informs
http://pubsonline.informs.org/journal/opre/

OPERATIONS RESEARCH
Vol. 65, No. 3, May–June 2017, pp. 621–634
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Relation Between EBA and Nested Logit Models

Rajeev Kohli,[a] Kamel Jedidi[a]

[a] Graduate School of Business, Columbia University, New York, New York 10027

Contact: rk35@columbia.edu (RK); kj7@columbia.edu (KJ)

Received: February 12, 2015
Revised: March 9, 2016; June 8, 2016; August 24, 2016
Accepted: September 23, 2016
Published Online in Articles in Advance: March 23, 2017

Subject Classifications: marketing: choice models, estimation/statistical technique, buyer behavior
Area of Review: Marketing Science

https://doi.org/10.1287/opre.2016.1580

Copyright: © 2017 INFORMS

Abstract. We show that elimination by aspects (EBA) generalizes nested logit and cross-nested logit models. The latter two models are equivalent to a special case of EBA called preference trees. The transformations between preference trees and nested logit models become more complex when the utilities of alternatives are functions of covariates. In this case, a simple model in one domain corresponds to a complex model in the other. An extended EBA model, in which the utilities of alternatives are functions of covariates, represents a two-stage choice process. Alternatives are first screened using a probabilistic lexicographic rule and then compared in terms of their compensatory utilities. We provide a typology of the relations between EBA and other logit models, and we discuss issues concerning estimation, statistical testing, and data collection. We describe an application illustrating (1) the process of constructing a preference tree with covariates and (2) the different implications obtained from a preference tree and a comparable nested logit model.

Keywords: elimination by aspects • preference trees • hierarchical elimination model • nested logit model • cross-nested logit model • two-stage choice model • order independence • independence of irrelevant alternatives

1. Introduction

Elimination by aspects (EBA) is a probabilistic choice model that generalizes the multinomial logit and rank-ordered logit models (Kohli and Jedidi 2015). We show that it also generalizes nested logit and cross-nested logit models. The latter two models are equivalent to a special case of EBA called preference trees (Tversky and Sattah 1979). The equivalence is simpler when the utilities of the alternatives are not functions of covariates. Otherwise, a simple preference tree corresponds to a complex nested logit model, and a simple nested logit model to a complex preference tree. A two-stage EBA-like model can be formulated in which options are first screened using a probabilistic lexicographic rule and then evaluated in terms of their compensatory utilities.

Our analysis extends the result obtained by Batley and Daly (2006), who showed that a nested logit model with three alternatives and two nests is equivalent to a preference tree. Although limited to a small problem, their result is notable because earlier research was either inconclusive about the relation between nested logit models and preference trees (McFadden 1981) or suggested that the two models were similar but distinct (Tversky and Sattah 1979). Like Batley and Daly (2006), we consider the equivalence between EBA and nested and cross-nested logit models, but not the entire class of Generalized Extreme Value (GEV) models, which is known to approximate any random utility model arbitrarily closely (Dagsvik 1995).

The present results may be useful for the following reasons. First, they provide a more complete account of the extent to which EBA generalizes logit models. Second, nested logit and cross-nested logit models assume a generalized extreme value distribution. EBA and preference trees do not require this assumption, but only assume that the aspects have independent, extreme value distributions. Third, simple formulations of nested logit models and preference trees are not equivalent when the utilities of the alternatives are functions of covariates. In these cases, both models can be estimated and compared. Fourth, by using attributes as covariates, EBA and preference trees can be extended to model a two-stage choice process. In the first stage, a subset of alternatives is screened using a probabilistic lexicographic rule. In the second stage, an alternative is selected if it has the highest multi-attribute utility among the screened alternatives. This formulation has the advantage that the parameters for both stages are simultaneously estimated using a common error structure.

Whether to use EBA, its generalization, or one of its special cases, depends on the same considerations that apply when selecting any general model or its special cases: the modeling objective and the adequacy of the restrictions imposed by a special case. A multinomial logit model is appropriate when it is known, or verified after estimation, that independence of irrelevant alternatives (IIA) violations do not occur. A nested logit model or a preference tree is suitable when violations of order independence are expected to occur across, but not within, nests.[1] The more general EBA is useful when substitutability is expected, or confirmed after estimation, to differ across pairs of alternatives. The extended EBA formulation, in which the utilities of alternatives are functions of covariates, is appropriate when a tree structure is too restrictive but the unrestricted EBA model has too many parameters. It is also useful for estimating a two-stage choice process.

Section 2 characterizes the equivalence of nested logit models and preference trees, with and without covariates. It also shows that preference trees can represent cross-nested logit models. Section 3 describes the known relations between EBA and logit models. Section 4 discusses estimation, statistical testing, and data collection for EBA. Section 5 describes an application in which the utilities of alternatives are functions of covariates. It illustrates the process of building an extended preference tree, compares the choice processes for a preference tree and a nested logit model, and contrast their marketing implications.

2. Equivalence

We describe preference trees and nested logit models and show how the parameters of one can be transformed into the parameters of the other. We then obtain a transformation mapping the parameters of a cross-nested logit model onto the parameters of a preference tree. Finally, we obtain a transformation of parameters between preference trees and nested logit models when the utilities of alternatives are functions of covariates.

We begin by summarizing EBA. Let A denote the set of aspects describing the alternatives in a choice set C. An aspect can be a discrete attribute level (say, the color blue) or a threshold value associated with a continuous attribute (say, a maximum price). Let aspect $k \in A$ have random utility $u_k = v_k + \epsilon_k$, where ϵ_k has an independent, extreme value distribution. EBA eliminates alternatives and aspects in stages. Let A_j denote the set of aspects that appear in one or more of the alternatives at stage j. EBA selects aspect $k \in A_j$ with probability $p_k = \exp(v_k)/\sum_{i \in A_j} \exp(v_i)$. It eliminates all alternatives that do not have aspect k and then advances to the next stage, stopping when only one alternative remains.

2.1. Preference Tree

A preference tree is a special case of EBA in which the aspects and alternatives can be represented by a tree (Tversky and Sattah 1979). Each node in a preference tree, except the root, is associated with a distinct aspect. Each path from the root to a terminal node describes an alternative. We say that a preference tree has $n \geqslant 1$ levels if each path from the root to a terminal node has n aspects.[2] We use the terms "node" and "aspect" interchangeably in the following development.

Let $S(0) = \{k_0\}$, where k_0 denotes the root of the preference tree. Let $S(k_0)$ be the set of nodes that are (immediate) successors of the root. Let $S(k_0, k_1)$ be the set of nodes that are successors of node $k_1 \in S(k_0)$. More generally, let $S(k_0, \ldots, k_r)$ be the set of nodes that are successors of node $k_r \in S(k_0, \ldots, k_{r-1})$, where $r = 1, \ldots, n$. Each path from the root to a terminal node in the tree represents an alternative. The nodes on the path correspond to the aspects that appear in the alternative. For brevity, we sometimes use the term "alternative k_n" to refer to an alternative associated with the terminal node $k_n \in S(k_0, \ldots, k_{n-1})$.

We define the following sets to facilitate the representations of preference trees and nested logit models with arbitrary numbers of levels. Let

$$\Omega_0 = \{S(0)\} \quad \text{and} \quad \Omega_1 = \{S(k_0)\}.$$

Node $k_1 \in S(k_0)$ is succeeded by the nodes in the set $S(k_0, k_1)$. The collection of these sets is

$$\Omega_2 = \{S(k_0, k_1) \mid k_1 \in S(k_0), S(k_0) \in \Omega_1\}.$$

More generally, node $k_{r-1} \in S(k_0, \ldots, k_{r-2})$ is succeeded by the nodes in the set $S(k_0, \ldots, k_{r-1})$. The collection of these sets is

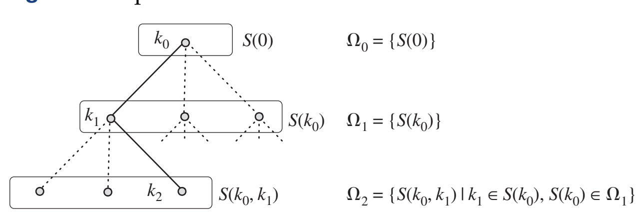
$$\Omega_r = \{S(k_0, \ldots, k_{r-1}) \mid k_{r-1} \in S(k_0, \ldots, k_{r-2}), S(k_0, \ldots, k_{r-2}) \in \Omega_{r-1}\}, \quad \text{for each } r = 2, \ldots, n.$$

Figure 1 shows the sets Ω_0, Ω_1 and Ω_2 for a tree with $n = 2$ levels.

Example 1. Let the first level of a preference tree have two nodes, representing ground and air travel. Let the node representing ground travel be succeeded by three nodes, representing train, bus, and car. Similarly, let the node representing air travel be succeeded by three nodes, representing first class, business class, and economy class. Then

$$\begin{aligned} S(0) &= \{k_0\}, \quad \Omega_0 = \{S(0)\} \\ S(k_0) &= \{\text{Ground, Air}\}, \quad \Omega_1 = \{S(k_0)\} = \{\{\text{Ground, Air}\}\} \\ S(k_0, \text{Ground}) &= \{\text{Train, Bus, Car}\}, \\ S(k_0, \text{Air}) &= \{\text{First class, Business class, Economy class}\} \\ \Omega_2 &= \{S(k_0, \text{Ground}), S(k_0, \text{Air})\} \\ &= \big\{\{\text{Train, Bus, Car}\}, \\ &\quad \{\text{First class, Business class, Economy class}\}\big\}. \end{aligned}$$

Figure 1. A preference tree with $n = 2$ levels



Notes. Every path from the root k_0 to a terminal node $k_2 \in S(k_0, k_1)$ identifies an alternative, for each $S(k_0, k_1) \in \Omega_2$.

Let aspect k have random utility $u_k = v_k + \epsilon_k$, where ϵ_k has an independent, extreme value distribution for each $k \in S$, where $S \in \Omega_r$ and $r = 1, \ldots, n$. Then the EBA choice probabilities can be computed in the manner described by Kohli and Jedidi (2015).

To obtain a transformation of parameters between preference trees and nested logit models, we use a result by Tversky and Sattah (1979) showing that the choice probabilities in a preference tree are equal to those in a hierarchical elimination model, which we discuss next.

Hierarchical Elimination Model. In the first stage, the hierarchical elimination model selects aspect $k_1 \in S(k_0)$ with probability

$$p(k_1 \in S(k_0)) = \frac{\exp(v_{k_1}) + J_{k_1}}{J_{k_0}},$$

where

$$J_{k_0} = \sum_{k \in S(k_0)} \{\exp(v_k) + J_k\},$$

and

$$J_{k_1} = \sum_{k \in S(k_0, k_1)} \{\exp(v_k) + J_k\}, \quad \text{for all } k_1 \in S(k_0) \text{ and } S(k_0) \in \Omega_1.$$

In the second stage, it selects an aspect $k_2 \in S(k_0, k_1)$ with probability

$$p(k_2 \in S(k_0, k_1)) = \frac{\exp(v_{k_2}) + J_{k_2}}{J_{k_1}},$$

where

$$J_{k_2} = \sum_{k \in S(k_0, k_1, k_2)} \{\exp(v_k) + J_k\}, \quad \text{for all } k_2 \in S(k_0, k_1) \text{ and } S(k_0, k_1) \in \Omega_2.$$

In general, if the first $r - 1$ stages select aspects $k_1, \ldots, k_{r-1}$, then the rth stage selects an aspect $k_r \in S(k_0, \ldots, k_{r-1})$ with probability

$$p(k_r \in S(k_0, \ldots, k_{r-1})) = \frac{\exp(v_{k_r}) + J_{k_r}}{J_{k_{r-1}}},$$

where

$$J_{k_r} = \sum_{k \in S(k_0, \ldots, k_r)} \{\exp(v_k) + J_k\}, \quad \text{for all } k_r \in S(k_0, \ldots, k_{r-1}) \text{ and } S(k_0, \ldots, k_{r-1}) \in \Omega_r,\ r = 1, \ldots, n-1.$$

The values of $J_{k_{n-1}}, \ldots, J_{k_0}$ are obtained recursively, using the initial values $J_{k_n} = 0$, for all $k_n \in S(k_0, \ldots, k_{n-1})$, where $S(k_0, \ldots, k_{n-1}) \in \Omega_n$. The unconditional probability that the hierarchical elimination model selects an alternative associated with the terminal node $k_n \in S(k_0, \ldots, k_{n-1})$ is

$$P(k_n \in S(k_0, \ldots, k_{n-1})) = \prod_{r=1}^{n} p(k_r \in S(k_0, \ldots, k_{r-1})).$$

Tversky and Sattah (1979) proved the following equivalence theorem.

Theorem 1. *The EBA choice probabilities for a preference tree are equal to the choice probabilities for a hierarchical elimination model.*

2.2. Nested Logit Model

A nested logit model has the same graphical representation as a preference tree. Each terminal node corresponds to an alternative. Each other node corresponds to a nest. We use the set notation introduced in the preceding subsection to represent sets of nodes.

Let $u'_k = w_k + e_k$ denote the random utility of alternative $k \in S(k_0, \ldots, k_{n-1})$, for all $S(k_0, \ldots, k_{n-1}) \in \Omega_n$. Here, w_k is a deterministic utility component, and e_k is a random component with a generalized extreme value distribution. The nested logit model associates a (nesting) parameter λ_k with each nest $k \in S(k_0, \ldots, k_{r-1})$, where $S(k_0, \ldots, k_{r-1}) \in \Omega_r$ and $r = 1, \ldots, n$. A sufficient condition for a nested logit model to be consistent with utility maximization is that the nesting parameters λ_k lie in the interval $(0, 1]$. To identify the model, let (1) $\lambda_{k_0} = 1$, and (2) $\lambda_{k_n} = 1$, for each terminal node $k_n \in S(k_0, \ldots, k_{n-1})$, where $S(k_0, \ldots, k_{n-1}) \in \Omega_n$. The nested logit model selects nest $k_1 \in S(k_0)$ with probability

$$q(k_1 \in S(k_0)) = \frac{\exp((\lambda_{k_1}/\lambda_{k_0}) I_{k_1})}{\exp(I_{k_0})}.$$

In this expression,

$$I_{k_0} = \ln\left\{\sum_{k \in S(k_0)} \exp\left(\frac{\lambda_k}{\lambda_{k_0}} I_k\right)\right\}$$

is the inclusive values associated with nest $k_0 \in S(0)$; and

$$I_{k_1} = \ln\left\{\sum_{k \in S(k_0, k_1)} \exp\left(\frac{\lambda_k}{\lambda_{k_1}} I_k\right)\right\}$$

is the inclusive value associated with nest $k_1 \in S(k_0)$, where $S(k_0) \in \Omega_1$.

Suppose nest $k_1 \in S(k_0)$ has been selected. Then the nested logit model selects nest $k_2 \in S(k_0, k_1)$ with probability

$$q(k_2 \in S(k_0, k_1)) = \frac{\exp((\lambda_{k_2}/\lambda_{k_1}) I_{k_2})}{\exp(I_{k_1})}, \quad \text{for all } k_2 \in S(k_0, k_1).$$

In this expression,

$$I_{k_2} = \ln\left\{\sum_{k \in S(k_0, k_1, k_2)} \exp\left(\frac{\lambda_k}{\lambda_{k_2}} I_k\right)\right\}$$

is the inclusive value associated with nest $k_2 \in S(k_0, k_1)$, where $S(k_0, k_1) \in \Omega_2$.

More generally, suppose that the sequence of nests $k_1, \ldots, k_{r-1}$ has been selected. Then the nested logit model selects nest $k_r \in S(k_0, \ldots, k_{r-1})$ with probability

$$q(k_r \in S(k_0, \ldots, k_{r-1})) = \frac{\exp((\lambda_{k_r}/\lambda_{k_{r-1}}) I_{k_r})}{\exp(I_{k_{r-1}})}.$$

In this expression,

$$I_{k_r} = \ln\left\{\sum_{k\in S(k_0,\ldots,k_r)} \exp\left(\frac{\lambda_k}{\lambda_{k_r}} I_k\right)\right\}$$

is the inclusive value associated with nest $k_r \in S(k_0,\ldots,k_{r-1})$, where $S(k_0,\ldots,k_{r-1}) \in \Omega_r$, for all $r = 1,\ldots,n-1$. Since $I_{k_{r-1}}$ is a function of I_{k_r}, the values of $I_{k_{n-1}},\ldots,I_{k_0}$ are obtained recursively, using the initial values $I_{k_n} = w_{k_n}$, for all $k_n \in S(k_0,\ldots,k_{n-1})$, where $S(k_0,\ldots,k_{n-1}) \in \Omega_n$. The unconditional probability that the nested logit model chooses an alternative associated with the terminal node $k_n \in S(k_0,\ldots,k_{n-1})$ is

$$Q(k_n \in S(k_0,\ldots,k_{n-1})) = \prod_{r=1}^{n} q(k_r \in S(k_0,\ldots,k_{r-1})).$$

2.3. Equivalence of Preference Trees and Nested Logit Models

We show that preference trees and nested logit models are equivalent. Given the parameter values for one model, there are parameter values for the other that obtain the same choice probabilities for all alternatives. We prove the following theorem in the appendix.

Theorem 2. *Preference trees and nested logit models are equivalent.*

The following example illustrates the proof of Theorem 2.

Example 2. Consider a tree with $n = 3$ levels. Let $J_{k_3} = 0$, $\lambda_{k_0} = \lambda_{k_3} = 1$ and $I_{k_3} = w_{k_3}$. Then the probabilities of selecting node $k_3 \in S$, where $S = S(k_0,k_1,k_2) \in \Omega_3$, have the following values for the preference tree and the nested logit model:

$$p(k_3 \in S) = \frac{\exp(v_{k_3}) + J_{k_3}}{J_{k_2}} = \frac{\exp(v_{k_3})}{J_{k_2}},$$
$$q(k_3 \in S) = \frac{\exp((\lambda_{k_3}/\lambda_{k_2})I_{k_3})}{\exp(I_{k_2})} = \frac{\exp((1/\lambda_{k_2})w_{k_3})}{\exp(I_{k_2})}.$$

Since $J_k = 0$ for $k \in S(k_0,k_1,k_2)$, the value of J_{k_2} in the preceding expression for $p(k_3 \in S)$ is

$$J_{k_2} = \sum_{k\in S(k_0,k_1,k_2)} \{\exp(v_k) + J_k\} = \sum_{k\in S(k_0,k_1,k_2)} \exp(v_k).$$

Similarly, since $\lambda_k = 1$ when $k \in S(k_0,k_1,k_2)$, the value of $\exp(I_{k_2})$ in the preceding expression for $q(k_3 \in S)$ is

$$\exp(I_{k_2}) = \sum_{k\in S(k_0,k_1,k_2)} \exp\left(\frac{\lambda_k}{\lambda_{k_2}} I_k\right) = \sum_{k\in S(k_0,k_1,k_2)} \left(\frac{1}{\lambda_{k_2}} w_k\right).$$

Equating the numerators of $p(k_3 \in S)$ and $q(k_3 \in S)$ gives

$$v_{k_3} = \frac{w_{k_3}}{\lambda_{k_2}}, \quad \text{for all } k_3 \in S = S(k_0,k_1,k_2).$$

It follows that $J_{k_2} = \exp(I_{k_2})$ and $p(k_3 \in S) = q(k_3 \in S)$, because the denominators of $p(k_3 \in S)$ and $q(k_3 \in S)$ are equal to the sum of their numerators across all $k_3 \in S$.

Similarly, the probabilities of selecting node $k_2 \in S$, where $S = S(k_0,k_1) \in \Omega_2$, have the following values for the preference tree and the nested logit model:

$$p(k_2 \in S) = \frac{\exp(v_{k_2}) + J_{k_2}}{J_{k_1}},$$
$$q(k_2 \in S) = \frac{C_2 \exp((\lambda_{k_2}/\lambda_{k_1})I_{k_2})}{C_2 \exp(I_{k_1})},$$

where C_2 is a positive constant associated with the second level of the tree. Equating the numerators of $p(k_2 \in S)$ and $q(k_2 \in S)$ gives

$$\exp(v_{k_2}) + J_{k_2} = C_2 \exp\left(\frac{\lambda_{k_2}}{\lambda_{k_1}} I_{k_2}\right), \quad \text{for all } k_2 \in S = S(k_0,k_1).$$

It follows that $J_{k_1} = C_2 \exp(I_{k_1})$ and $p(k_2 \in S) = q(k_2 \in S)$, because the denominators of $p(k_2 \in S)$ and $q(k_2 \in S)$ are equal to the sum of their numerators across all $k_2 \in S$. We use the relation $J_{k_2} = \exp(I_{k_2})$ to write the preceding expression in either of the following two forms:

$$v_{k_2} = \ln\left\{C_2 \exp\left(\frac{\lambda_{k_2}}{\lambda_{k_1}} I_{k_2}\right) - \exp(I_{k_2})\right\},$$

and

$$\lambda_{k_2} = \frac{\lambda_{k_1}}{\ln J_{k_2}} \{\ln\{\exp(v_{k_2}) + J_{k_2}\} - \ln C_2\},$$

Finally, the probabilities of selecting node $k_1 \in S$, where $S = S(k_0) \in \Omega_1$, are given by the following expressions for the preference tree and nested logit model:

$$p(k_1 \in S) = \frac{\exp(v_{k_1}) + J_{k_1}}{J_{k_0}},$$
$$q(k_1 \in S) = \frac{C_1 \exp((\lambda_{k_1}/\lambda_{k_0})I_{k_1})}{C_1 \exp(I_{k_0})} = \frac{C_1 \exp(\lambda_{k_1} I_{k_1})}{C_1 \exp(I_{k_0})},$$

where $\lambda_{k_0} = 1$ and C_1 is a positive constant associated with the second level of the tree. Equating the numerators of $p(k_1 \in S)$ and $q(k_1 \in S)$ gives

$$\exp(v_{k_1}) + J_{k_1} = C_1 \exp(\lambda_{k_1} I_{k_1}), \quad \text{for all } k_1 \in S = S(k_0).$$

Since the denominators of $p(k_1 \in S)$ and $q(k_1 \in S)$ are equal to the sum of their numerators across all $k_1 \in S$, this condition implies that $J_{k_0} = C_1 \exp(I_{k_0})$, and thus $p(k_1 \in S) = q(k_1 \in S)$. We use the relation $J_{k_1} = C_2 \exp(I_{k_1})$ to write the preceding expression in either of the following two forms:

$$v_{k_1} = \ln\{C_1 \exp(\lambda_{k_1} I_{k_1}) - C_2 \exp(I_{k_1})\}, \quad \text{for all } k_1 \in S(k_0);$$

and

$$\lambda_{k_1} = \frac{1}{\ln J_{k_1} - \ln C_2}\{\ln\{\exp(v_{k_1}) + J_{k_1}\} - \ln C_1\}, \quad \text{for all } k_1 \in S(k_0).$$

The constants C_1 and C_2 can have any positive values for which $\lambda_{k_1}, \lambda_{k_2} \in (0,1]$, and the logarithms in the expressions for v_{k_1} and v_{k_2} have positive arguments.

2.4. Cross-Nested Logit Model

Cross-nested logit models relax the assumption that each alternative belongs to a single nest. These models have been developed only for trees with two levels. We consider the generalized nested logit (GNL) model by Wen and Koppelman (2001), which subsumes the cross-nested logit models by Papola (2004), Vovsha and Bekor (1998), and Vovsha (1997) as special cases. Bierlaire (2006) showed that GNL is a member of the family of generalized extreme value models. We show that it is a special case of EBA.

Recall that each terminal node $k_2 \in S(k_0, k_1)$ in a two-level tree corresponds to an alternative. GNL assumes that each alternative is a partial member of each set (nest) $S \in \Omega_2$. A parameter $\theta_k(S)$ characterizes the degree of membership of alternative k in nest S, where $\sum_{S\in\Omega_2} \theta_k(S) = 1$. Alternative k does not belong to nest S if $\theta_k(S) = 0$, and it belongs to only nest S if $\theta_k(S) = 1$. GNL associates the following deterministic utility with alternative k in nest $S \in \Omega_2$:

$$w_k(S) = w_k + \ln\{\theta_k(S)\}, \quad \text{for all } S = S(k_0, k_1) \in \Omega_2.$$

Observe that $w_k(S) \leqslant w_k$ because $\ln\{\theta_k(S)\} \leqslant 0$. The unconditional choice probability for alternative k is given by $\sum_{S\in\Omega_2} p_k(S)$, where $p_k(S)$ is the probability that alternative k in nest S is chosen by the preference tree. GNL reduces a cross-nested logit model to a nested logit model in which each alternative is a partial member of each nest. Thus, the transformation given in the preceding section allows the representation of a cross-nested logit model by a preference tree.

2.5. Extended Preference Trees

Nested logit models allow the deterministic utility components of the alternatives to be functions of covariates. We show that such a model is equivalent to an extension of a preference tree in which all aspect utilities are functions of covariates. Similarly, a preference tree in which the utilities of the alternatives are functions of covariates is equivalent to a nested logit model in which all the parameters, including the nesting parameters, are functions of covariates.

Consider a nested logit model in which alternative $k_n \in S$, $S \in \Omega_n$, has the deterministic utility

$$w_{k_n} = \boldsymbol{\beta}'_{k_n}\mathbf{z}_{k_n} = \beta_{0k_n} + \beta_{1k_n} z_{1k_n} + \cdots,$$

where $\mathbf{z}_{k_n} = (1, z_{1k_n}, \ldots,)'$ is a column vector of covariates, and $\boldsymbol{\beta}'_{k_n} = (\beta_{0k}, \beta_{1k}, \ldots)$ is a row vector of parameters, for alternative k_n. If the parameters are common across the alternatives, then $\boldsymbol{\beta}'_{k_n} = \boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots)$. As described in the proof of Theorem 2, an equivalent preference tree has the parameters

$$v_{k_n} = \frac{w_{k_n}}{\lambda_{k_{n-1}}}, \quad \text{for all } k_n \in S \text{ and } S \in \Omega_n, \quad \text{and}$$

$$v_{k_r} = \ln\left\{C_r \exp\left(\frac{\lambda_{k_r}}{\lambda_{k_{r-1}}} I_{k_r}\right) - C_{r+1}\exp(I_{k_r})\right\}, \quad \text{for all } k_r \in S \text{ and } S \in \Omega_r.$$

Since w_{k_n} is a function of covariates, $v_{k_n} = w_{k_n}/\lambda_{k_{n-1}}$ and

$$I_{k_{n-1}} = \ln\left\{\sum_{k\in S(k_0,\ldots,k_{n-1})} \exp\left(\frac{w_k}{\lambda_{k_{n-1}}}\right)\right\}$$

are also functions of covariates. The other I_{k_r} values are also functions of the covariates because they are recursively obtained by using the relation

$$I_{k_r} = \ln\left\{\sum_{k\in S(k_0,\ldots,k_r)} \exp\left(\frac{\lambda_k}{\lambda_{k_r}} I_k\right)\right\}.$$

Thus, if the utilities of the alternatives are functions of covariates in a nested logit model, then all aspect utilities are functions of covariates in an equivalent preference tree.

Similarly, consider a preference tree in which aspect $k_n \in S$, $S \in \Omega_n$, has the deterministic utility

$$v_{k_n} = \boldsymbol{\mu}'_{k_n}\mathbf{z}_{k_n} = \mu_{0k_n} + \mu_{1k_n} z_{1k_n} + \cdots,$$

where $\boldsymbol{\mu}'_{k_n} = (\mu_{0k_n}, \mu_{1k_n}, \ldots)$ denotes a row vector of parameters. Since k_n is a terminal node, it appears in only one alternative. If the parameters are common across alternatives, then $\boldsymbol{\mu}'_{k_n} = \boldsymbol{\mu}' = (\mu_0, \mu_1, \ldots)$. As described in the proof of Theorem 2, an equivalent nested logit model has the parameters $w_{k_n} = \lambda_{k_{n-1}} v_{k_n}$ and

$$\lambda_{k_r} = \frac{1}{\ln J_{k_r} - \ln C_{r+1}}\{\ln\{\exp(v_{k_r}) + J_{k_r}\} - \ln C_r\}, \quad \text{for all } k_r \in S,\ S \in \Omega_r \text{ and } r = 1, \ldots, n-1.$$

Since λ_{k_r} is a recursive function of v_{k_n}, each λ_{k_r} and I_{k_r} is also a function of the covariates.

To illustrate, consider the correspondence between a two-level preference tree and a nested logit model. Let $v_{k_2} = \boldsymbol{\mu}'_{k_2}\mathbf{z}_{k_2}$; that is, let the unique aspect of each alternative in the preference tree have a deterministic utility that is a linear function of covariates. Then the correspondence implies the following.

(1) The deterministic utilities of the alternatives in an equivalent nested logit model are functions of covariates:

$$w_{k_2} = \lambda_{k_1} v_{k_2} = \lambda_{k_1}\boldsymbol{\mu}'_{k_2}\mathbf{z}_{k_2} = \boldsymbol{\beta}'_{k_2}\mathbf{z}_{k_2}.$$

(2) The inclusive value parameters in a nested logit model are functions of covariates:

$$\lambda_{k_1} = \frac{\ln(e^{v_{k_1}} + J_{k_1}) - \ln C_1}{\ln J_{k_1}},$$

where

$$J_{k_1} = \sum_{k_2 \in S(k_0, k_1)} e^{v_{k_2}} = \sum_{k_2 \in S(k_0, k_1)} e^{\mu'_{k_2} \mathbf{z}_{k_2}}.$$

Thus, if we generate data from a preference tree with covariates, then a nested logit model will obtain a perfect fit to the data only if the inclusive value parameters are functions of covariates. Similarly, as shown above, if we generate data from a nested logit model with covariates, then a preference tree will obtain a perfect fit to the data only if all the aspect utilities are functions of covariates.

3. Relations Among Models

Figure 2 shows the relations between EBA and several other choice models.[3] We discuss these relations below.

(1) EBA is equivalent to a probabilistic lexicographic rule: to select an alternative, a person probabilistically arranges the aspects in decreasing order of their utilities, then uses them in a lexicographic rule. Tversky (1972a, b) described the conceptual correspondence between EBA and a probabilistic lexicographic rule. Kohli and Jedidi (2015) formalized it by showing that if the aspect utilities have extreme value distributions, then the set of lexicographic sequences can be partitioned into mutually exclusive and collectively exhaustive subsets, each corresponding to an EBA instance. Suppose an EBA problem is defined over n aspects, and that an EBA instance uses a subset of $m \leqslant n$ aspects to select an alternative A. Without loss of generality, let $1, 2, \ldots, m$ denote the aspect sequence used by the EBA instance. Let $\mathcal{S}$ denote the set of ordered sequences over all n aspects in which aspect i precedes aspect j, for all $1 \leqslant i < j \leqslant m$. Then all sequences $s \in \mathcal{S}$ choose alternative A by eliminating the same alternatives in the same sequence. The sum of the occurrence probabilities for the sequences in $\mathcal{S}$ is equal to the probability that the EBA instance chooses A.

(2) EBA generalizes the rank ordered logit model. If an EBA instance uses all n aspects to eliminate alternatives,[4] then it occurs with a probability given by the rank-ordered logit model (Beggs et al. 1981). More generally, as discussed above, the probability of an EBA instance is obtained by adding the probabilities of occurrence for a subset of aspect sequences, and each aspect sequence occurs with a probability given by the rank ordered logit model.

(3) A probabilistic lexicographic rule reduces to a deterministic lexicographic rule over a sequence of aspects, say $1, \ldots, n$, when $v_k - v_{k+1}$ is sufficiently large for all $k = 1, \ldots, n-1$. For example, suppose $v_1 = (n-1)v$ and $v_{k+1} = v_k - v$, for each $k = 2, \ldots, n$. Then

$$\frac{\exp(v_k)}{\sum_{j=k}^{n} \exp(v_j)} = \frac{1}{\sum_{j=k}^{n} \exp(v_j - v_k)} > \frac{1}{1 + (n-k)\exp(-v)}.$$

Thus, the probabilistic lexicographic rule selects the aspect sequence $1, \ldots, n$ with probability

$$\prod_{k=1}^{n-1} \frac{\exp(v_k)}{\sum_{j=k}^{n} \exp(v_j)} > \prod_{k=1}^{n-1} \frac{1}{1 + (n-k)\exp(-v)}.$$

The right-hand side of this expression, and thus the probability of choosing the aspect sequence $1, \ldots, n$, can be made to be arbitrarily close to one by choosing a sufficiently large value for $v > 0$.

(4) Extended elimination by aspects refers to an EBA model in which (i) each alternative has a unique aspect, and (ii) the utility of a unique aspect is a function of covariates. Thus, a unique aspect is a composition of covariates, including product attributes, that are associated with an alternative. In a preference tree, the unique aspects are represented by terminal nodes.

Extended EBA captures a consideration-then-choice process, but without separately formulating a consideration stage and a choice stage. Choosing a unique aspect at any stage of elimination corresponds to choosing an alternative. Any preceding elimination stages represent a consideration phase in which alternatives are eliminated using the shared aspects. The number of elimination stages in the consideration phase can differ from one choice occasion to another, as can the specific sequence of aspects used for eliminating alternatives. Only one error model is used, since all aspects, including the unique aspects, have independent random utilities with extreme value distributions.

(5) As discussed in Section 2, a preference tree is a special case of EBA. It is equivalent to a nested logit model in which the utilities of the alternatives are not functions of covariates. An extended preference tree is a special case of extended EBA. It is equivalent to a nested logit model in which the utilities of alternatives are functions of covariates. In the latter case, (i) the aspect utilities in a preference tree are functions of the covariates used in a nested logit model, and (ii) the inclusive value parameters in a nested logit model are functions of the covariates used in a preference tree.

There is also a conceptual difference between the extended preference tree and the nested logit model. An extended preference tree not only describes a two-phased decision process but also allows the same probabilistic elimination by aspects that is used by EBA. As a consequence, the number of steps in the elimination phase is not fixed. An alternative may be chosen at the first step, with a probability that depends on its utility and the utilities of all other alternatives and aspects; or it may be selected after one or more steps, with a probability that depends on its utility, and the utilities of all other surviving alternatives and aspects. A nested logit model, like the hierarchical elimination model, only allows a sequential elimination of alternatives. Tversky and Sattah (1979, p. 548) observed that the probabilistic elimination process associated with a preference tree is more likely to be used for decisions like choosing a restaurant or a movie, which have no fixed sequence of choice points (they call this free access). A sequential elimination process, described by nested logit and hierarchical elimination models, is more likely to be used for decisions that have a natural hierarchy of choice points (they call this sequential access). For example, sequential access occurs when a person first decides whether to travel by train or airplane, and only then evaluates the options in a schedule of trains or flights.

(6) EBA allows a distinct aspect to be associated with each subset of alternatives in a choice set. Thus, the number of parameters in an unrestricted EBA problem increases exponentially with the number of alternatives in a choice set, and the expressions for the EBA choice probabilities become increasingly more complex. In contrast, the number of parameters in a preference tree increases only linearly with the number of alternatives. The choice probabilities have simple expressions for a hierarchical elimination model, and thus for a preference tree, with any number of levels.

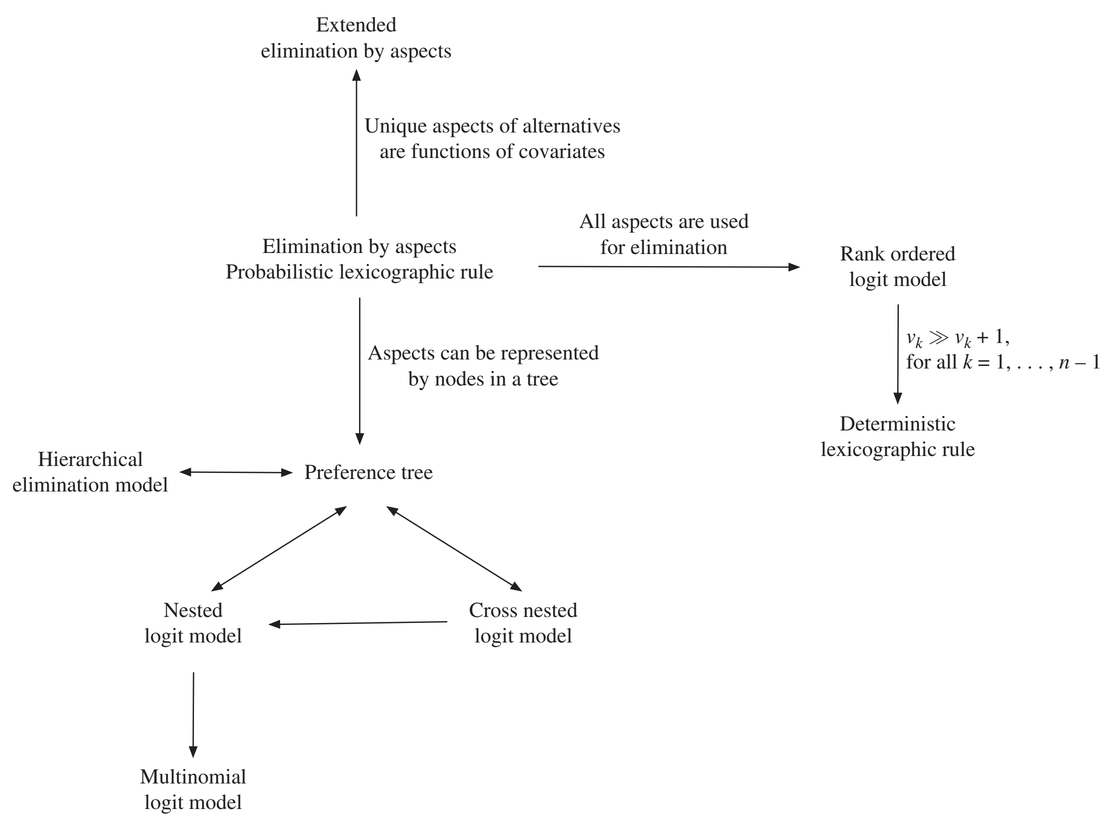
(7) A cross-nested logit model generalizes a nested logit model by allowing alternatives to be partial members of multiple nests. We showed in Section 2 that a cross-nested logit model can be represented by a preference tree. An advantage of the preference tree representation is that it does not require the assumption of a generalized extreme value distribution. A related benefit is that the latter error structure has been used to construct cross-nested logit models with two levels. In contrast, a preference tree representing a cross-nested logit model can be obtained for any number of levels by assuming that the aspect utilities have independent, extreme value distributions.

(8) It is well known that a nested logit model reduces to a multinomial logit model when each nesting parameter is equal to unity. A multinomial logit model in which the utilities of alternatives are functions of covariates is also a special case of an extended preference tree.

4. Estimation

The relations among the models shown in Figure 2 can be useful for both confirmatory and exploratory

Figure 2. Choice models related to elimination by aspects (EBA)



Notes. Extended elimination by aspects is the only model with covariates in this figure. Similar relations exist between extended EBA and version of the other models with covariates (except the rank-ordered logit model and the deterministic lexicographic model).

analysis of choice data. In a confirmatory context, a hypothesized preference structure can be estimated and compared to any more or less restricted specifications. In an exploratory context, an aspect can be associated with each subset of alternatives in a choice set. Likelihood ratio tests and model selection criteria, such as Akaike information criterion (AIC) and Bayesian information criterion (BIC), can be used to identify the aspects retained in the model. The interpretation of a retained aspect needs to be based on the commonalities among the alternatives in which it appears. As discussed below, this approach is practical only for problems with a small number of alternatives. Bentley and Seetharaman (2016) illustrated it for a problem with five brands, using data on the choices made by a panel of households.

As noted, an extended preference tree can be used for modeling a two-stage choice process. It also offers a compromise between the highly constrained preference tree and the unconstrained EBA. Like a preference tree, it can require fewer parameters than an unrestricted EBA model; like a nested logit model, it allows the utilities of the alternatives to be functions other additional aspects and covariates.

We now discuss (1) how the number of aspects affects estimation, (2) the identification of important aspects, (3) the representation of preference heterogeneity, and (4) issues concerning the data used for estimating EBA models.

Number of Aspects

An EBA model with m alternatives can have $O(2^m)$ parameters. Including all possible aspects is reasonable only when m is small; otherwise, too many parameters may need to be estimated. For example, a full EBA model has over one thousand parameters when there are $m = 10$ alternatives and over one million parameters when there are $m = 20$ alternatives. It may also not be necessary to estimate a full EBA model if the purpose is to assess the effects of manipulated variables, such as product attributes in a conjoint choice experiment, on consumer choice. In such cases, the aspects should be primarily defined in terms of the attributes.

Even for an exploratory model, it is better to begin with a simple nested logit model/preference tree and then add more aspects by (1) sequentially introducing more levels in a tree, (2) selectively adding aspects across nests, and (3) specifying the utilities of the alternatives to be functions of covariates. Observe that the model can be estimated even if the same aspects are used for nesting and as covariates, because the differences in the utilities of pairs of alternatives is unaffected by identical values of their nesting aspects. Typically, the deterministic utilities of alternatives will have alternative-specific constants. As in a nested logit model, the parameters of the covariates may or may not be constant across alternatives. Differences in likelihood values can be used for assessing the marginal benefit of adding one or more aspects and for comparing alternative preference trees. Likelihood ratio tests can be used to prune trees and compare alternative tree models. If useful, selected aspects across nests can then be added.

One advantage of a preference tree over a nested logit model is that the choice probabilities are simple functions of the deterministic aspect utilities. In contrast, the choice probabilities for a nested logit model are recursive functions of the inclusive values. For small problems, there may be no computational advantage of using one or the other model. But for problems with several nesting levels, the preference tree may be easier to specify and estimate.

Assessing Aspect Importance

Suppose two aspects, with deterministic utilities v_1 and v_2, are available at an elimination stage. The probability that aspect 1 is used for eliminating alternatives at the stage is $1/(1 + \exp(v_2 - v_1))$. Both v_1 and v_2 may be large negative values, but if $v_1 = v_2$, each attribute will be selected with equal probability. Thus, an aspect with a large negative, deterministic utility is not necessarily an unimportant aspect. It is better to assess the importance of an aspect by comparing the effect of its exclusion on the maximum likelihood value. This is feasible for the more general EBA model but not for a preference tree. In the latter case, removing an aspect also eliminates all alternatives in which it appears. For example, if "ground" and "air" are two aspects in a preference tree describing transportation choices, then neither aspect can be removed without eliminating alternatives. We can only test one preference tree against another.

Modeling Preference Heterogeneity

There are several ways of introducing heterogeneity in EBA. First, in a preference tree, the parameters of the utility functions may be allowed to be different for each $S \in \Omega_{n-1}$. This approach is practical only if $|\Omega_{n-1}|$ is small. Otherwise, the parameters may be allowed to differ only across the first two or three levels of the tree. Second, the utilities of the alternatives can be functions of demographic variables. Alternatively, the parameters associated with the attributes can be functions of demographic or other person-specific variables. Finally, random effects and latent class formulations can be used to represent unobserved heterogeneity in consumer preferences.

Data for EBA Models

EBA is useful when choices violate IIA or the weaker order-independence condition. A data set may show no violations of these conditions for two reasons. First, consumer preferences may be consistent with IIA and order independence. Second, preferences may violate the conditions but not in the choice sets used to estimate the model. For example, suppose we collect data for the red-bus-—blue-bus problem. A logit model will fit the data well if we only consider a choice set with two buses and a car. It will also fit the data well if we consider only one bus and one car. But a single logit model will not fit well if the data contain choices from both types of choice sets. Thus, it is better to collect data by fixing pairs of alternatives across multiple choice sets and varying the other alternatives in these sets. Any independence violations are then more likely to be captured in the data. Otherwise, there may be little difference in the fits obtained by a multinomial logit model and EBA, not because independence violations do not occur, but because the data do not have a sufficient number of such instances.

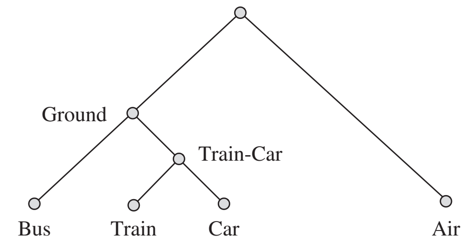
5. Application

We illustrate the approach to building and testing an EBA/preference tree model for a problem concerning the prediction of transportation choices by 210 nonbusiness travelers between Sydney, Canberra, and Melbourne (Louviere et al. 2000). A traveler could drive, fly, or ride a train or bus, to a destination. Several flights, trains, and buses, with different prices and travel times, were available to passengers traveling between each pair of cities. Some of the options had stops and transfers, which added to the total travel time.

Process of Model Construction

The number of alternatives across city pairs and travel modes was too large to construct a full EBA model. We started by estimating a nested logit model and a preference tree using intuitively reasonable groupings of alternatives. Both models used the covariates discussed below. Each tree partitioned the alternatives into two groups. The first tree had a public-private partition (bus-train and air-car), the second a ground-air partition (car-train-bus and air).[5] Both structures have been previously used to construct nested logit models (Louviere et al. 2000, Hensher and Greene 2002). We selected the tree with the ground-air partition based on a likelihood ratio test relative to the MNL model ($\chi^2(1) = 5.62$, $p < 0.05$ for nested logit model; and $\chi^2(1) = 45.67$, $p < 0.001$ for the preference tree). For the nested logit model, the public-private partition (bus-train and air-car) achieved better fit than the ground-air partition, but obtained inclusive value estimates that were larger than one (the estimate was much larger than one for the private nest). Constraining the inclusive values to lie between zero and one resulted in a substantial reduction in the likelihood value. A likelihood ratio test suggested that the constrained model was no different from the multinomial logit model ($\chi^2(2) = 0.87$; $p > 0.5$). For the preference tree, the ground-air partition had one less parameter, but significantly better fit, than the public-private partition.

Figure 3. Structure of preference tree



Next, we examined the three ways in which the ground alternatives could be partitioned into two nests: (1) car and bus-train, (2) train and bus-car, and (3) bus and train-car. We selected the last partition (bus and train-car), which is shown in Figure 3, based on the results of likelihood ratio tests ($\chi^2(2) = 10.22$, $p < 0.01$ for the nested logit model; and $\chi^2(2) = 47$, $p < 0.001$ for the preference tree). The nested logit model allows for the choice of bus only by the sequential selection of ground and bus, and the choice of train or car by the sequential selection of ground and train-car. The corresponding preference tree allows the bus to be chosen in two ways: (1) in one step, without eliminating any of the other three alternatives; and (2) after choosing the ground aspect. Similarly, it allows train or car to be chosen in three ways: (1) in one step, without eliminating any of the other three alternatives; (2) after choosing the train-car aspect in the first step; and (3) after choosing the ground aspect in the first step and the train-car aspect in the second step. As discussed below, these differences in the choice processes for the nested logit model and the preference tree lead to different interpretations and implications of the results.

Covariates

Travel cost, traveling time, and waiting time could be potentially used as aspects in a preference tree. We used them as covariates because the data set did not have information on the origin and destination cities for the travelers. Without this information, it is not meaningful to associate an aspect weight with a particular price or traveling/waiting time. For example, a \$120 air ticket may be inexpensive for a flight between Sydney and Canberra but expensive for a flight between Sydney and Melbourne (the latter pair of cities is more distant but typically has cheaper flights). For this reason, a single aspect weight cannot be associated with the travel cost across pairs of cities. Instead, we used travel cost and traveling/waiting times as covariates in the present analysis.

We also used the number of traveling companions and family income as covariates. These variables cannot be used as aspects because they are characteristics of travelers, not alternatives. We included them to model the effect of (observed) heterogeneity in traveler characteristics on the utilities of alternatives. Altogether, we used the following five covariates in our analysis:

(1) (in-vehicle) travel cost across all stages of a journey (in Australian dollars)

(2) the traveling time in a vehicle

(3) the waiting time at a terminal (equal to zero for car)

(4) the number of companions a traveler had on a trip (party size)

(5) the household income (in thousands of Australian dollars).

Estimation and Results

We estimated the preference tree and the corresponding nested logit model shown in Figure 3. For comparison, we also estimated a multinomial logit model. Ninety percent of the choice sets were randomly selected to estimate the models. The other ten percent were used for holdout validation. The procedure was repeated one hundred times. All parameter estimates were obtained using the full information maximum likelihood method implemented using the Proc NLP routine in SAS.

We began by confirming the equivalence of the preference tree and the nested logit model in the absence of covariates (that is, when the models had only intercept terms). Both models obtained identical log-likelihood values (LL = −283.76). The parameter estimates could be transformed from one model to the other in the previously described manner.

Using covariates improved the fits and predictions of all three models. Table 1 shows (1) the log-likelihood values, (2) the ρ^2 values, (3) the BIC values, and (4) the average in-sample and out-of-sample hit rates. The ρ^2 value, which measures the improvement in the log-likelihood value over the multinomial logit model, is 3% for the nested logit model and 13% for the (extended) preference tree. The BIC value, which penalizes for over-parametrization, indicates that the preference tree is the best model. It also obtains better hit rates than the nested logit model, both in sample (73.8% versus 81.3%) and out of sample (70.3% versus 81.6%).[6]

Table 2 shows the parameter estimates for the three models. For model identification, we set the constant for car to zero. Thus, all (including constants for ground and car-train constants in the preference tree model) should be interpreted relative to car. Following Louviere et al. (2000), we estimated the income parameters for only plane and train, and the party size parameter for only plane. All parameter estimates have the expected signs. The inclusive values for the nested logit model are between zero and one. Mostly the same variables are significant in all three models. In-vehicle (travel) cost is not significant, probably because it varies a great deal across pairs of cities.[7] The

Table 1. Model Performance Statistics

Model	No. of par.	Log likelihood	ρ^2	BIC	Average hit rate (%)	
					In-sample	Holdout
Multinomial logit	9	−174.22	—	396.56	73.3	73.3
Nested logit	11	−169.11	0.03	397.04	73.8	70.3
Preference tree	11	−150.74	0.13	360.30	81.3	81.6

Table 2. Parameter Estimates for Multinomial Logit, Nested Logit, and Preference Tree

Variable	Multinomial logit model		Nested logit model		Preference tree	
	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
Plane constant	**5.847**	1.128	**3.656**	1.411	1.728	1.293
Train constant	**5.613**	0.653	**4.064**	0.864	**5.375**	0.758
Bus constant	**3.626**	0.491	**2.931**	0.612	**3.471**	0.570
Ground constant	—	—	—	—	**-5.341**	1.081
Car-train constant	—	—	—	—	**-15.774**	2.560
In-vehicle cost	−0.009	0.008	−0.008	0.006	0.003	0.008
In-vehicle time	**-0.004**	0.001	**-0.004**	0.001	**-0.016**	0.002
Terminal time	**-0.102**	0.011	**-0.073**	0.015	**-0.088**	0.012
HH income (plane)	0.0140	0.012	0.016	0.011	**0.025**	0.011
HH income (train)	**-0.049**	0.014	**-0.034**	0.012	**-0.056**	0.017
Party size (plane)	**-0.973**	0.249	**-0.788**	0.235	**-0.644**	0.253
Inclusive value (ground)	—	—	**0.451**	0.115	—	—
Inclusive value (car-train)	—	—	**0.765**	0.205	—	—

Note. Numbers in bold are parameter estimates that are significant at the 5% level.

constant for plane is significant for the multinomial logit and nested logit models, but not for the preference tree. The parameter estimate for car-train is much smaller than for any other aspect in the preference tree, which implies that there is a very small probability that a person uses train-car to screen alternatives in the first stage. Household income has a significant and positive effect on air travel in the preference tree but not in the other two models. It also has a significant and negative effect on train travel in all three models. Both in-vehicle time and terminal time have significant, negative, effects.

Comparison of Preference Structures

The preference tree and the nested logit model predict similar total shares of choices for the four travel modes, but assume different choice processes. The nested logit model assumes hierarchical elimination, and the preference tree nonhierarchical elimination, of the alternatives.

Table 3 compares the choice processes implied by the two models. It decomposes the choice probabilities into direct and indirect components. Direct probabilities correspond to choices made from a set with all four alternatives. Indirect probabilities correspond to choices made after eliminating one or more alternatives. In both models, air travel can only be chosen directly, with a predicted choice probability equal to 27.62%. The preference tree additionally predicts direct choice probabilities of 13.79% for train, 4.39% for bus and 10.51% for car. Thus, the preference tree predicts an additional 13.79 + 4.39 + 10.51 = 28.69% of direct choices for train, bus, and car, from a choice set consisting of all four alternatives.

Indirect choices after the elimination of air account for 44% of the choices for the preference tree and 62.38% of the choices for the nested logit model. There are no consumers who directly screen for train-car at the first step (this occurs by definition for the nested logit model and has a probability of only 0.0013% for the preference tree). Thus, in both models, all consumers who do not make a direct choice first screen using ground as an aspect. In the preference tree, 96.9% of all choices are then made directly by choosing among train, bus, and car. In contrast, the nested logit model predicts that only 13.95% of all choices (for bus) are made directly after the ground aspect is selected. It requires train and car to be chosen only after the elimination of bus.

Thus the preference tree and nested logit models provide substantially different decompositions of the choice probabilities because they assume different choice processes. Which model provides a more accurate description of the actual choice process cannot be assessed based on the fits and predictions alone, although the statistics in Table 1 favor the preference tree in the present application. Such an assessment requires additional process-tracing information.

Elasticities for Travel Time

Table 4 shows the own and cross elasticities for travel time, for each of the four alternatives. The first row in the panel labeled "preference tree" shows that a 1% reduction in the travel time for air increases the choice probability for air by 0.91%, and decreases the choice probabilities for train, bus, and car by 0.28%, 0.40%, and 0.39%, respectively. The second row shows the effect of a 1% reduction in the travel time for train on the percentage changes in the choice probabilities for each of the four alternatives. The other rows have similar interpretations.

The data in Table 4 shows three discernible patterns. First, in all three models, the own price elasticities are

Table 3. Comparison of Preference Structure: A Decomposition of Preference Tree and Nested Logit Model Choice Probabilities (%)

Preference tree					
		Ground (44%)			
Alternative	Direct (66%)	Direct (96.9%)	Train-car (3.1%)	Train-car (0.0013%)	Total
Air	27.62	—	—	—	27.62
Train	13.79	14.26	1.02	0.0006	29.08
Bus	4.39	9.62	—	—	14.01
Car	10.51	17.55	1.24	0.0007	29.29

Nested logit model				
		Ground (62.38%)		
Alternative	Direct (27.62%)	Direct (13.95%)	Train-car (86.05%)	Total
Air	27.62	—	—	27.62
Train	—	—	29.91	29.91
Bus	—	13.95	—	13.95
Car	—	—	28.53	28.53

Table 4. Own and Cross Elasticities for Waiting Time, for Each of the Four Alternatives

	Air	Train	Bus	Car
		Preference tree		
Air	0.91	−0.28	−0.40	−0.39
Train	−0.18	2.63	−1.80	−1.58
Bus	−0.08	−1.00	4.35	−1.01
Car	−0.28	−1.49	−1.74	2.58
		Nested logit		
Air	0.25	−0.07	−0.11	−0.12
Train	−0.36	1.15	−0.73	−0.49
Bus	−0.27	−0.34	2.12	−0.42
Car	−0.52	−0.43	−0.80	1.34
		Multinomial logit		
Air	0.24	−0.06	−0.10	−0.11
Train	−0.33	0.95	−0.48	−0.44
Bus	−0.25	−0.23	1.53	−0.29
Car	−0.51	−0.38	−0.52	1.17

the highest for bus, followed by train and car, and then air. Second, the magnitudes of the own elasticities are substantially higher for the preference tree than for the nested logit and multinomial logit models. That is, the preference tree implies that the market is much more responsive to marginal reductions in travel time than is suggested by the two logit models. The own elasticity for bus is more than twice as large in the preference tree (4.65) than in the nested logit model (2.12), and three times as large than in the multinomial logit model (1.53). Even the own elasticity for air, which is in a separate branch in the preference tree and nested logit models, is almost four times as large for the preference tree (0.91) than for the nested logit model (0.25) and the multinomial logit model (0.24). Third, the magnitudes of the cross elasticities are different for the preference tree and the other two logit models. For example, the preference tree predicts that a 1% reduction in travel time for bus has virtually no effect on the demand for air; the nested logit model predicts a decline in the choice probability for air of 0.27%, which is close to the 0.25% decline predicted by the multinomial logit model. On the other hand, the preference tree predicts a larger increase in the share for train when its travel time is reduced by 1% than does a nested logit model (2.63% versus 1.15%). The resulting loss in the share for air travel is smaller for the preference tree than the nested logit model (0.18% versus 0.36%). The overall pattern is that the elasticities for the nested logit model are quite similar to those for the multinomial logit model, but different from those for the preference tree.

Implications

The choice processes assumed by the nested logit model and the preference tree have different implications for the structure of competition in the market. The

Table 5. Effect of 10% Less Travel Time for Train on Choice Probabilities

		Change in average choice probability (%)		
Alternative	Base choice probability (%)	Multinomial logit	Nested logit	Preference tree
Plane	27.60	−0.96	−1.07	−0.57
Train	30.00	2.96	3.62	8.42
Bus	14.30	−0.72	−1.07	−2.80
Car	28.10	−1.28	−1.48	−5.05

preference tree suggests far greater direct competition among all four alternatives than does the nested logit model. It also implies that the three ground alternatives compete directly against each other for the most part. In contrast, the nested logit model assumes that train and car compete directly with each other, and as a group against bus.

To further assess the differences in the implications of the two models, we simulated the effect of a 10% reduction in the travel time for train on the choice probabilities of the four alternatives. Table 5 compares the predicted changes in these probabilities for the multinomial logit model, the nested logit model, and the preference tree. The multinomial logit and nested logit models predict increases of 2.96% and 3.62% in the share of train travel. The preference tree predicts a much larger increase of 8.42% in share. It also predicts that 60% ($5.05 \times 100/8.42$) of the share gain for train is attained at the expense of car travel. The corresponding share gain from car is about 43% for both the multinomial logit and nested logit models.

Table 6 shows the ratios of the choice probabilities for air and bus, and air and car, before and after the 10% change in the rail travel time. The ratios are approximately equal across the three models before the travel time change. After the change, they are substantially higher for the preference tree than for the other two models. The most noticeable change is in the ratio of the choice probabilities for air and bus travel. In both the nested logit model and the preference tree, a shorter train journey reduces the choice probability for air to a lesser extent than it does the choice probabilities of bus and car. But the effect is more pronounced in the preference tree, which implies a greater violation of IIA than does the nested logit model.

Table 6. Effect of 10% Less Travel Time For Train on Probability Odds

Probability odds	Multinomial logit		Nested logit		Preference tree	
	Before	After	Before	After	Before	After
Air/bus	1.93	1.96	1.98	2.06	1.97	2.41
Air/car	0.98	0.99	0.97	0.98	0.94	1.12

6. Conclusion

Elimination by aspects (EBA) is a choice model that allows violations of order independence. It generalizes the multinomial logit and rank ordered logit models (Kohli and Jedidi 2015). This paper shows that it also generalizes nested logit and cross-nested logit models. EBA can be extended to represent a consideration-then-choice process. Preference trees with covariates are a special case of the extended model. We illustrated the process of developing a preference tree with covariates, compared it with a nested logit model, decomposed the choice probabilities into direct and indirect components, and assessed the market response to changes in product offerings.

Acknowledgments

The authors thank Yanyan Li for help with the data analysis. They also thank two anonymous reviewers and the editor for their encouragement and helpful comments.

Appendix

A.1. Proof of Theorem 2

The probabilities of selecting node $k_r \in S$, where $S = S(k_0, \ldots k_{r-1}) \in \Omega_r$, have the following expressions for a preference tree and a nested logit model:

$$p(k_r \in S) = \frac{\exp(v_{k_r}) + J_{k_r}}{J_{k_{r-1}}},$$

$$q(k_r \in S) = \frac{C_r \exp((\lambda_{k_r}/\lambda_{k_{r-1}}) I_{k_r})}{C_r \exp(I_{k_{r-1}})}, \quad \text{where } C_r > 0.$$

In the above expressions,

$$J_{k_{r-1}} = \sum_{k \in S(k_0,\ldots,k_{r-1})} \{\exp(v_k) + J_k\};$$

and

$$I_{k_{r-1}} = \ln\left\{\sum_{k \in S(k_0,\ldots,k_{r-1})} \exp\left(\frac{\lambda_k}{\lambda_{k_{r-1}}} I_k\right)\right\}$$

is the inclusive value associated with node $k_{r-1} \in S(k_0, \ldots, k_{r-2})$, where $S(k_0, \ldots, k_{r-2}) \in \Omega_{r-1}$ for all $r = 1, \ldots, n$.

Observe that we have multiplied and divided the numerator and denominator of $q(k_r \in S)$ by a constant C_r, which is associated with level r of the tree. Let $C_n = 1$. We obtain conditions on the other values of C_r, $r = 1, \ldots, n-1$, below.

Let the numerators of $p(k_r \in S)$ and $q(k_r \in S)$ be equal; that is, let

$$\exp(v_{k_r}) + J_{k_r} = C_r \exp\left(\frac{\lambda_{k_r}}{\lambda_{k_{r-1}}} I_{k_r}\right), \quad \text{for all } k_r \in S,\ S \in \Omega_r,\ r = 1, \ldots, n, \tag{A.1}$$

where $\lambda_{k_n} = \lambda_{k_0} = 1$, $I_{k_n} = w_{k_n}$ and $J_{k_n} = 0$. Then

$$J_{k_{r-1}} = C_r \exp(I_{k_{r-1}}), \quad \text{for all } k_{r-1} \in S(k_0, \ldots, k_{r-2}),\ r = 2, \ldots, n,$$

because the denominators of $p(k_r \in S)$ and $q(k_r \in S)$ are equal to the sum of their numerators across all $k_r \in S$. It follows that $p(k_r \in S) = q(k_r \in S)$ if the condition in Equation (A.1) is satisfied. To obtain a mapping of the parameters of a nested logit model onto the parameters of a preference tree, we rewrite Equation (A.1) as

$$v_{k_r} = \ln\left\{C_r \exp\left(\frac{\lambda_{k_r}}{\lambda_{k_{r-1}}} I_{k_r}\right) - J_{k_r}\right\}, \quad k_r \in S,\ S \in \Omega_r, \text{ for all } r = 1, \ldots, n-1.$$

Substituting $J_{k_r} = C_{r+1} \exp(I_{k_r})$ gives

$$v_{k_r} = \ln\left\{C_r \exp\left(\frac{\lambda_{k_r}}{\lambda_{k_{r-1}}} I_{k_r}\right) - C_{r+1} \exp(I_{k_r})\right\}, \quad \text{for all } k_r \in S, \text{ where } S \in \Omega_r \text{ and } r = 1, \ldots, n-1,$$

where $S = S(k_0, \ldots, k_{r-1})$. The values of C_r, $r = 1, \ldots, n-1$, are arbitrary except that they must ensure that the logarithm on the right-hand side is defined. For $r = n$, we have $J_{k_n} = 0$, $I_{k_n} = w_{k_n}$ and $C_n = 1$, which gives

$$v_{k_n} = \frac{w_{k_n}}{\lambda_{k_{n-1}}}, \quad k_n \in S,\ S \in \Omega_n.$$

To obtain a mapping of the parameters of a preference tree onto the parameters of a nested logit model, we rearrange A.1 to obtain

$$\lambda_{k_r} = \frac{\lambda_{k_{r-1}}}{I_{k_r}} \{\ln\{\exp(v_{k_r}) + J_{k_r}\} - \ln C_r\}.$$

We use the relation $J_{k_r} = C_{r+1} \exp(I_{k_r})$ to substitute $I_{k_r} = \ln J_{k_r} - \ln C_{r+1}$ in the preceding expression to obtain

$$\lambda_{k_r} = \frac{1}{\ln J_{k_r} - \ln C_{r+1}} \{\ln\{\exp(v_{k_r}) + J_{k_r}\} - \ln C_r\}, \quad \text{for all } k_r \in S, \text{ where } S \in \Omega_r \text{ and } r = 1, \ldots, n-1.$$

Since $J_{k_n} = 0$, $I_{k_n} = w_{k_n}$ and $C_n = 1$, we obtain $w_{k_n} = \lambda_{k_{n-1}} v_{k_n}$, for all $k_n \in S$, where $S \in \Omega_n$.

Endnotes

[1] Order independence is a weaker condition than IIA. The latter requires the ratio of the choice probabilities for two alternatives to be independent of the other alternatives in a choice set. Order independence requires the ordering, but not necessarily the ratio, of the choice probabilities for two alternatives to be independent of other alternatives in a choice set.

[2] If necessary, we can extend a path using dummy nodes, each of which is selected with probability one if its preceding aspect is selected.

[3] We thank a reviewer for suggesting the diagram shown in Figure 2.

[4] This can occur only if there are at least n alternatives in a choice set.

[5] We first searched for the best partition of the four modes of transportation from a total of 10 possible partitions, obtained by placing one or two distinct alternatives in one partition and the rest in the other partition. Based on likelihood ratio tests, the two best partitions that emerged were (i) public (train-bus) and private (air-car), and (ii) ground (train-bus-car) and air.

[6] We also estimated a cross-nested logit model with two nests. We do not report the results because we encountered convergence issues in the maximum likelihood estimation of this model. Some of the parameter estimates had no standard errors, and the gradient values at the optimum were not equal to zero.

[7] As noted previously, we do not have information matching costs with city pairs.

References

Batley R, Daly A (2006) On the equivalence between elimination-by-aspects and generalized extreme value models of choice behavior. *J. Math. Psych.* 50(5):456–467.

Beggs S, Cardell S, Hausman J (1981) Assessing the potential demand for electric cars. *J. Econometrics* 17(1):1–19.

Bentley T, Seetharaman PB (2016) Identifying unobserved similarity: Estimating a generalized EBA model on scanner data. Working paper, University of Texas at Austin, Austin, TX.

Bierlaire M (2006) A theoretical analysis of the cross nested logit model. *Ann. Oper. Res.* 144(1):287–300.

Dagsvik JK (1995) How large is the class of generalized extreme value random utility models? *J. Math. Psych.* 39(1):90–98.

Hensher DA, Greene WH (2002) Specification and estimation of the nested logit model: Alternative normalisations. *Transportation Res. Part B: Methodological* 36(1):1–17.

Kohli R, Jedidi K (2015) Error theory for elimination by aspects. *Oper. Res.* 63(3):512–526.

Louviere JJ, Hensher DA, Swait J (2000) *Stated Choice Methods: Analysis and Applications in Marketing: Transportation and Environmental Valuation* (Cambridge University Press, Cambridge, UK).

McFadden D (1981) Econometric models of probabilistic choice. Manski C, McFadden D, eds. *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MA), 198–272.

Papola A (2004) Some developments on the cross nested logit model. *Transportation Res. Part B: Methodological* 38(9):833–851.

Tversky A (1972a) Elimination by aspects: A theory of choice. *Psych. Rev.* 79(4):281–299.

Tversky A (1972b) Choice by elimination. *J. Math. Psych.* 9(4):341–367.

Tversky A, Sattah S (1979) Preference trees. *Psych. Rev.* 86(6):542–573.

Vovsha P (1997) Application of cross nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Res. Record* 1607:6–15.

Vovsha P, Bekor S (1998) Link-nested logit model of route choice: Overcoming route overlapping problem. *Transportation Res. Record* 1645:133–142.

Wen CH, Koppelman FS (2001) The generalized nested logit model. *Transportation Res. Part B: Methodological* 35(7):627–641.

Rajeev Kohli is the Ira Leon Rennert Professor of Business at Columbia Business School, New York. His research interests are in choice models, product design, emerging markets, combinatorial optimization and algorithms. He has published research on these topics in leading journals in marketing, operations research, mathematical programming, mathematical psychology and discrete mathematics.

Kamel Jedidi is the John Howard Professor of Business and the Chair of the Marketing Division at Columbia Business School, New York. He has extensively published in leading marketing and statistical journals. His research interests include pricing, product positioning, and market segmentation. He was awarded the 1998 IJRM Best Article Award and the Marketing Science Institute 2000 Best Paper Award. He was also finalist for 2009 Paul Green Award for the Journal of Marketing Research and for the 2009 Long-term Impact Paper Award for Marketing/Management Science.