second segment (62.7%) considers the Town Car, Acura, Mercedes, Continental, Seville, and Volvo. This can be verified via the predicted latent utility scores and consideration probabilities reported in Table 4.

4. Discussion

A. *Managerial Implications*

The results obtained from the proposed MDS methodology provides the basis for a simultaneous approach for segmentation and positioning based on consideration set composition. Figures 1 and 2 can be gainfully employed in the creation of a marketing strategy to benefit a particular nameplate. These joint space maps help managers understand who they are competing against and how this competition varies by segment. With appropriate supplementary data (e.g., demographics and psychographics), a descriptive profile of the resulting market segments can be derived by relating such information to

TABLE 4
Latent Utility and Consideration Set Inclusion Probabilities by Derived Market Segment—Luxury Cars

Nameplate	Latent Utility Scores Segment 1	Segment 2
Continental	−5.228	2.472*
Seville	−4.104	1.670*
Riviera	−0.092	−3.116
Olds 98	−3.495	0.567*
Town Car	−10.038	6.718*
Mercedes	2.764*	4.437*
BMW	29.847*	−9.906
Volvo	6.397*	2.541*
XJ6	6.785*	−1.982
Acura	−2.021	6.710*
Nameplate	**Consideration Probabilities Segment 1**	**Segment 2**
Continental	0.234	0.618*
Seville	0.284	0.580*
Riviera	0.495	0.353
Olds 98	0.314	0.527*
Town Car	0.082	0.792*
Mercedes	0.649*	0.704*
BMW	1.000*	0.115
Volvo	0.813*	0.621*
XJ6	0.827*	0.405
Acura	0.389	0.791*

* Designates nameplates predicted to be in market segment's consideration set.

the estimated posterior probabilities of segment membership (such data were not available in this study) via logistic regression (or direct reparameterization of the prior/mixing proportions as in Dayton and Macready (1988)). Such descriptions can be utilized for segment targeting. In addition, the dimensions provide a theoretical mechanism for discussing repositioning strategy by market segment. For example, given a target segment, the resulting joint space maps reveal the basis on which to attempt to beneficially move brands in a designated direction. Movement in the space vis a vis repositioning can be modeled either by property fitting designated attributes onto the space via regression methods, or by extending the procedure to include linear restrictions on the brand coordinates as proposed by DeSarbo and Rao (1986). With such calibration, one could then perform policy simulations, as well as optimal positionings, by market segment.[3]

B. *Limitations*

There are several obvious limitations involving the use of the proposed methodology. As mentioned in the text, the proposed MDS based methodology is a descriptive procedure, not a normative one. It does not provide the basis of modeling the dynamic process by which consideration sets are actually formed. Rather, it provides a descriptive paramorphic representation of consideration sets and resulting market segments at any

[3] A number of competing methods, including MDPREF, MULTICLUS, PRINCALS, and correspondence analysis, were also utilized for both commercial applications. See the working paper version of this manuscript for details.

given static point in time. Obviously, given the use of such a stochastic framework, the distributional assumptions of the proposed model must be adhered to. In this case, the three major assumptions concern the use of normal mixtures, conditional independence, and the vector model. The use of normal mixtures in this context resembles a semi-parametric approach for estimating an unknown distribution. As such, it is reasonably robust except for cases where the segment centroids are not well separated as we ascertained in our preliminary Monte Carlo simulations. Conditional independence is assumed with all such latent structure models and appears reasonable given the market segmentation interpretation provided with these support points. The most serious limitation of the three appears to revolve around the use of the vector (vs. unfolding) model. Given the "more the better" utility structure assumed by the procedure, applications involving a single peaked utility structure may be problematic. Thus, extension of this MDS procedure to accommodate ideal points is an area for future research.

C. *Future Research*

We have mentioned a number of gainful areas for future research above including the accommodation of linear restrictions on the brand coordinates, incorporating Dayton and Macready's (1988) reparameterization of the mixing proportions, and the extension to an ideal point framework. Additional applications with validation samples and individual/brand background data are also needed. More Monte Carlo testing and comparison with traditional methods is also required. Finally, further research on model selection criteria in such latent structure models is also needed.[4]

Acknowledgments. The authors wish to thank John Hauser, Gary Lilien, and two anonymous reviewers for their insightful comments on this manuscript.

[4] This paper was received December 31, 1993, and has been with the authors 3 months for 2 revisions. Processed by Gary L. Lilien.

Appendix

1. *The "Major" E-Step*

Using Bayes theorem, we can show that:

$$E(z_{ig} | \hat{\Delta}_i^*, \hat{\mathbf{y}}, \hat{\mathbf{x}}, \hat{\Sigma}, \hat{\mathbf{w}}) = \frac{\hat{w}_g f(\hat{\Delta}_{ij}^* | \hat{\mathbf{y}}^g, \hat{\mathbf{x}}, \hat{\sigma}_g^2)}{\sum_{k=1}^G w_k f(\hat{\Delta}_{ij}^* | \hat{\mathbf{y}}^k, \hat{\mathbf{x}}, \hat{\sigma}_k^2)}, \tag{A-1}$$

which is equal to $\hat{P}_{ig}$ given in equation (7). Therefore, the conditional expectation of $\ln L_c(\cdot)$ in (11) with respect to $\mathbf{z}_i$ evaluated at the provisional estimates $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$, $\hat{\Sigma}$, $\hat{\mathbf{w}}$, and $\hat{\Delta}^*$ (ignoring the constant term without loss of generality) is:

$$E_{\mathbf{z}}(\ln L_c | \hat{\Delta}^*, \hat{\mathbf{y}}, \hat{\mathbf{x}}, \hat{\Sigma}, \hat{\mathbf{w}}) = -\frac{1}{2} \sum_{i=1}^{I} \sum_{g=1}^{G} \sum_{j=1}^{J} \hat{P}_{ig} \ln(\sigma_g^2) + \sum_{i=1}^{I} \sum_{g=1}^{G} \hat{P}_{ig} \ln(w_g) - \frac{1}{2} \sum_{g=1}^{G} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\hat{P}_{ig}}{\sigma_g^2} (\hat{\Delta}_{ij}^* - \mathbf{x}_j \mathbf{y}'^g)^2. \tag{A-2}$$

Thus, the expectation phase amounts to replacing the nonobserved data $\mathbf{z} = ((z_{ig}))$ in (A-2) by their estimated conditional expectation, $\hat{\mathbf{P}} = ((\hat{P}_{ig}))$.

2. *The "Major" M-Step*

In this step we maximize $E_{\mathbf{z}}$ ($\ln L_c$) in (A-2) with respect to $\mathbf{y}$, $\mathbf{x}$, Σ, and $\mathbf{w}$ subject to the restrictions $w_g > 0$ and $\sum_{g=1}^G w_g = 1$ and conditional on the new provisional estimates of z_{ig}, $\hat{P}_{ig}$. The structure of this maximization problem lends itself to the application of an *E-M* procedure nested within the major *E-M* algorithm since Δ^* is partially unobserved. In the *E*-Step of this minor *E-M* procedure, we compute the conditional expectation of $E_{\mathbf{z}}(\ln L_c)$ in (A-2) with respect to Δ^* given the observed data (Δ) and provisional estimates for $\mathbf{y}$, $\mathbf{x}$, Σ, $\mathbf{w}$, and $\mathbf{z}$ parameters. In the "Minor" *M*-Step, we maximize such conditional expectation with respect to $\mathbf{y}$, $\mathbf{x}$, Σ, and $\mathbf{w}$, given $\hat{\mathbf{z}}$ and $\hat{\Delta}^*$.

3. *The "Minor" E-Step*

The conditional expectation of (A-2) with respect to $\boldsymbol{\Delta}^*$ is:

$$E_{\Delta}[E_z(\ln L_c|\boldsymbol{\Delta}^*, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \mathbf{w})|\boldsymbol{\Delta}] = \sum_{i=1}^{I}\sum_{g=1}^{G} z_{ig}\ln(w_g)$$
$$-\frac{1}{2}\sum_{i=1}^{I}\sum_{g=1}^{G}\sum_{j=1}^{J} z_{ig}\left[\ln(\sigma_g^2) + \frac{1}{\sigma_g^2}E((\Delta_{ij}^* - \mathbf{x}_j\mathbf{y}'^g)^2|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})\right]. \quad \text{(A-3)}$$

It can be shown that:

$$E[(\Delta_{ij}^* - \mathbf{x}_j\mathbf{y}'^g)^2|\mathbf{z}_i, \mathbf{y}^g, \mathbf{x}, \sigma_g^2, \Delta_{ij}] = \begin{cases} (\Delta_{ij} - \mathbf{x}_j\mathbf{y}'^g)^2 & \text{iff} \quad \Delta_{ij} > 0 \\ [E(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) - \mathbf{x}_j\mathbf{y}'^g]^2 & \\ \quad + \operatorname{Var}(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) & \text{iff} \quad \Delta_{ij} = 0, \end{cases} \quad \text{(A-4)}$$

where:

$$E(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) = \sum_{g=1}^{G} z_{ig}\left[\mathbf{x}_j\mathbf{y}^g - \frac{\sigma_g\phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)}{1 - \Phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)}\right], \quad \text{(A-5)}$$

and

$$\operatorname{Var}(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) = \sum_{g=1}^{G} z_{ig}\left[\sigma_g^2 + \frac{\mathbf{x}_j\mathbf{y}^{g\prime}\sigma_g\phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)}{1 - \Phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)} - \left(\frac{\sigma_g\phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)}{1 - \Phi(\mathbf{x}_j\mathbf{y}'^g/\sigma_g)}\right)^2\right]. \quad \text{(A-6)}$$

Let:

$$\hat{\Delta}_{ij}^* = \begin{cases} \Delta_{ij} & \text{iff} \quad \Delta_{ij} > 0 \\ E(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) & \text{iff} \quad \Delta_{ij} = 0. \end{cases} \quad \text{(A-7)}$$

Hence, we can rewrite equation (A-3) as:

$$Q = E_{\Delta^*}[E_z \ln L_c(\cdot)] = \sum_{g=1}^{G}\sum_{i=1}^{I} z_{ig}\ln(w_g)$$
$$-\frac{1}{2}\sum_{i=1}^{I}\sum_{g=1}^{G}\sum_{j=1}^{J} z_{ig}\left[\ln(\sigma_g^2) + \frac{1}{\sigma_g^2}(\hat{\Delta}_{ij}^* - \mathbf{x}_j\mathbf{y}'^g)^2 + \frac{(1-\lambda_{ij})}{\sigma_g^2}\operatorname{Var}(\Delta_{ij}^*|\mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0)\right], \quad \text{(A-8)}$$

where λ_{ij} is as defined in (2). Thus, this minor E-Step amounts to replacing the non-positive values of $\boldsymbol{\Delta}^*$ by their estimated expected values $\hat{\boldsymbol{\Delta}}^*$.

4. *The "Minor" M-Step*

In this step, we maximize Q with respect to $\mathbf{y}$, $\mathbf{x}$, $\boldsymbol{\Sigma}$ and $\mathbf{w}$ subject to the restrictions $w_g > 0$ and $\sum_{g=1}^{G} w_g = 1$, and conditional on the new provisional estimates of z_{ig} and the non-positive values of $\boldsymbol{\Delta}^*$. To estimate $\mathbf{w}$, it suffices to maximize the augmented function:

$$\sum_{i=1}^{I}\sum_{g=1}^{G}\hat{P}_{ig}\ln(w_g) + \theta\left(\sum_{g=1}^{G} w_g - 1\right), \quad \text{(A-9)}$$

where θ denotes a Lagrangian multiplier. By differentiating (A-9) with respect to θ and w_g and setting the derivatives equal to zero, we can show that:

$$\hat{w}_g = \sum_{i=1}^{I}\frac{\hat{P}_{ig}}{I}. \quad \text{(A-10)}$$

Given $\hat{\mathbf{P}} = ((\hat{\mathbf{P}}_{ig}))$, $\hat{\boldsymbol{\Delta}}^*$, and $\hat{\mathbf{w}}$, the maximization of Q with respect to $\mathbf{y}$, $\mathbf{x}$, and $\boldsymbol{\Sigma}$ requires the computation of their partial derivatives. These are given below:

$$\frac{\partial Q}{\partial \mathbf{y}^g} = \sum_{i=1}^{I}\frac{z_{ig}}{\sigma_g^2}(\mathbf{x}'\hat{\Delta}_j^*) - \left[\sum_{i=1}^{I}\frac{z_{ig}}{\sigma_g^2}(\mathbf{x}'\mathbf{x})\right]\mathbf{y}^g, \quad \text{(A-11)}$$

$$\frac{\partial Q}{\partial \mathbf{x}} = \sum_{i=1}^{I}\sum_{g=1}^{G}\frac{z_{ig}}{\sigma_g^2}(\hat{\Delta}_i^* - \mathbf{x}\mathbf{y}'^g)\mathbf{y}'^g, \quad \text{(A-12)}$$

$$\frac{\partial Q}{\partial \sigma_g^2} = -\frac{1}{2} \sum_{j=1}^{J} \sum_{i=1}^{I} z_{ig} \left[\frac{1}{\sigma_g^2} - \frac{1}{\sigma_g^4} (\hat{\Delta}_{ij}^* - \mathbf{x}_j \mathbf{y}'^g)^2 - \frac{1}{\sigma_g^4} (1 - \lambda_{ij}) \operatorname{Var}(\Delta_{ij}^* \,|\, \mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0) \right]. \quad \text{(A-13)}$$

Equating equations (A-11), (A-12), and (A-13) to zero and solving for $\mathbf{y}^g$, $\mathbf{x}$, and σ_g^2, respectively, produces the following closed form or analytical solutions estimates for these parameters:

$$\mathbf{y}^g = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' \left(\sum_{i=1}^{I} z_{ig} \hat{\Delta}^* \right) \Big/ \sum_{i=1}^{I} z_{ig}, \quad \text{(A-14)}$$

$$\mathbf{x} = \left(\sum_{i=1}^{I} \sum_{g=1}^{G} z_{ig} (\mathbf{y}^g \hat{\Delta}^{*\prime}) \right) \left(\sum_{i=1}^{I} \sum_{g=1}^{G} z_{ig} \mathbf{y}^g \mathbf{y}'^g \right)^{-1}, \quad \text{(A-15)}$$

and

$$\sigma_g^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} z_{ig} [(\hat{\Delta}_{ij}^* - \mathbf{x}_j \mathbf{y}'^g)^2 + (1 - \lambda_{ij}) \operatorname{Var}(\Delta_{ij}^* \,|\, \mathbf{z}_i, \mathbf{y}, \mathbf{x}, \boldsymbol{\Sigma}, \Delta_{ij} = 0)] \Big/ \sum_{i=1}^{I} \sum_{j=1}^{J} z_{ig}. \quad \text{(A-16)}$$

It is important to note from equation (A-15) that $\mathbf{x}$ is estimable if $G \geq T$. Also solutions with dimensionality $T > G$ will not improve the log-likelihood function over those corresponding to $T = G$ since G vectors can be fitted perfectly in a G dimensional space.

The estimates obtained in the "Major" M-Step serve as the new provisional estimates for the next "Major" E-Step iteration. We keep alternating between the "Major" E- and "Major" M-Steps until convergence. Once convergence occurs, we obtain final estimates for $\mathbf{y}$, $\mathbf{x}$, $\boldsymbol{\Sigma}$, and $\mathbf{w}$. One can now compute the asymptotic variance-covariance matrix, $\mathbf{H}$, of the segment level censored vector MDS model parameter estimates using the asymptotic efficiency of the MLE (Berndt, Hall, Hall, and Hausman 1974) via:

$$\mathbf{H} = \frac{1}{I} \left[\sum_{i=1}^{I} \left(\frac{\partial \ln L_i}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln L_i}{\partial \boldsymbol{\theta}} \right)' \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}^{-1}, \quad \text{(A-17)}$$

where $\ln L_i$ is Equation (11) evaluated for observation i only, $\boldsymbol{\theta}$ is a vector stacking all the free parameters of the model, and $\hat{\boldsymbol{\theta}}$ contains their corresponding maximum likelihood estimates.

5. *Model Selection*

When estimating the segment level censored vector MDS model, the number of dimensions T and the number of market segments G are usually unknown a priori and therefore has to be inferred from the data. We make such inference by running the estimation procedure for varying number of dimensions (T) and varying number of segments (G), and select the T and G that best represent the data. As in MULTICLUS (DeSarbo, Howard, and Jedidi 1991), solutions with $T > G$ are not identified. Since the regularity conditions for the validity of the conventional tests based on the likelihood ratio statistics (e.g., the chi-squared test) are violated in our case, we rely on several information criteria and goodness-of-fit measures in the selection of the appropriate G and T.

Consider first the information criteria. One approach is to choose G and T to minimize the Akaike (1974) Information Criterion (AIC) defined by:

$$\text{AIC}_G = -2 \ln L_G + 2M_G, \quad \text{(A-18)}$$

where M_G is the effective number of parameters estimated in a G-segment, T-dimensional solution taking into account the rotation and scale indeterminacies:

$$M_G = GT + JT + 2K - T^2. \quad \text{(A-19)}$$

One problem with the AIC is that it fails to sufficiently penalize overparametrization. Bozdogan (1987) suggests that G be chosen to minimize the consistent AIC defined by:

$$\text{CAIC}_G = -2 \ln L_G + M_G(\ln I + 1). \quad \text{(A-20)}$$

Another measure that is similar in nature to this CAIC criterion in appropriately penalizing overparametrization is the Baysian Information Criterion (Schwartz 1978) defined as:

$$\text{BIC}_G = -2 \ln L_G + M_G \ln I. \quad \text{(A-21)}$$

We shall use both the CAIC and BIC criteria because they are less likely to lead to the choice of an excessively larger number of segments and dimensions. It should be noted here that the AIC, CAIC, and BIC are heuristics.

In addition to the information criteria, we also employ several goodness of fit measures to support the selection of the appropriate G and T. The first measure is a correlation measure (CORR) assessing the degree of fit between Δ and $\hat{\Delta}^*$. The second is a matching coefficient between the dichotomized Δ and $\hat{\Delta}^*$. It is defined as:

$$\text{MATCH} = \frac{N(0, -) + N(+, +)}{I}, \quad \text{(A-22)}$$

where $N(0, -)$ denotes the number of instances of zero values in Δ and negative values in $\hat{\Delta}^*$, and $N(+, +)$ is the number of positive values in both Δ and $\hat{\Delta}^*$. The last measure we output is the correlation between Δ and the probabilities of consideration, $P(\Delta_{ij}^* > 0)$. Finally to assess the separation of the market segments, we use an Entropy-based measure defined as:

$$E_G = 1 - \left[\sum_i \sum_g - \hat{P}_{ig}\right] \ln \hat{P}_{ig}/I \ln G. \tag{A-23}$$

A value of E_G close to zero should be cause for concern as it indicates that the latent class centroids are not sufficiently separated (see DeSarbo, Wedel, Vriens, and Ramaswamy 1992).

References

Akaike, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, 6, 716–723.

Alba, J. W. and J. W. Huthchinson (1987), "Dimensions of Consumer Expertise," "Journal of Consumer Expertise," *Journal of Consumer Research*, 13, 411–454.

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

Baker, F. B. and L. J. Hubert (1976), "A Graph-Thoeretic Approach to Goodness-of-Fit in Complete Link Hierarchical Clustering," *Journal of the American Statistical Association*, 71, 810–878.

Berndt, E. K., B. H. Hall, and J. A. Hausman (1974), "Estimation and Inference in Non-Linear Structural Models," *Annals of Economic and Social Measurement*, 3, 653–665.

Bettman, J. (1979), *An Information Processing Theory of Consumer Choice*, Reading, MA: Addison-Wesley.

———, E. J. Johnson, and J. W. Payne (1989), "Consumer Decision Making," in T. S. Robertson and H. H. Kassarjian (Eds.), *Handbook of Consumer Behavior*, Prentice Hall: Englewood Cliffs, NJ.

Bozdogan, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, 52, 345–370.

Brisoux, Jacques E. and Emmanuel Cheron (1990), "Brand Categorization and Product Involvement," in *Advances in Consumer Research*, Marvin E. Goldberg and Gerald Gorn (Eds.), Provo, UT: Association for Consumer Research, 101–109.

——— and Michel Laroche (1980), "A Proposed Consumer Strategy of Simplification for Categorizing Brands," in *Evolving Marketing Thought for* 1980, John H. Summey and Ronald D. Taylor (Eds.), Carbondale, IL: Southern Marketing Association, 112–114.

——— and Michel Laroche (1981), "Evoked Set Formation and Composition: An Empirical Investigation Under a Routinized Response Behavior Situation," in *Advances in Consumer Research*, Kent B. Monroe (Ed.), Provo, UT: Association for Consumer Research, 357–361.

Brown, J. J. and Wildt, A. R. (1992), "Consideration Set Measurement," *Journal of the Academy of Marketing Science*, 3, 235–243.

Campbell, Brian M. (1969), "The Existence and Determinants of Evoked Set in Brand Choice Behavior." Dissertation, Columbia University, New York.

Carroll, J. D. (1980), "Models and Methods for Multidimensional Analysis of Preferential Choice Data," in *Similarity and Choice*, E. D. Lantermann and H. Feger (Eds.), Bern: Hans Huber.

Chang, W. C. (1983), "On Using Principal Components before Separating a Mixture of Two Multivariate Normal Distributions," *Applied Statistics*, 32, 267–275.

Chatterjee, R. and W. S. DeSarbo (1992), "Accommodating the Effects of Brand Unfamiliarity in the Multidimensional Scaling of Preference Data," *Marketing Letters*, 3, 85–99.

Cooper, L. G. (1983), "A Review of Multidimensional Scaling in Marketing Research," *Applied Psychological Measurement*, 7, 427–450.

Crowley, Aye E. and John H. Williams (1991), "An Information Theoretic Approach to Understanding the Consideration Set/Awareness Set Proportion," in *Advances in Consumer Research*, Rebecca H. Holman and Michael R. Solomon (Eds.), Provo, UT: Association for Consumer Research, 780–787.

Dayton, C. M. and G. B. Macready (1988), "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Society*, 83, 173–178.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society*, *B*, 39, 1–22.

DeSarbo, W. S., A. K. Manrai, and L. A. Manrai (1994), "Latent Class Multidimensional Scaling: A Review of Recent Development in the Marketing and Psychometric Literature," *Handbook of Marketing Research*, R. Bagozzi (Ed.), Blackwell Publishing, London, UK.

———, M. Wedel, V. Ramaswamy, and M. Vriens (1992), "Latent Class Metric Conjoint Analysis," *Marketing Letters*, 3, 273–288.

———, D. J. Howard, and K. Jedidi (1991), "MULTICLUS: A New Method for Simultaneously Performing Multidimensional Scaling and Cluster Analysis," *Psychometrika*, 56, 121–136.

——— and V. R. Rao (1986), "A New Constrained Unfolding Model for Product Positioning," *Marketing Science*, 5, 1–19.

DeSarbo, W. S., and V. R. Rao (1984), "GENFOLD2: A Set of Models and Algorithms for the General Unfolding Analysis of Preference/Dominance Data," *Journal of Classification* 1, 146–185.

Gensch, Dennis (1987), "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Science*, 6, 223–231.

Green, P. E., D. S. Tull, and G. Albaum (1988), *Research for Marketing Decisions, 5th edition*, Prentice Hall: Englewood Cliffs, NJ.

Greenacre, Michael, J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press: London.

Greene, W. H. (1993), *Econometric Analysis* (2nd Ed.), Macmillan Publishing: New York.

Hauser, John R. and B. Wernerfelt (1990), "An Evaluation Cost Model of Evoked Sets," *Journal of Consumer Research*, 16, 393–408.

Heath, Timothy B. and Subimal Chatterjee (1991), "How Entrants Affect Multiple Brands: A Dual Attraction Mechanism," in *Advances in Consumer Research*, Rebecca H. Holman and Michael R. Solomon (Eds.), Provo, UT: Association for Consumer Research, 768–772.

Hoffman, D. L. and G. R. Franke (1986), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research," *Journal of Marketing Research*, 23, 213–227.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*, New York: John Wiley and Sons.

Jedidi, K., V. Ramaswamy, and W. S. DeSarbo (1993), "A Maximum Likelihood Method for Latent Class Regression Involving a Censored Dependent Variable," *Psychometrika*, 58, 375–394.

Johnson, E. and R. Meyer (1984), "Compensatory Choice Models of Non-Compensatory Processes: The Effect of Varying Context," *Journal of Consumer Research*, 11, 528–541.

Klenosky, D. B. and A. J. Rethang (1989), "The Formation of Consumer Choice Sets," in *Advances in Consumer Research*, M. Houston (Ed.), ACR, 13–17.

Ling, R. F. (1971), *Cluster Analysis*, Ann Arbor, MI: University Microfilms, No. 71-22356.

Maddala, G. S. (1992), *Introduction to Econometrics* (2nd Ed.), Macmillan: New York.

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

Muthen, B. O. (1989), "Tobit Factor Analysis," *British Journal of Mathematical and Statistical Psychology*, 42, 241–250.

Narayana, Chem L. and Rom J. Markin (1975), "Consumer Behavior and Product Performance: An Alternative Conceptualization," *Journal of Marketing*, 39, 1–6.

Nedungadi, Prakash (1987), *Formation and Use of a Consideration Set: Implications for Marketing and Research on Consumer Choice*. Gainesville, FL: University of Florida, Unpublished Ph.D. Dissertation.

——— (1990), "Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations." *Journal of Consumer Research*, 17 (December), 263–276.

——— and V. Kanetkar (1992), "Incorporating Consideration Sets into Models of Brand Choice." *Advances in Consumer Research*, 19, 251–252.

Ratneshwar, S. and Allan D. Shocker (1991), "The Role of Usage Context in Product Category Structures," *Journal of Marketing Research*, 28, 3.

Roberts, John (1989), "A Grounded Model of Consideration Set Size and Composition," in *Advances in Consumer Research*, Thomas K. Srutt (Ed.), Provo, UT: Association for Consumer Research, 749–757.

Roberts, John H. and James M. Lattin (1991), "Development and Testing of a Model of Consideration Set Formation," *Journal of Marketing Research*, 28, 429–440.

Schmee J. and G. J. Hahn (1979), "A Simple Method for Regression Analysis with Censored Data," *Technometrics*, 21, 417–432.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Shocker, Allan D., Moshe Ben-Akia, Bruno Boccara, and Prakash Nedungadi (1991), "Consideration Set Influences on Customer Decision-Making and Choice: Issues, Models, and Suggestions," *Marketing Letters*, 2, 181–198.

Silk, Alvin J. and Glen L. Urban (1978), "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15, 171–191.

Stigler, G. (1961), "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

Titterington, D. M., A. F. Smith, and W. E. Makov (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrika*, 26, 24–36.

Urban, G. L. and J. R. Hauser (1993), *Design and Marketing of New Products* (2nd Ed.), Clifton Heights, N.J.: Prentice-Hall.

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

Yakowitz, S. J., and J. D. Spragins (1968), "On the Identifiability of Finite Mixtures," *Annals of Mathematical Statistics*, 39, 209–214.