Donald R. Lehmann
Columbia University

William L. Moore
University of Utah

This article proposes a model that nests both a strict tree model and the Luce choice model. The multiplicative formulation allows for easy estimation using least-squares procedures. The model is shown to be more parsimonious than the hierarchical elimination method and in a small illustration, to significantly out-perform Luce in predicting soft-drink preferences.

Introduction

Traditionally most preference and choice models in marketing have assumed the property of simple scalability where the relative preference or choice probability of an object in a choice set is a function of a scale value that represents that object's desirability. This is true of both models that assume that this scale value is a function of an overall gestalt (Luce, 1959; Thurstone, 1927; Zajonc, 1980, 1984) and ones that assume that the scale value is a function of the alternative's underlying attributes (McFadden, 1974). Simply scalable models assume that the relative probability of choosing between two alternatives is unaffected by the presence of other alternatives, a context invariant property referred to as independence from irrelevant alternatives.

However, it has been shown that in many instances the introduction of a new alternative has a greater effect on the choice probabilities of the alternatives with which it is most similar (e.g., Debreu, 1960; Restle, 1961; Becker et al., 1963; Rumelhart and Greeno, 1971; Luce, 1977; Batsell and Polking, 1985). For example, the introduction of a new diet drink might be expected to have a greater impact on the other diet drinks than it does on non-diet drinks. Sometimes, this is referred to as the similarity effect. It can be illustrated by the red bus–blue bus problem. Here it is assumed that if a person is indifferent between driving a car and taking

---

Address correspondence to: Donald Lehmann, Graduate School of Business, Columbia University, New York, NY 10027.

The authors wish to thank Lee G. Cooper, Dennis Gensch, Rashi Glazer, Sunil Gupta, Joel Huber, Wes Hutchinson, Jordan Louviere, and Edgar A. Pessemier for comments on a previous draft.

a blue bus to work, s/he would have a 50% probability of taking each. If a second bus line is introduced that is identical to the first except that its buses are colored red, one would assume that the person would be indifferent between either color of bus and also between taking the second bus and driving. A simply scalable model would predict that each form of transportation would be used approximately one third of the time. However, common sense would suggest that the probability of driving a car would not decrease by the addition of a second bus line that offered nothing over the first, and choice probabilities for the car and each bus would be approximately 0.50, 0.25, and 0.25, respectively.

A second type of counter example to simply scalable models has been termed the dominance effect. It can be illustrated with the following example. Assume a person is indifferent between a free trip to Paris and a free trip to London. If a third alternative of a free trip to Paris plus one dollar is added to the choice set, a person would probably be indifferent between this new alternative and the trip to London (as the additional dollar would not have an impact). Given these two pieces of information, a simply scalable model would predict that the person would be indifferent between a trip to Paris and the trip to Paris plus one dollar; however, s/he would have a high probability of choosing the new alternative over the original (and equivalent) trip to Paris.

A number of choice models have been developed to model these choice set effects. These include Tversky's (1972) elimination by aspects (EBA), the more parsimonious hierarchical elimination method (HEM), and elimination by tree (EBT) developed by Tversky and Sattath (1979), and the generalized extreme value models, sometimes referred to as nested logit models (McFadden, 1978, 1981). Unfortunately, these approaches lead to models that are relatively difficult to estimate. This article presents a model (referred to as AT for augumented tree) that assumes that choice is based on a combination of a simply scalable model and a particular hierarchical choice model. The model nests both the simply scalable model and that hierarchical model. Furthermore, due to its functional form and parsimony, it is easy to estimate at the individual level.

This article proceeds as follows. First, a brief description is given of approaches to modeling choice hierarchically. Next, the proposed model is presented and its properties, relation to HEM, and estimation are described in some detail. An illustrative application is then presented.

Hierarchical Models

One of the most comprehensive and widely recognized attempts to deal with data where the relative probability of choosing between two alternatives depends on a noncompensatory process is Tversky's EBA model (1972). In EBA, an individual selects an aspect (i.e., a feature or an attribute) with probability proportional to its importance or "measure" and eliminates all alternatives not possessing that feature or not exceeding a minimum level on that attribute. This procedure continues until one alternative is left. While elegant, this approach requires estimation of a large number of parameters ($2^n - 3$, where n is the number of alternatives) and is of limited practical value.

In order to reduce the number of parameters to be estimated, Tversky and Sattath (1979) developed a constrained version of EBA called elimination by tree

(EBT), which assumed that the pattern of shared aspects resulted in a hierarchical partitioning of the brands. The choice process in EBT was the same as that in EBA. They also developed the HEM, which incorporated a fixed sequence of aspect selection and happened to be mathematically equivalent to EBT. Both of these models require that $2n - 3$ free parameters be estimated where n is the number of alternatives. Although they are more parsimonious than EBA, their mathematical structure leads to nontrivial estimation problems (Moore et al., 1986). The advantage of hierarchical models is their ability to handle the similarity and dominance effects.

Other Approaches

One attempt to circumvent the independence of irrelevant alternatives property was developed by Batsell and Polking (1985). This approach augments the Luce model by first adding terms that account for the joint presence of each pair of items in the choice set, then each group of 3 items, etc., in a nested framework that examines the impact of "interactions" of items on choice probabilities. This approach is more flexible than AT but pays for the flexibility by having more parameters to estimate and not producing an explanation for the results.

Another approach for dealing with non-simply scalable preferences has been nested logit models. Recent applications include those of Dubin (1986), Buckley (1988), and Moore and Lehmann (1989). These models are viable alternatives to the AT model presented here but require a more complex estimation procedure and implicitly assume a relatively more complex choice process on the part of consumers. In addition, most of these methods are applied at the aggregate rather than individual level.

Recently, Jain and Bass (1989) developed an extended logit model that accounts for similarity effects by making the coefficients in a simple logit model of attribute levels depend on the items in the choice set, so then the basic logit model:

$$V(i) = \Sigma\, B_l X_{il}$$

becomes

$$V(i) = \Sigma\, (B_l + B_l^k Z_k + B_l^{ks} Z_{ks}) X_{il}$$

where $Z_k = 1$ if k is in the choice set, 0 otherwise; and $Z_{ks} = 1$ if k and s are in the choice set, 0 otherwise. This extension, which includes the pairwise interaction term used by Louviere and Woodworth (1983), retains the explanatory basis of attributes but pays for this by requiring fairly extensive data and by potentially generating coefficient values *(B_l^k's, B_l^{ks}'s), which are nonintuitive.*

The AT Model

In Luce's model, the probability of choosing alternative X_i, in a comparison with X_j, is given by:

$$\frac{P(X_i\text{: } X_i, X_j)}{P(X_j\text{: } X_i, X_j)} = \frac{\exp v_i}{\exp v_j}$$

where v_i and v_j are single values representing the relative worth of the 2 alternatives. It is possible to directly incorporate attributes in this structure by allowing v_i to be a weighted (by importances) combination of attributes, i.e., $v_i = \Sigma B_k X_{ik}$, leading to the logit model.

The model presented here assumes that individuals engage in a top-down processing strategy for eliminating alternatives. Top-down processing of choice tasks involving multiple items from a product class has been supported theoretically and empirically by Howard (1977), Johnson (1989), and Park and Smith (1989). The model (hereafter referred to as AT) assumes that the value of an alternative X_i in a paired comparison of X_i and X_j is:

$$\text{Value } (X_i{:}X_i,X_j) = \exp a_i \exp v_i = \exp(a_i + v_i) \tag{1}$$

where a_i = value of the *most important* aspect (attribute) possessed by X_i and not X_j, and v_i = value of alternative X_i which is not contained in the attributes of the model (sometimes referred to as brand image, equity or goodwill). Thus, the AT model is equal to the Luce model with an addition of an a_i term. From Equation (1) it immediately follows that the relative probability of choosing X_i over X_j is:

$$\frac{P(X_i{:}X_i,X_j)}{P(X_j{:}X_i,X_j)} = \frac{P_i}{P_j} = \frac{\exp(a_i + v_i)}{\exp(a_j + v_j)} \tag{2}$$

or

$$1n(P_i/P_j) = (a_i - a_j) + (v_i - v_j) \tag{3}$$

When a_i equals a_j (that is, the tree part of the model is dropped), Equation (3) reduces to Luce's simply scalable model. Similarly when the v_js are equal, the model becomes a strict decision tree model, in which an individual begins at the top and continues down the tree until the first time s/he reaches a branch where the 2 alternatives differ (possess different aspects). Choice is then determined by the relative value of the aspects (a_is).

For example, assume that an individual was choosing between 2 soft drinks, Classic Coke and Diet 7-Up using the choice process described by the hierarchical tree in Figure 1, where the first branch is based on calories and the second on flavor (cola versus lemon-lime). The choice probability of Classic Coke (versus Diet 7-Up) would depend on the difference in the aspect values of diet versus non-diet (i.e., the impact of calories) plus the brand-specific effects of Coke and Diet 7-Up (i.e., those characteristics not accounted for by calories and flavor type).

In choosing between 3 alternatives, choice depends on both the brand values and the most important aspect on which they differ. Assuming X_i and X_j take on one value on this aspect (a_{ij}) and X_k a different value (a_k), then:

$$\frac{P(X_i \text{ or } X_j : X_i, X_j, X_k)}{P(X_k : X_i, X_j, X_k)} = \frac{P_{ij}}{P_k} = \frac{\exp\,(a_{ij} + v_i + v_j)}{\exp\,(a_k + v_k)}$$

or

$$1n\,(P_{ij}/P_k) = (a_{ij} - a_k) + (v_i + v_j - v_k)$$

Such trinary comparisons are necessary for estimating the model parameters and are discussed in detail later.

The proposed AT model in Equation (3) has several useful features and inter-

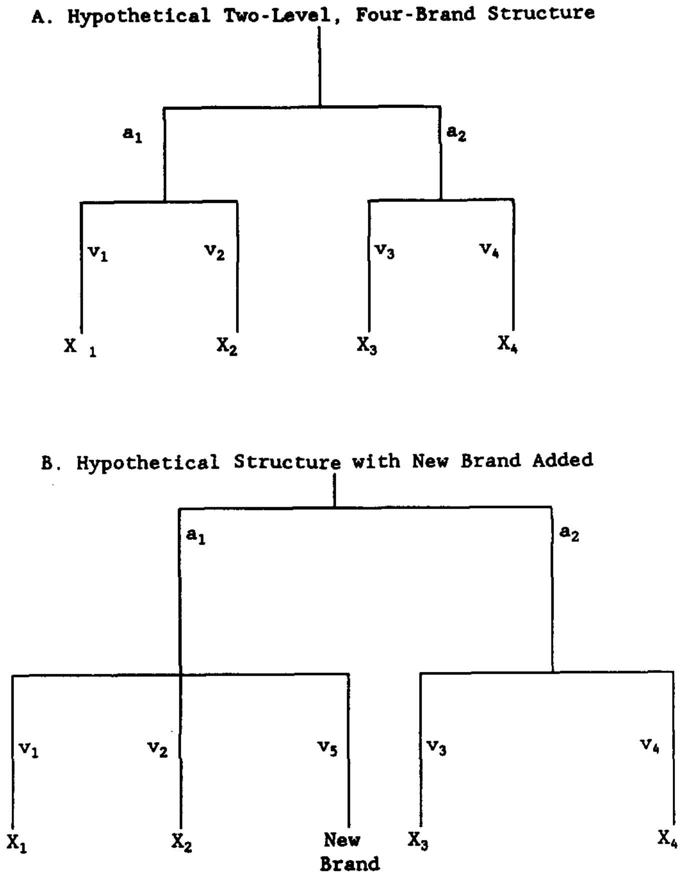


Figure 1

pretations. For example, the larger v_i is, the more positive is the benefit associated with that brand. Also consider the size of the differences involving the a_js versus those of the v_is. If the a_js have larger differences than the v_is, choice is largely determined by attributes and the product category is likely to be a market for which product aspects are much more important than image. By contrast, if the differences among the v_is are larger than those among the a_js, then (assuming that the appropriate attributes have been specified) brand image (equity, goodwill) matters more and the alternatives are likely to be intangible. Finally, since the model nests both a strict decision tree model and a simply scalable model, we can formally test whether the model improves on either of these simpler models.

The AT model assumes that an individual has or at least acts as if they have a hierarchical tree representation of the alternatives. (This tree may be either recalled from memory or constructed at the time of decision.) The particular structure to be analyzed can either come from prior theory or be derived empirically. Many methods have been proposed for deriving trees such as those based on brand switching data (Kalwani and Morrison, 1977; Lehmann, 1972; Rao and Sabavala, 1981; Vilcassim, 1989), similarity judgments (Sattath and Tversky, 1977), and concept learning schemes (Currim et al., 1988).

It is also possible to use the AT model itself to determine the best structure. Different structures can be estimated and the "best" chosen based on a combination of parsimony, fit, and examination of model parameters. In deciding which of several structures is most appropriate, the model parameters should be plausible and interpretable. Moreover, it is desirable that the difference in the aspect values (a_js) decreases as one moves down the tree. Thus, the tree structure can be determined by either a separate analysis or chosen after comparing several alternative forms of the AT model. (This later approach is demonstrated in the illustration presented later.)

Handling Similarity and Dominance Effects

The basic advantage of a hierarchical model over Luce is it can handle choice when the alternatives break into subcategories with preference largely dependent on subcategory. To see how AT handles the similarity effect, consider the red bus–blue bus problem. There are substantial differences between buses and cars, but much smaller differences between different buses or between different cars. Here, assume:

$$a_{\text{bus}} = a_{\text{car}} = 1 \text{ (buses and cars are equally preferred)}$$

and

$$v_{\text{red bus}} = v_{\text{blue bus}} = 0.01, \text{ and } v_{\text{car}} = 0.01.$$

Then

$$P(\text{car: car, blue bus}) = P(\text{car: car, red bus}) = 0.5$$

as

$$1n \frac{P(\text{car: car, blue bus})}{P(\text{blue bus: car, blue bus})} = 1 - 1 + 0.01 - 0.01 = 0$$

However,

$$P(\text{car: car, bus}) = \text{approximately } 0.4975$$

as

$$1n \frac{P(\text{car: car, red bus, blue bus})}{P(\text{red or blue bus: car, red bus, blue bus})}$$

$$= 1 - 1 + 0.01 - 0.01 - 0.01 = 0.01$$

Furthermore, the probability of chosing the red bus is equal to the probability of choosing a bus (0.5025) times the probability of choosing the red bus over the blue bus (0.5) or 0.25125.

The dominance problem can be handled in the following manner. Consider again the choice between a trip to Paris, a trip to Paris plus $1, and a trip to London where the second Paris trip dominates the first. Assume the following: a_{Paris} =

$a_{\text{London}} = 10$, and $v_i = 1$ for all alternatives. Also the two Paris trips differ on the attribute "money added" (\$1 versus \$0). Let $a_{\$1} = 10$ and $a_{\$0} = 0$. Then, one is indifferent between either of the Paris trips and the London trip as:

$$1n \frac{P(\text{Paris: Paris, London})}{P(\text{London: Paris, London})} = 1n \frac{P(\text{Paris} + \$1\text{: Paris} + \$1\text{, London})}{P(\text{London: Paris} + \$1\text{, London})}$$

$$= a_{\text{Paris}} - a_{\text{London}} + v_P - v_L = 10 - 10 + 1 - 1 = 0.$$

However, the trip Paris plus \$1 is greatly preferred over just the trip to Paris as they differ on the attribute \$1 versus \$0. In this case, the logit is:

$$1n \frac{P(\text{Paris} + \$1\text{: Paris} + \$1\text{, Paris})}{P(\text{Paris: Paris} + \$1\text{, Paris})} = a_{\$1} - a_{\$0} + v_{P+\$1} - v_P$$

$$= 10 - 0 + 1 - 1 = 10.$$

The corresponding probability of choosing Paris plus \$1 over just Paris in a paired comparison is thus essentially 1.

Incorporating a New Alternative

The impact of the incorporation of a new alternative in the model is fairly straightforward and helps demonstrate one of the key features of the model. Notice that adding a new alternative only requires estimation of one additional parameter, so this can be done quite easily. Consider the example in Figure 1A. If we add an alternative similar to X_1 and X_2 with unique aspect value v_5, we would have Figure 1B.

Before the new alternative was introduced, we would have the natural logarithm of the relative share of Segment I (X_1 and X_2) given by:

$$ln \frac{(\text{Share of Segment I})}{(\text{Share of Segment II})} = a_1 + v_1 + v_2 - a_2 - v_3 - v_4 \tag{4}$$

After the new alternative was introduced, the relative shares of the two segments would be given by:

$$ln \frac{(\text{Share of Segment I})}{(\text{Share of Segment II})} = a_1 + v_1 + v_2 + v_5 - a_2 - v_3 - v_4 \tag{5}$$

Relation to HEM

There are several important differences between the proposed AT model and HEM. In the classic case of 2 similar versus one dissimilar alternative (In Fig. 2, two cola-flavored soft drinks versus orange juice), the overall choice probabilities are given in Table 1. An important difference among the models is highlighted in the ratio of the paired comparison probability of Coke to that of orange juice. For the proposed AT model, this ratio is strictly multiplicative. Hence, the model can be linearized by taking logarithms and is easier to estimate than the additive ratio for HEM, which has a sum in the numerator. This ease in estimation is a direct consequence of the form of the value function in Equation (1).

To make the distinction between the models clearer, consider the entree choice

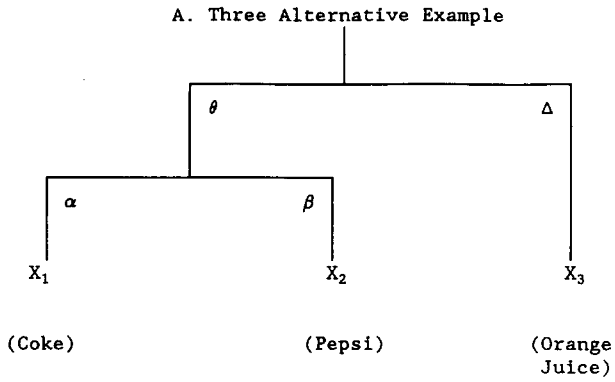


Figure 2

problem of Tversky and Sattath (1979) shown in Figure 3.According to HEM, when the sum of all the measures (λ, μ, etc.) equals one, the probability of choosing steak (S) from the set of alternatives A is:

$$P(S{:}A) = (\alpha + \beta + \gamma + \theta + \lambda)\frac{\alpha + \beta + \theta}{\alpha + \beta + \gamma + \theta}\;\frac{\alpha}{\alpha + \beta} \tag{6}$$

In a choice between only steak and trout, the ratio of the choice probabilities under HEM is:

$$\frac{P(S{:}S,T)}{P(T{:}S,T)} = \frac{\lambda + \theta + \alpha}{\mu + \phi} \tag{7}$$

Notice that this form is not conducive to least-squares estimation.

Now consider the AT model. Here the probability of choosing steak from the set A is:

$$P(S{:}A) = \frac{\exp\alpha}{\exp\alpha + \exp\beta}\cdot\frac{\exp(\theta + \alpha + \beta)}{\exp(\theta + \alpha + \beta) + \exp\gamma}\cdot\frac{\exp(\lambda + \alpha + \beta)}{\exp(\lambda + \alpha + \beta) + \exp(\mu + \Delta + \phi)} \tag{8}$$

While this unconditional probability is no simpler in form than the HEM model probability, the conditional probabilities of paired or trinary comparisons are easier to use. Under AT, the probability of choosing steak over trout is given by:

Table 1. Choice Probabilities for Different Models Associated with Figure 2

Model	$P(C{:}C,P,OJ)$	$\frac{P(C{:}\ C,\ OJ)}{P(OJ{:}\ C,\ OJ)}$	$\frac{P(C{:}\ C,\ P)}{P(P{:}\ C,\ P)}$
HEM	$\frac{\alpha}{\alpha + \beta}\;\frac{\alpha + \beta + \theta}{\alpha + \beta + \theta + \Delta}$	$\frac{\alpha + \theta}{\Delta}$	$\frac{\alpha}{\beta}$
Proposed model	$\frac{\exp\alpha}{\exp\alpha + \exp\beta}\cdot\frac{\exp(\theta + \alpha + \beta)}{\exp(\theta + \alpha + b) + \exp\Delta}$	$\frac{\exp(\theta + \alpha)}{\exp\Delta}$	$\frac{\exp\alpha}{\exp\beta}$

Tversky and Sattath's (1979) Figure 2

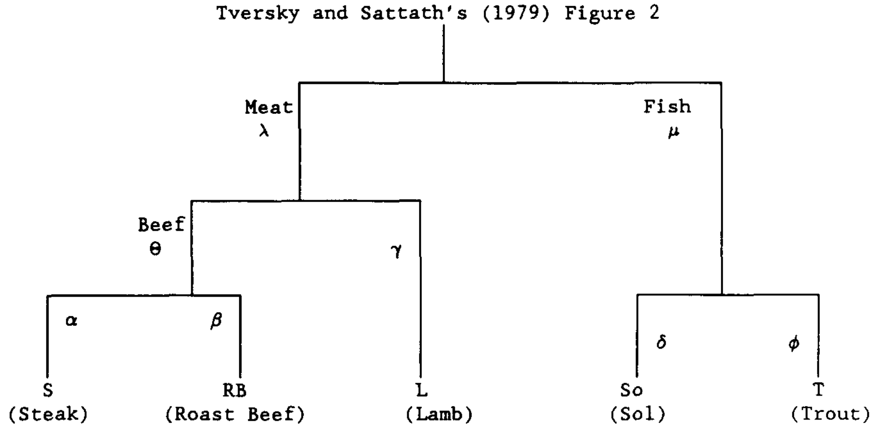


Figure 3

$$\frac{P(S{:}S,T)}{P(T{:}S,T)} = \frac{\exp(\lambda + \alpha)}{\exp(\mu + \phi)} \tag{9}$$

By taking the logarithms of each side, we have an easily estimable equation.

Similarly for a trinary comparison among steak, roast beef, and trout, the probability of choosing steak or roast beef over trout is given by:

$$\frac{P(S,RB{:}S,RB,T)}{P(T{:}S,RB,T)} = \frac{\exp(\lambda + \alpha + \beta)}{\exp(\mu + \phi)}$$

Again by taking logarithms of both sides we have a linear form that can be easily estimated by least squares. Therefore, a least-squares procedure can be used to estimate the parameters of the AT model given binary and trinary preferences or choice data.

Thus, the proposed model differs from HEM in 2 important ways. First, the AT model assumes that only the most salient (important) attribute on which the alternatives differ affects a choice between 2 alternatives. Second, the AT model is much easier to estimate than HEM. The multiplicative model allows a least-squares estimation. This means that the constraint on the number of brands to be modeled becomes adequate data (and consequently respondent burden and/or the assumption of stationary preferences) and not estimation.

Notice that if a brand has no unique value (adds nothing to a particular category), its v_i parameter will be 0. Thus, if roast beef adds nothing to the beef category, β will equal 0 and:

$$\frac{P(ST \text{ or } RB : ST,RB,TR)}{P(TR : ST,RB,TR)} = \frac{P(ST : ST,TR)}{P(TR : ST,TR)}$$

Also in the comparison between steak and roast beef,

$$ln\,\frac{P(ST : ST,RB)}{P(RB : ST,RB)} = \alpha - \beta = \alpha$$

Table 2. Hypothetical Choice Probabilities

Binary comparisons					
1) Pepsi	0.70	7pUp	0.30		
2) Pepsi	0.80	Diet Pepsi	0.20		
3) Pepsi	0.80	Diet 7-Up	0.20		
4) 7-Up	0.80	Diet Pepsi	0.20		
5) 7-Up	0.80	Diet 7-Up	0.20		
6) Diet Pepsi	0.90	Diet 7-Up	0.10		
Trinary comparisons					
1) Pepsi	0.56	7-Up	0.24	Diet Pepsi	0.20
2) Pepsi	0.56	7-Up	0.24	Diet 7-Up	0.20
3) Pepsi	0.80	Diet Pepsi	0.18	Diet 7-Up	0.02
4) 7-Up	0.80	Diet Pepsi	0.18	Diet 7-Up	0.02

If roast beef is *never* chosen, α will approach minus infinity, making estimation impossible. In order to obtain estimates, we can make $P(ST)/P(RB)$ equal to some arbitrarily large number (e.g., 200, implying the probability of roast beef being chosen is less than 1%. This is essentially the same as the procedure of setting the frequency of cells with no observations equal to a small fraction such as 0.005 in log-linear model estimation). Alternatively, if roast beef is never chosen, it can simply be eliminated from the tree, since it apparently never enters the decision process.

Estimation

The AT model can be estimated using dummy variable regression. There are n brand value parameters plus one parameter for each pair of the aspect nodes in the tree to estimate. The input data required to estimate the model are paired and trinary comparisons of the alternatives.

Consider the situation represented by Figure 1A, where X_1 is Pepsi, X_2 is 7-Up, X_3 is Diet Pepsi, and X_4 is Diet 7-Up. Assume the v_js $= 0$ and that the probability of choosing a non-diet drink is 0.8, the probability of choosing a cola given a non-diet drink is 0.7, and the probability of choosing a cola given a diet drink is 0.9. The data generated by this model are given in Table 2. These data would be coded as in Table 3. In order to estimate the model, the estimation of 4 brand level (v_1, v_2, v_3, and v_4) plus 3 aspect (a_1–a_2, a_3–a_4, and a_5–a_6) parameters is required. Notice here that we estimate only the contrasts between aspect values (e.g., cola versus lemon-lime) rather than the aspect values separately due to their perfect collinearity (i.e., cola only appears when lemon-lime does). Then we arbitrarily assign half the difference, appropriately signed, to each branch.

The model cannot be estimated on paired comparison data alone. In equation (1) we can determine only the sum of $ln\ a_j + ln\ v_j$, but not the individual estimates. More generally, in a multiple level tree we cannot separate the a_j values in the lowest level of the tree from the v_is. Therefore, data on a number of trinary comparisons are needed.

Each trinary comparison provides 2 pieces of information. First, assuming that X_i and X_j share some aspects that X_k doesn't, the model would predict the following:

$$ln\ [Pr(X_i \text{ or } X_j\text{: } X_i, X_j, X_k)/Pr(X_k\text{: } X_i, X_j, X_k)] \\ = a_{ij} - a_k + v_i + v_j - v_k \tag{10}$$

Table 3. Hypothetical Example: Data Coding

		Attributes						Brand Values			
	$ln(P_{ij}/P_{ji})$	1	2	3	4	5	6	1	2	3	4
Binary comparisons											
1) Pepsi/7-Up	0.847					1	−1	1	−1		
2) Pepsi/Diet Pepsi	1.386	−1	1					1		−1	
3) Pepsi/Diet 7-Up	1.386	−1	1					1			−1
4) 7-Up/Diet Pepsi	1.386	−1	1						1	−1	
5) 7-Up/Diet 7-Up	1.386	−1	1						1		−1
6) Diet Pepsi/Diet 7-Up	2.197			1	−1					1	−1
Trinary comparisons											
1a) Pepsi/7-Up	0.847					1	−1	1	−1		
1b) (Pepsi & 7-Up)/Diet Pepsi	1.386	−1	1					1	1	−1	
2a) Pepsi/7-Up	0.847					1	−1	1	−1		
2b) (Pepsi & 7-Up)/Diet 7-Up	1.368	−1	1					1	1		−1
3a) Diet Pepsi/Diet 7-Up	2.197					1	−1			1	−1
3b) (Diet Pepsi & Diet 7-Up)/ Pepsi	−1.386	1	−1					−1		1	1
4a) Diet Pepsi/Diet 7-Up	2.197			1	−1					1	−1
4b) (Diet Pepsi & Diet 7-Up)/ 7-Up	−1.386	1	−1						−1	1	1

where a_{ij} is the measure of the first aspect shared by X_i and X_j, but not X_k, and the other terms are defined as in equation (1).

A second piece of independent information is the paired comparison involving the two most similar brands, X_i and X_j, in the example just given, which is treated as in equation (1). (If all three brands are joined at the same node, the information from the trinary comparison is broken into two paired comparisons.)

Notice that it is the presence of these trinary comparisons that allows for estimation of a unique value for each brand (rather than only the differences between the values of brands such as Coke and Pepsi when two brands share a common attribute structure), and makes the model distinguishable from a simply scalable model. The estimation of unconstrained values for the brand effects is important since constraining the *v*s to sum to zero damages the ability of the model to represent the relative importance of aspects and brand values through comparing the relative sizes of the *a*s and the *v*s. In fact, only trinary comparisons are necessary for estimating the model.

Method

Perceptual and preference data were collected on 12 brands of soft drinks from 15 students in order to illustrate the model. Soft drinks were chosen because they are relevant to the subjects and because they have a long tradition of use in choice and preference studies (e.g., Bass et al., 1972). The brands involved were Pepsi Cola, Coca-Cola, 7-Up, Sprite, Pepsi-Free, Caffeine-Free Coke, Diet Pepsi-Free, Caffeine-Free Diet Coke, Diet Pepsi, Diet Coke, Diet 7-Up, and Sugar-Free Sprite. The products contained 8 colas and 4 lemon-lime flavors, 6 diet and 6 non-diet drinks, and 4 drinks with and 8 without caffeine. The students provided constant sum (100 points) preference judgments on all 66 possible pairs. Constant sum preferences (100 points) were also collected on 50 selected triples.

Table 4. Comparisons of Augmented Decisions Tree via R^2s

Subject	Flavor		Diet		Caffeine	
	Luce	AT	Luce	AT	Luce	AT
1	0.76	0.79[a]	0.80	0.83[a]	0.75	0.77[b]
2	0.78	0.82[a]	0.78	0.86[a]	0.77	0.78
3	0.80	0.82[b]	0.80	0.81	0.79	0.80
4	0.82	0.84[b]	0.81	0.83[b]	0.81	0.83[b]
5	0.77	0.77	0.76	0.77	0.79	0.79
6	0.79	0.81[b]	0.79	0.89[a]	0.78	0.79
7	0.67	0.68	0.68	0.73[a]	0.69	0.71[b]
8	0.88	0.88	0.87	0.88	0.86	0.87
9	0.73	0.77[a]	0.72	0.75[a]	0.73	0.75[b]
10	0.58	0.60[b]	0.61	0.64[a]	0.61	0.62
11	0.73	0.80[a]	0.76	0.78[b]	0.76	0.79[a]
12	0.70	0.74[a]	0.72	0.72	0.69	0.72[a]
13	0.52	0.56[a]	0.55	0.59[a]	0.56	0.57
14	0.87	0.88	0.92	0.95[a]	0.87	0.89[b]
15	0.53	0.56[a]	0.51	0.52	0.51	0.54[a]
Average	0.73	0.75	0.74	0.77	0.73	0.75

[a]Fits better than Luce at 0.05 level.
[b]Fits better than Luce at 0.10 level.

Theoretically, one would like a long string of actual choice data with which to estimate these models. However, in experimental situations, demand effects and nonstationarity are likely to affect the results. Scanner data suffer from the impact of multiple users, changes in the market place, and the inability to manipulate the choice set. Therefore, we have used preference data to calibrate the model.

In general, the number of data points to be collected depends on the number of parameters to be estimated. In the 12-brand example presented here, that means 11 + 5 = 16 parameters. Consequently, estimates are possible with a small number of data points and 30–50 binary and trinary comparisons would be enough for reasonably stable estimates, suggesting that respondent burden is not a major problem. Thus, for purposes of estimating the models, far fewer than the 66 paired comparisons and 50 trinary comparisons are needed. However, since the subjects provided these data, all the data were used in the analysis.

Results

The tree model was applied to the 15 repondents from the soft drink study. There are 2 possible trees for each primary split (e.g., Flavor/Diet/Caffeine and Flavor/Caffeine/Diet). However, in keeping with the illustrative nature of this example, only one of these two possibilities was run for each primary split. Specifically, the three aspect orders tested were Flavor/Diet/Caffeine, Diet/Flavor/Caffeine, and Caffeine/Diet/Flavor. The 2 comparisons implied by the trinary comparisons were different for each of the primary splits, so the Luce model was estimated three times for each subject. The results of these regressions are given in Table 4.

Table 4 leads to several interesting conclusions. First, as usual, the Luce model fits the data quite well. Second, the AT model fits significantly (at the 0.10 level) better than the Luce model in 11/15, 10/15, and 8/15 cases for the Flavor-, Diet-, and Caffeine-based trees respectively. Each AT model explains about 8% of the variance unexplained by the Luce model and each one fits significantly better. Also, for 13 of the 15 subjects at least 1 of the 3 forms of the proposed AT model fits significantly better than Luce (based on an F test for the additional parameters). This suggests that even with the strong performance of the Luce model, the proposed model is superior. Third, it is possible based on the fit of the model to classify individuals by the importance of attributes. For example, subject 2 is clearly concerned about calorie content, while subject 9 is more concerned about flavor. Still, the over-riding result here is that most of the models fit the data reasonably well.

It is not reasonable to expect the AT model to predictively outperform Luce by a wide margin in most real markets. First, the Luce model has proven to be a robust and useful model in a wide variety of situations. Second, it captures most of the variance (here an average 73%) so that given data with error, it may be close to the upper bound in terms of explained variance. What is important is that it does significantly improve predictive power in the context of a parsimonious and interpretable model. When this occurs, the *interpretation* of the choice process can be radically different (i.e., it is very different to conclude that choice is based on the brand images of Coke and Pepsi than to assume it is based on the desire for cola). It also points out systematic deviations from the Luce model.

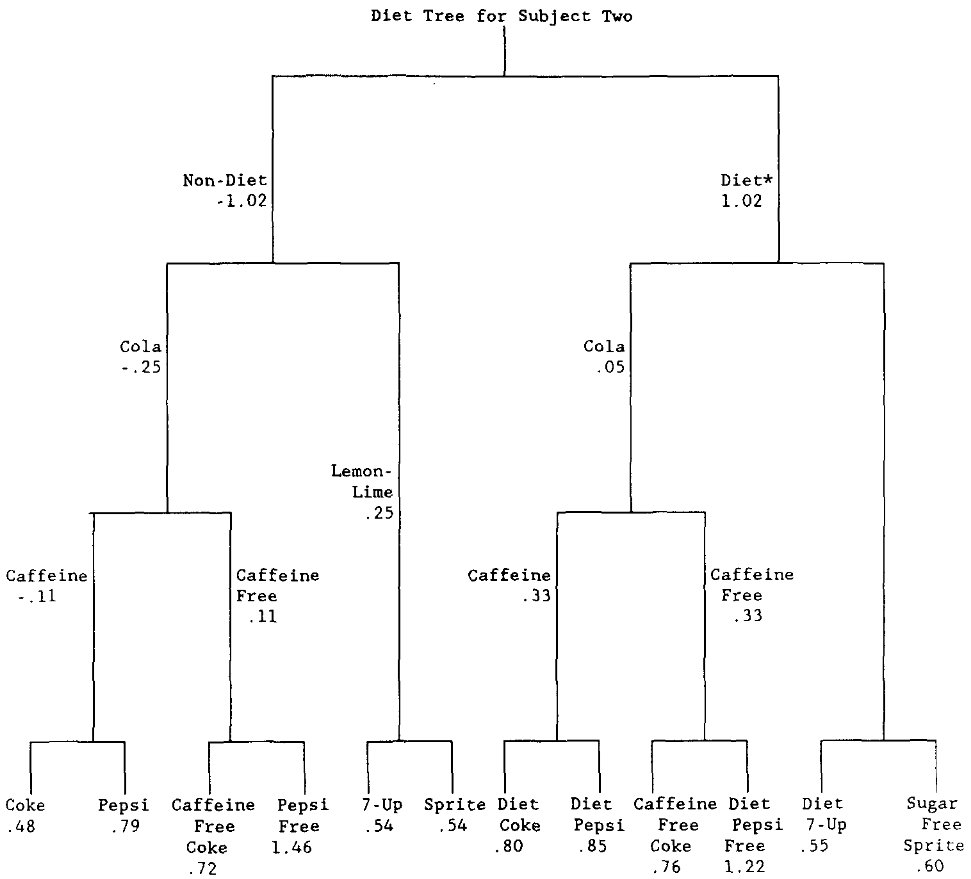
As an example of how to interpret an individual's results, the diet-based tree for subject 2 (the calorie-conscious subject) is shown in Figure 4. Since the sum of the aspect coefficients that connect to a node is zero, we arbitrarily put half the estimated coefficient (which measures the difference between the values) on each branch with the appropriate signs. The magnitude of the various aspect coefficients suggests that this person is most interested in calories, likes Pepsi-Free in either diet or non-diet form, and thinks that the other brands *per se* are fairly equal in value. Notice that subject 2 is also fit quite well by the Luce model. Yet since the AT model fits significantly better, it appears that the process the individual follows involves some hierarchical attribute-based decisions, here involving calories. Hence, marketers who approach consumers such as subject 2 with pure image advertising (aimed at v_i) may be disappointed in the effect of the advertising. Thus, the major difference between AT and Luce tends to be in interpretation rather than fit.

This 12-brand example also highlights the relative parsimony of the approach. EBA, which assumes no particular structure, would require that over 4,000 parameters be estimated. With Batsell and Polking's (1985) approach and looking at only first-, second-, and third-order effects, 78 parameters would have to be estimated. Thus, for reasonably large choice sets, it seems obvious that some *a priori* structure is needed.

When the number of parameters of the simpler models are compared, it is seen in this example that Luce has 11, the proposed AT model 16, and HEM 21. Thus, AT not only blends hierarchical and brand-based choice, but it does so with fewer parameters than the best known hierarchical model.

Discussion

This article has demonstrated a relatively straightforward way to incorporate the similarity and dominance effects into hierarchical preference models using augmented decision tree models. The model appears to be more tractable than current



* = aspect coefficient had a t greater than 2
(all the brand values had t's greater than 2)

Figure 4

methods for estimation of preference tree models as well as the alternative methods that do not assume simple scalability. The illustration utilized a large number of brands to demonstrate the relative ease of data collection and parameter estimation, and showed that the model fit the data well.

We expect the AT model will prove most beneficial when distinct physical features differentiate the product (as in VHS vs Super 8 camcorders or various chemicals), and the benefits will be primarily in interpretation rather than fit. However, given the robustness of the Luce model and the existence of other hierachical models such as HEM and nested logit, considerable further testing is required to establish the value of the AT model. Future research should examine the applicability of the model to other situations, such as durables and industrial goods (e.g., chemicals), as well as studying actual choice as opposed to preference. Also, work is needed to see how stable the estimates are on small data sets. Still, the relative parsimony of the model along with its good fit suggest that it is a useful candidate for further work.

References

Bass, Frank M., Pessemier, Edgar A. and Lehmann, Donald R. Experimental Study of Relationships Between Attitudes, Brand Preference and Choice," *Behavioral Science,* 17 (1972):352–541.

Batsell, Richard R. and Polking, John C. A New Class of Market-Share Models, *Marketing Science,* 4 (Summer 1985):177–198.

Becker, Gordon M., DeGroot, M. H. and Marschak, J. Probabilities of Choice Among Very Similar Objects, *Behavioral Science,* 8 (1963):306–311.

Buckley, Patrick G. Nested Multinominal Logit Analysis of Scanner Data for a Hierarchical Choice Model, *Journal of Business Research,* 17 (1988):133–154.

Currim, Imran S., Meyer, Robert J., and Lee, Nhan I. Disaggregate Tree-Structural Modeling of Consumer Choice Data, *Journal of Marketing Research,* 25 (August 1988):253–265.

Debreu, G. Review of R. D. Luce, Individual Choice Behavior: A Theoretical Analysis, *American Economic Review,* 50 (1960):186–188.

Dubin, Jeffrey A. A Nested Logit Model of Space and Water Heat System Choice, *Marketing Science,* 5 (Spring 1986):112–124.

Hauser, John Agendas and Consumer Choice, *Journal of Marketing Research,* 23 (August 1986):199–212.

Howard, John *Consumer Behavior: Application of Theory,* New York: John Wiley, 1977.

Jain, Dipak L. and Bass, Frank M. Effect of Choice Set Size on Choice Probabilities: An Extended Logit Model, *International Journal of Research in Marketing,* 6 (1989):1–11.

Johnson, Michael D. The Differential Processing of Product Category and Noncomparable Choice Alternatives, *Journal of Consumer Research,* 16 (December 1989):300–309.

Kalwani, Manohar, and Morrison, Donald G. A Parsimonious Description of the Hendry Model, *Management Science,* 23 (January 1977):467–477.

Lehmann, Donald R. Judged Similarity and Brand-Switching Data as Similarity Measures, *Journal of Marketing Research,* 9 (August 1972):331–334.

Louviere, Jordan J. and Woodworth, George Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research,* 20 (November 1983):350–367.

Luce, R. Duncan *Individual Choice Behavior,* New York: John Wiley & Sons, Inc. 1959.

Luce, R. Duncan The Choice Axiom After Twenty Years, *Journal of Mathematical Psychology,* 15 (1977):215–233.

McFadden, Daniel Conditional Logit Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics,* P. Zaremba, ed., Academic Press, New York, 1974, pp. 105–142.

McFadden, Daniel Modelling the Choice of Residential Location, *Transportation Research Record,* 673 (1978). 72–78.

McFadden, Daniel Econometric Models of Probabilistic Choice, in *Structural Analysis of Discrete Data with Econometric Applications,* C. F. Manski and D. McFadden (eds.), Cambridge, MA: The MIT Press, 1981.

Moore, William L., and Donald R. Lehmann A Paired Comparison Nested Logit Model of Individual Preference Structures *Journal of Marketing Research,* 26 (November 1989): 460–468.

Moore, William L., Pessemier, Edgar A., and Lehmann, Donald R. "Hierarchical Representations of Market Structures and Choice Processes Through Preference Trees," *Journal of Business Research,* 14 (November 1986):371–386.

Park, C. Whan, and Smith, Daniel C. Product Level Choice: A Bottom-Up or Top-Down Process, *Journal of Consumer Research,* 16 (December 1989): 289–299.

Rao, Vithala R. and Sabavala, Darius C. Inference Hierarchial Choice Processes from Panel Data, *Journal of Consumer Research,* 8 (June 1981):85–96.

Restle, F. *Psychology of Judgment and Choice: A Theoretical Essay,* New York: John Wiley & Sons, Inc., 1961.

Rumelhart, Donald L., and Greeno, J. G. Similarity Between Stimuli: An Experimental Test of the Luce and Restle Choice Models, *Journal of Mathematical Psychology,* 8 (1971):171–191.

Sattath, Shmuel and Tversky, Amos, Additive Similarity Trees, *Psychometrika,* 42 (1977):319–45.

Thurstone, Leon L. A Law of Comparative Judgment, *Psychological Review,* 34 (1927):273–286.

Tversky, Amos Elimination by Aspects: A Theory of Choice, *Psychological Review,* 79 (4) (1972):281–289.

Tversky, Amos, and Sattath, Shmuel Preference Trees, *Psychological Review,* 86 (November 1979):542–593.

Vilcassim, Naufel J. Extending the Rotterdam Model to Test Hierarchical Market Structures, *Marketing Science,* 8 (Spring 1989):181–190.

Zajonc, Robert Feeling and Thinking: Preferences Need No Inferences, *American Psychologist,* 35 (February 1980):151–175.

Zajonc, Robert On the Primacy of Affect, *American Psychologist,* 39 (February 1984):117–123.