

Representation and Inference of Lexicographic Preference Models and Their Variants

Rajeev Kohli

Graduate School of Business, Columbia University, 506 Uris Hall, New York, New York 10027,
rk35@columbia.edu

Kamel Jedidi

Graduate School of Business, Columbia University, 518 Uris Hall, New York, New York 10027,
kj7@columbia.edu

The authors propose two variants of lexicographic preference rules. They obtain the necessary and sufficient conditions under which a linear utility function represents a standard lexicographic rule, and each of the proposed variants, over a set of discrete attributes. They then: (i) characterize the measurement properties of the parameters in the representations; (ii) propose a nonmetric procedure for inferring each lexicographic rule from pairwise comparisons of multiattribute alternatives; (iii) describe a method for distinguishing among different lexicographic rules, and between lexicographic and linear preference models; and (iv) suggest how individual lexicographic rules can be combined to describe hierarchical market structures. The authors illustrate each of these aspects using data on personal-computer preferences. They find that two-thirds of the subjects in the sample use some kind of lexicographic rule. In contrast, only one in five subjects use a standard lexicographic rule. This suggests that lexicographic rules are more widely used by consumers than one might have thought in the absence of the lexicographic variants described in the paper. The authors report a simulation assessing the ability of the proposed inference procedure to distinguish among alternative lexicographic models, and between linear-compensatory and lexicographic models.

Key words: lexicographic preferences; noncompensatory preference models; linear models; optimization techniques; greedy algorithm; approximation algorithms; utility theory; conjoint analysis; hierarchical clustering; market segmentation; hierarchical market structures

History: This paper was received January 18, 2005, and was with the authors 11 months for 3 revisions; processed by Wayne DeSarbo.

1. Introduction

A lexicographic rule orders alternatives over attributes in the same way that a dictionary orders words over letters. A consumer using the rule evaluates alternatives first on the most important attribute, and if there are ties, on the secondmost important attribute, and so forth. For example, a person buying a personal computer (PC) displays lexicographic preferences if he or she strictly prefers Microsoft Windows to any other operating system; among Windows-based systems, he or she always prefers PCs equipped with the Intel's latest microprocessor; and then uses, in sequence, brand name, memory, and hard-disk space to break ties among still-tied alternatives.

A lexicographic rule only orders alternatives. It therefore does not require that consumers make, or be able to make, cardinal (interval-scaled) judgments. However, it demands stronger judgments than are necessary for binary classifications, such as those obtained using conjunctive, disjunctive, or subset-conjunctive rules (Gilbride and Allenby 2004, Jedidi and Kohli 2005).

There is good evidence that people use lexicographic rules. Drolet and Luce (2004) note that consumers use them when they have emotional reasons to avoid trade-offs. Slovic (1975) reports the use of lexicographic rules for breaking ties among equally valued alternatives. Tversky et al. (1988) examine the rules that consumers use for choice and matching tasks involving public policies, job applicants, and benefit plans. Although each alternative in their study is described using only two attributes, they find that choice is more lexicographic than matching. Yee et al. (2007) report that approximately two-thirds of their subjects use lexicographic rules for evaluating Smart-Phones. A growing literature in psychology also documents the use of these rules in the formation of judgments over perceptual cues (e.g., Gigerenzer et al. 1991; Martignon and Hoffrage 1999, 2002; Broder 2000). For more evidence of, and details about, the use of lexicographic rules in consumer research, we refer the reader to Colman and Stirk (1999), Dhar and Nowlis (1999), Roedder-John (1999), Gonzalez-Vallejo et al. (1996), Kahn and Baron (1995), Westenber and Koele (1994), Ford et al. (1989), and Payne et al. (1988).

In contrast with the above literature characterizing the use of lexicographic rules, there is only limited work in marketing and consumer research on the mathematical representation, inference, and testing of lexicographic preference structures. A result known at least since Debreu (1954) says that it is impossible to construct a utility function representing lexicographic preferences over two or more real-valued attributes.¹ However, lexicographic utility functions can exist over discrete attributes. Martignon and Schmitt (1999) give a numerical sequence for representing lexicographic preferences over binary attributes. Martignon and Hoffrage (2002) prove that a sequence of weights w_1, \dots, w_n represents a lexicographic order over binary cues x_1, \dots, x_n , if $w_i > w_{i+1} + \dots + w_n$, for all $i = 1, \dots, n - 1$. In particular, they prove that if $w_i = 1/2^i$, then a lexicographic ordering of cues is identical to the ordering of the scores produced by the function $w_1x_1 + \dots + w_nx_n$. Kohli (1999) observes that all number systems are lexicographic, and can be used to represent, among other structures, lexicographic preferences over discrete attributes. He shows that number systems in which the base (radix) changes from one digit to the next can represent alternatives defined over attributes with varying numbers of levels. Substantial theoretical research has also examined conditions under which continuous lexicographic utility functions can exist (e.g., Fishburn 1974, 1975), the possibility of representing such preferences by multiple functions (e.g., Bridges 1983, Chateaufneuf 1987, Wakker 1988, Knoblauch 2000), and the formulation of models for probabilistic lexicographic preferences (e.g., Tversky 1972, Manrai and Sinha 1989).

The present paper obtains the general conditions under which a linear model is necessary and sufficient for representing lexicographic preferences over discrete and/or finitely divisible attributes. The representations described by Martignon and Schmitt (1999) and Kohli (1999) are obtained as special cases. We show that these representations—indeed, all representations of lexicographic preferences by a linear model over a finite number of discrete attributes—are not unique, in the sense that they allow a larger set of transformations for the parameter values than are permitted in a linear model representing interval-scaled preferences. We obtain the associated invariance conditions for the parameters. We then introduce two

variants of lexicographic preference models. We call these satisficing and binary lexicographic models.² Consumers can use the two variants separately or together. A linear utility function continues to be sufficient for representing each of these variants. We describe how each of the lexicographic models are related to each other, and to a linear compensatory model, in a partially nested structure. We propose a method for inferring each of the four lexicographic preference models (standard, satisficing, binary, and binary-satisficing) from ranking or paired-comparisons data. Computationally, the proposed procedure is substantially less demanding than an enumeration of all possible lexicographic orderings of the attributes and levels. We then propose a method for assigning a linear model, or one of the lexicographic models, to a consumer based on a sequence of tests comparing nested models. Finally, we examine how individual-level lexicographic rules can be combined to construct hierarchical clusters (segments) and aggregate market structures. There is a substantial marketing literature on market structure analysis using brand-switching data (e.g., Grover and Srinivasan 1987, Kannan and Wright 1991, Russell and Kamakura 1994, Urban et al. 1984). To our knowledge, the present approach is the first to use conjoint data for constructing aggregate, hierarchical market structures.

Organization of the Paper. Section 2 discusses the representation of lexicographic preferences over discrete attributes by a linear utility function, and discusses the invariance properties of the parameters. Section 3 describes the two variants of lexicographic preferences and their representation by utility functions. Section 4 examines the inference of lexicographic preference structures from paired-comparisons data, which can be directly elicited from consumers or inferred from rankings or ratings of multiattribute alternatives. Section 5 reports empirical tests of the alternative lexicographic models; it also describes the proposed approach for constructing hierarchical clusters (segments) and aggregate market structures using conjoint data. Section 6 presents the results of a simulation designed to assess the accuracy of the proposed algorithm for identifying the lexicographic rules. It examines the accuracy of a procedure for distinguishing among alternative lexicographic rules, and between lexicographic and linear models. It also reports the computational efficiency of the algorithm used for inferring the lexicographic rules.

¹ A formal proof for this result is available in Varian (1984). An intuitive explanation is as follows. Every multiattribute utility function is associated with indifference curves in the attribute space. These curves determine the marginal rate of substitution between pairs of attributes. A lexicographic rule has an infinite marginal rate of substitution. It therefore allows no indifference curves, and no utility function, over two or more real-valued attributes.

² While writing this paper, it has come to our knowledge that Yee et al. (2007) have also independently examined the binary variant, which they call “lexicographic by aspect.”

2. Representation Over Finite Attributes

Let $m \geq 2$ denote the number of attributes. We assign the integers $1, \dots, m$, to the attributes in decreasing order of their importance to a given consumer. Let attribute k have $n_k \geq 2$ levels, for all $k = 1, \dots, m$. We arrange the levels of attribute k in increasing preference order, sequentially assigning to them the values

$$x_k = a_1^k, \dots, a_{n_k}^k,$$

where

$$0 \leq a_1^k < \dots < a_{n_k}^k, \quad \text{for all } k = 1, \dots, m.$$

Let

$$X_k = \{a_1^k, \dots, a_{n_k}^k\}$$

denote the (ordered) set of values assigned to the levels of attribute k .

Let M denote a set of alternatives. Let \mathbf{x} and \mathbf{x}' denote two alternatives in M . Let x_k and x'_k denote the k th elements of \mathbf{x} and \mathbf{x}' , respectively. We say that \mathbf{x} is lexicographically preferred to \mathbf{x}' if there exists some k , $1 \leq k \leq m$, such that

$$x_k > x'_k \quad \text{and} \quad x_l = x'_l, \quad \text{for all } l = 1, \dots, k - 1. \quad (1)$$

That is, \mathbf{x} is preferred to \mathbf{x}' on the most important attribute for which the two alternatives are not equally preferred.

Let

$$\min \Delta x_k = \min \{a_j^k - a_{j-1}^k \mid j = 2, \dots, n_k\}$$

denote the smallest difference in the successive values of x_k , and let

$$\max \Delta x_k = a_{n_k}^k - a_1^k$$

denote the difference between the largest and smallest values of x_k , for all $k = 1, \dots, m$. The following theorem characterizes the necessary and sufficient conditions under which a linear model represents lexicographic preferences. A proof for the theorem appears in the appendix.

THEOREM 1. *Let*

$$u(\mathbf{x}) = \beta_1 x_1 + \dots + \beta_m x_m, \quad (2)$$

where x_k is the k th-most important attribute, $k = 1, \dots, m$. Then $u(\mathbf{x})$ represents lexicographic preferences over the m attributes if, and only if,

$$\beta_k \min \Delta x_k > \sum_{j=k+1}^m \beta_j \max \Delta x_j, \quad \text{for all } k = 1, \dots, m. \quad (3)$$

We illustrate the main point of Theorem 1 with an example using $m = 3$ attributes, each with $n_k = 3$ levels. For each attribute k , we can select any increasing sequence of nonnegative values a_1^k, a_2^k , and a_3^k . Suppose we set $a_1^k = a_1, a_2^k = a_2$, and $a_3^k = a_3$. Then

$$\min \Delta x_k = \min \{a_2 - a_1, a_3 - a_2\} \quad \text{and} \quad \max \Delta x_k = a_3 - a_1.$$

Let x_1 denote the most important attribute, x_2 the secondmost important attribute, and x_3 the least important attribute. Consider the utility function

$$u(x_1, x_2, x_3) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \quad (4)$$

where $\beta_k > 0$ and $x_k \in \{a_1, a_2, a_3\}$, for $k = 1, 2, 3$. Theorem 1 says that (4) represents lexicographic preferences if, and only if, the following conditions are satisfied:

$$\beta_2 \min \Delta x_2 > \beta_3 \max \Delta x_3; \quad \text{and} \quad (5)$$

$$\beta_1 \min \Delta x_1 > \beta_2 \max \Delta x_2 + \beta_3 \max \Delta x_3. \quad (6)$$

Condition (5) requires that the smallest change in utility for attribute 2 be no smaller than the maximum possible change in utility obtained by changing x_3 . Condition (6) requires that the smallest change in utility for attribute 1 be no smaller than the maximum possible change in utility obtained when both x_2 and x_3 are changed at the same time. If satisfied, these conditions ensure that an alternative with a preferred level of attribute 1 has a higher utility value, regardless of the levels of the less-preferred attribute.

Note that the specific values of a_1, a_2, a_3 play no role in the above example. We can, if we wish, choose these values from any increasing sequence of nonnegative, real numbers, and then constrain the β_k values so that the conditions in (5) and (6)—more generally, the conditions in (3)—are satisfied. For example, suppose we choose $a_1 = 0, a_2 = 10$, and $a_3 = 99$ in the above example, where $k = n_k = 3$. Then $\min \Delta x_1 = \min \Delta x_2 = a_2 - a_1 = 10, \max \Delta x_2 = \max \Delta x_3 = a_3 - a_1 = 99$, and so the utility function (4) represents lexicographic preferences if $10\beta_2 > 99\beta_3$ (i.e., $\beta_2 > 9.9\beta_3$) and if $10\beta_1 > 99\beta_2 + 99\beta_3$ (i.e., $\beta_1 > 9.9(\beta_2 + \beta_3)$). The important thing to note is that this is a weaker relationship between the parameters than is obtained in a linear model representing interval-scaled preferences, where for any fixed measurement scales for x_k and x_l , the ratio β_k/β_l must be a constant, for all $k, l = 1, 2, 3$. The reason for this difference, which we discuss in more detail below, is that (2) and (3) specify an ordinal utility function. As a result, the β parameters permit not only multiplicative transformations, which alone are allowed for cardinal (interval-scaled) utility functions, but the larger class of transformations described by (3). We therefore caution against interpreting the values of $\beta_k x_k$ as the part worths in

a conjoint model. Loosely put, one cannot extract the same information from the parameters of an ordinal, lexicographic utility function as one can from the parameters of a cardinal utility function. To further illustrate this point, consider the following two utility functions, each of which represents the same (lexicographic) ordering of alternatives:

$$u = \frac{x_1}{n_1} + \dots + \frac{x_m}{n_1 \dots n_m},$$

where $x_k \in \{0, \dots, n_k - 1\}$, for all $k = 1, \dots, m$; (7)

and

$$u = \frac{x_1}{2^{n_1}} + \dots + \frac{x_m}{2^{n_1} \dots 2^{n_m}},$$

where $x_k \in \{0, \dots, 2^{n_k} - 1\}$, for all $k = 1, \dots, m$. (8)

The successive, within-attribute values of $\beta_k x_k$ differ in (7) by the same amount, $1/(n_1 \dots n_k)$; but they differ in (8) by the unequal amounts, $2^{j-1}/(2^{n_1} \dots 2^{n_k})$, for all $j = 1, \dots, n_k$. Thus, a part-worths interpretation for the $\beta_k x_k$ is not appropriate for a lexicographic utility function.

Kohli (1999) notes that (7) represents a number system in which the base (radix) changes from one digit to the next, and gives the example of time measurements (hours, minutes, seconds) to illustrate its use in an actual measurement system. He also shows how lexicographic structures can be represented by wave functions. Martignon and Hoffrage (2002) examine the special case of (7) in which $n_i = 2$, for all $i = 1, \dots, m$. They prove that (7) assigns values to alternatives so that these are always lexicographically ordered; and that any lexicographic ordering over binary cues can be represented by the function $w_1 x_1 + \dots + w_m x_m$, provided $w_i > w_{i+1} + \dots + w_m$, for all $i = 1, \dots, m - 1$; a special case of this function is $x_1/2 + \dots + x_m/2^m$.

Implications for Model Estimation. The fact that (2) and (3) specify an ordinal utility function for lexicographic preferences has implications for how we estimate the parameters of the model. We cannot use least-squares regression to infer lexicographic rules because the method requires interval-scaled data. We also cannot use standard nonmetric scaling algorithms, such as LINMAP (Srinivasan and Shocker 1973), MONANOVA (Kruskal 1965), and the polyhedral method (Toubia et al. 2004), because these require an interval-scaled utility function, albeit at the unobserved (latent) level. For this reason, the measures of fit minimized by these methods (stress, mean absolute error, minimum error) require that the parameters be unique up to, and only up to, multiplication by a positive constant. In contrast, (3) comprises a larger class of transformations. This larger class includes as a

proper subset multiplicative transformations allowed for interval-scaled utility functions. Each of the additional, infinitely many nonmultiplicative transformations (e.g., transformations satisfying the constraints in (5) and (6)) represent alternative parameterizations in a linear utility function. In other words, one cannot obtain a unique set of parameters from a linear model assuming interval-scaled preferences when the actual preferences have a lexicographic structure. The addition of judgment and response error complicates the discussion. In §6, we report the results of a simulation that examines the performance of one nonmetric scaling method, LINMAP, when it is used with lexicographic preference data containing error. In most of these problems, the linear model obtains zero part worths for all but one (and never more than two) attributes. The inclusion of error therefore does not appear to resolve the degeneracy that occurs because of the violation of the assumption that consumers have cardinal utility functions.

3. Nonstandard Lexicographic Preferences

The above discussion assumes that a person evaluates the alternatives one attribute at a time. Here, we examine two variants of a lexicographic model, which we call satisficing and binary lexicographic models. The two variants can be combined to form a binary-satisficing lexicographic model.

A *satisficing lexicographic model* allows indifference among attribute levels. It can occur if a person finds that any higher value of an ordered attribute, such as the amount of computer memory, is no more valuable beyond an upper threshold (Simon 1956). Analogously, there can be a minimum threshold below which an attribute has no value to a consumer; computer memory is again an example. For nominal attributes, indifference among attribute levels can reflect absence of preference over a subset of attribute levels—for example, among different colors or physical styles of a product. An attribute with n levels can have anywhere between 1 to n possible indifference classes, the former corresponding to the case where there are no preferences over the attribute levels, and the latter to the case in which each level is its own indifference class. The case with n indifference classes is, of course, just the standard lexicographic model, and is in this sense a special case of a satisficing model with s indifference classes, where s is an integer ranging from 1 to n .

A *binary lexicographic model* relaxes the assumption that a consumer evaluates alternatives one attribute at a time. A person using the rule has an importance ordering over the attribute levels, rather than over the attributes. This decoupling of the levels from the

Figure 2 Example of a Consumer Switching Across Attribute Levels
 D-LP-MP-HP-C-U

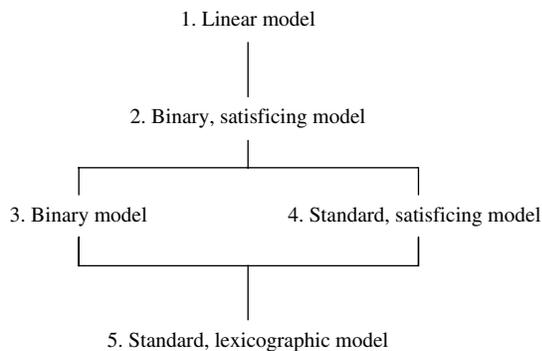
Rating	Brand	Price	Brand
90	D	LP	
85	D	MP	
80	D	HP	
75	D	HP	
<hr/>			
70		LP	C
65		LP	C
60		LP	U
50		LP	U
<hr/>			
45		MP	C
40		MP	U
35		MP	U
30		MP	U
<hr/>			
25		HP	C
20		HP	C
10		HP	U
5		HP	U

D = Dell LP = Low price
 MP = Med price C = Compaq
 HP = High price U = Unbranded

(pooled) levels, possibly switching across attributes in the process. This is the most general lexicographic model we examine. Figure 3 shows the hierarchical relationship among the four lexicographic models and a linear model.

Representations. The binary lexicographic model is obtained by treating each attribute level as a distinct (binary) variable in a utility function. Let $N = \sum_{k=1}^m n_k$ denote the total number of levels across the m attributes. Let $l = 1, 2, \dots, N$ denote an ordering of the N attribute levels, in increasing preference order. We associate the 0-1 integer x_l with level l , and set $x_l = 1$ if level l appears in an alternative; otherwise,

Figure 3 Relationship Between Linear Model and Four Lexicographic Models



$x_l = 0$. In this way, we can describe each multiattribute alternative by a vector (x_1, \dots, x_N) in which $x_l = 1$ for (some) m values of l ; and $x_l = 0$ for all other $N - m$ values of l . We denote the associated utility for an alternative as $u(x_1, \dots, x_N)$. Then

$$u(x_1, \dots, x_N) = \sum_{l=1}^N \beta_l x_l \tag{9}$$

represents a lexicographic utility function if

$$\beta_l > \sum_{t=l+1}^N \beta_t. \tag{10}$$

For example, we can use the values

$$\beta_l = \frac{1}{2^l}, \quad \text{for all } 1 \leq l \leq N, \tag{11}$$

because

$$\sum_{t=l+1}^N \frac{1}{2^t} < \sum_{t=l+1}^{\infty} \frac{1}{2^t} = \frac{1}{2^l} = \beta_l, \quad \text{for all } 1 \leq l \leq N-1. \tag{12}$$

The binary lexicographic model subsumes the standard lexicographic model in the following sense. We can associate the first n_1 terms of (9) with the levels of attribute 1, arranged in descending preference order; the following n_2 terms with the levels of attribute 2, also arranged in descending preference order, etc. Then the binary lexicographic model produces the same rankings for a set of alternatives as does the corresponding standard lexicographic model in which the alternatives are evaluated one attribute at a time.

The satisficing condition is represented in both the standard and binary lexicographic models by assigning the same integer x_k to two or more tied levels of an attribute. Which common integer value is assigned to the tied levels in a nonbinary model does not matter. The only requirement is that we assign the numbers $0, \dots, n_k - 1$ to the levels in nondecreasing preference order.

4. Inference

We describe a procedure for inferring lexicographic rules from data on pairwise comparison of alternatives, obtained directly, or inferred from rankings or ratings of a set of alternatives.

Let Ω denote the set of pairwise comparisons. As in the above discussion, we assume that each alternative $\mathbf{x} \equiv (x_1, x_2, x_3, \dots, x_m)$ is defined over m attributes, and that attribute k has n_k levels. We write $(\mathbf{x}, \mathbf{x}') \in \Omega$ to denote that \mathbf{x} is preferred to \mathbf{x}' . We assess the goodness of fit of a candidate lexicographic solution by a concordance measure that counts the number of pairs of alternatives for which the actual and estimated preference orderings are the same. Maximizing

this measure is equivalent to maximizing the value of Kendall’s tau and Goodman and Kruskal’s gamma between the actual and estimated rankings for a set of alternatives.³ The problem of finding a lexicographic solution that maximizes this concordance measure is a combinatorial optimization problem, which cannot be efficiently solved by enumeration unless the number of possible orderings is small. Schmitt and Martignon (1999) show that the inference problem is NP-Hard when the preference data have error. Efficient solution procedures are therefore unlikely for problems in this class, unless $P = NP$ (see, e.g., Garey and Johnson 1979). We therefore use the following polynomial-time approximation algorithm to find a solution to the problem of inferring a lexicographic preference structure from pairwise comparisons data.

We begin by describing the proposed algorithm for a standard lexicographic rule over ordered attributes. After that, we describe the changes to the algorithm that allow nominal attributes, and that permit the inference of binary, satisficing, and binary-satisficing lexicographic rules.

Initialization step. Let

$$S_0 = \{1, \dots, m\} \tag{13}$$

denote the set of attributes. Arrange the levels of each attribute $k \in S_0$ in increasing preference order, and assign to them the sequence of integers $0, \dots, n_k - 1$. Each level of attribute $k \in S_0$ is thus identified by a unique value $x_k \in \{0, \dots, n_k - 1\}$. Let $\mathbf{x} = (x_1, \dots, x_m)$ denote a product profile in which attribute $k \in S_0$ has the level associated with x_k . For each alternative \mathbf{x} , compute⁴

$$u_{1k}(\mathbf{x}) = \frac{x_k}{n_k}, \quad \text{for all } k \in S_0. \tag{14}$$

Let Ω denote the set of all pairwise comparisons $(\mathbf{x}, \mathbf{x}')$ in which \mathbf{x} is preferred to \mathbf{x}' . For each $(\mathbf{x}, \mathbf{x}') \in \Omega$, for all $k \in S_0$, compute

$$d_{1k}(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if } u_{1k}(\mathbf{x}) < u_{1k}(\mathbf{x}'), \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Let

$$Z_{1k} = \sum_{(\mathbf{x}, \mathbf{x}') \in \Omega} d_{1k}(\mathbf{x}, \mathbf{x}'), \quad \text{for all } k \in S_0. \tag{16}$$

³ Kendall’s tau = $1 - (2Z/NC)$, where Z is the number of reversals and NC is the total number of pairwise comparisons.

⁴ As discussed in §2, the β parameters do not have metric properties and are unique only up to the class of transformations in (3). In this algorithm, we use the utility function representation in (7) for simplicity. In general, however, we can use any other representation that satisfies the conditions in (3).

Let k_1 denote an attribute for which Z_{1k} has the smallest value across all $k \in \{1, \dots, m\}$. That is,

$$Z_{1k_1} = \min\{Z_{1k} \mid k \in S_0\}. \tag{17}$$

Select attribute k_1 as the first lexicographic attribute, arbitrarily breaking ties if necessary. Set

$$u_1(\mathbf{x}) = \frac{x_{k_1}}{n_{k_1}}, \tag{18}$$

and

$$S_1 = S_0 \setminus \{k_1\}. \tag{19}$$

Thus, S_1 is the set of attributes remaining after k_1 is eliminated from S_0 .

Recursion step. For each alternative \mathbf{x} , compute

$$u_{tk}(\mathbf{x}) = u_{t-1}(\mathbf{x}) + \frac{x_k}{n_{k_1} n_{k_2} \dots n_{k_{t-1}} n_k}, \quad \text{for all } k \in S_{t-1}, \tag{20}$$

where

$$S_{t-1} = \{1, \dots, m\} \setminus \{k_1, \dots, k_{t-1}\} \tag{21}$$

is the set of attributes still not selected after step $t - 1$, and

$$u_{t-1}(\mathbf{x}) = \frac{x_{k_1}}{n_{k_1}} + \dots + \frac{x_{k_{t-1}}}{n_{k_1} \dots n_{k_{t-1}}}. \tag{22}$$

For each $(\mathbf{x}, \mathbf{x}') \in \Omega$, and for all $k \in S_{t-1}$, let

$$d_{tk}(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if } u_{tk}(\mathbf{x}) < u_{tk}(\mathbf{x}'), \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

Compute

$$Z_{tk} = \sum_{(\mathbf{x}, \mathbf{x}') \in \Omega} d_{tk}(\mathbf{x}, \mathbf{x}'), \quad \text{for all } k \in S_{t-1}. \tag{24}$$

Let k_t denote an attribute for which Z_{tk} has the smallest value across all $k \in S_{t-1}$. That is,

$$Z_{tk_t} = \min\{Z_{tk} \mid k \in S_{t-1}\}. \tag{25}$$

Select attribute k_t as the t th lexicographic attribute, arbitrarily breaking ties if necessary. Set

$$u_t(\mathbf{x}) = \frac{x_{k_1}}{n_{k_1}} + \dots + \frac{x_{k_t}}{n_{k_1} \dots n_{k_t}}, \tag{26}$$

and

$$S_t = S_{t-1} \setminus \{k_t\}. \tag{27}$$

Termination step. Stop when there are no further alternatives left (in which case the attributes not considered are irrelevant to the problem), or if $t = m$ (in which case all the attributes have been considered).

We call the above procedure a greedy algorithm. The following theorem characterizes the solution obtained by the procedure. It is valid for all

lexicographic rules we have considered in this paper. A proof of the theorem for a standard lexicographic rule appears in the appendix. With slight modification of terminology, the same proof can be used to prove analogous results for the binary, satisficing, and binary-satisficing lexicographic rules.

THEOREM 2. *Suppose there exists an attribute ordering that perfectly reproduces an input rank ordering of product profiles; then the greedy algorithm finds such an ordering.*

The algorithm is not guaranteed to find an optimal solution if the preference data contain error. In §6, we evaluate how well it recovers a lexicographic preference model when the data contain error.

Incorporating Nominal Attributes. Consider the recursion step of the algorithm. If attribute k is nominal, we compute (20)–(24) for each possible preference ordering of the attribute levels, and select that ordering for which the value of (25) is the smallest. This enumeration over all possible permutations of attribute levels is feasible because n_k is typically small for conjoint analysis. For example, the application described in the next section has no more than three levels for any attribute. In this case, enumerating all $3! = 6$ possible attribute-level orderings is simple. In the unlikely instance when n_k is large, we can use a greedy algorithm, similar to the one described above, for selecting the optimal ordering of an attribute level.

Inferring Nonstandard Lexicographic Models. The greedy algorithm can be modified in a straightforward manner to accommodate inferences about the structures of the nonstandard lexicographic models discussed earlier. The binary lexicographic model simply requires the use of the above greedy algorithm by treating each of the $N = \sum_{k=1}^m n_k$ attribute levels as a binary attribute (indicating the presence or absence of the attribute level). The satisficing versions of the standard and binary lexicographic models additionally compute Z_{ik} for all possible ways of pooling of attribute levels, provided we maintain at least two pooled levels for each attribute. This requires less computational effort for ordered attributes than it does for nominal attributes, because one can then restrict the pooled levels to a sequence of successively preferred attribute levels. For nominal attributes, one can enumerate all possible groupings if the number of levels, n_k , is small. Otherwise, one needs an algorithm for grouping the levels. We again use a greedy heuristic: First, select n_k groupings, each with a single attribute level. Next, eliminate one grouping, placing two levels in one common subset; to do so, evaluate all possible $n_k(n_k - 1)$ pairs of attribute-level groupings. This successive, greedy merging of levels continues until only two groupings remain. The partition for which the number of reversals is minimized is then selected.

5. Empirical Comparisons of Models

We examine the following four questions in this section: (1) Do preference data support the use of one or another lexicographic model by consumers? (2) If so, is the standard lexicographic model adequate, or does one or another of the proposed variants provide a better description of the decision process? (3) Is there a single lexicographic model that appears to provide the best description of the data? (4) Can we infer a hierarchical market structure by combining information about the lexicographic rules inferred for different consumers? We address these questions using data from a conjoint study for laptop computers. We use the data to estimate each of the individual-level lexicographic models described above. For comparison, we estimate part-worth utility functions for each consumer using LINMAP (Srinivasan and Shocker 1973). We also examine the adequacy of inferring lexicographic structures based on importance weights derived using a traditional conjoint approach.

Recall that there are always parameterizations of a linear model that can represent any given lexicographic model. For this reason, one cannot expect a lexicographic model to provide better fit to data than does a linear model: All one has to do is choose a linear model that corresponds to a parameterization of an estimated lexicographic rule. In practice, it is possible that a linear model might not find a parameterization that exactly reproduces a lexicographic function. There can be two reasons for this. First, the true model might be linear and not lexicographic. Second, even if it is lexicographic, the presence of error in the data and the assumption that there is an underlying metric utility function might lead to parameter estimates in a linear model that do not reflect a lexicographic structure. How well an estimated linear model approximates the lexicographic model is an empirical question that we examine with actual preference data in this section, and with simulated data in the next section.

We use the partial hierarchy in Figure 3 to assign the linear model, or one of the lexicographic models, to each individual. We assess the adequacy of this assignment for both the estimation data and for hold-out data, for which we obtain preference predictions using the estimates for the assigned model. We then illustrate how the lexicographic analysis can be used for a hierarchical clustering of consumers (segmentation) and for inferring market structure. We briefly discuss the estimation of a linear model, and the rule inference procedure, before describing the application.

Linear Model. We use LINMAP (Srinivasan and Shocker 1973) to estimate the linear model because, like the methods in §4, it uses ordinal (pairwise

comparisons) data. The difference is that LINMAP minimizes the sum of absolute errors in utilities, whereas the proposed greedy algorithm minimizes the number of reversals. This could lead to a lower fit for the linear model and may adversely affect the accuracy of our model assignment procedure, which we discuss below.⁵ Simulation evidence, however, suggests that (i) the fit of the linear model is never worse than that of the lexicographic models when the data are generated using the linear model, and (ii) our procedure does quite well in discriminating among the models. See §6 and Table 3.

Rule Assignment. We use the nested structure in Figure 3 for rule assignment. For ease of exposition, we refer to the various models by the numbers 1–5 assigned to them in this figure. We allow for the possibility that people differ in which model they use. We assign a model to a subject in the following way. We start by comparing Model 1 with Model 2. If Model 1 has a (statistically) better fit, we assign it to the subject. Otherwise, we compare Model 2 *sequentially* with Models 5, 4, and 3. If all three comparisons yield statistically significant differences in fits, we assign Model 2 to the subject. Otherwise, we assign to the subject the *last* (i.e., the most constrained) model tested against Model 2 for which there is no statistically significant difference in fit. We assess statistical significance using a standard *t*-test for the equality of proportions with Bonferroni-adjusted significance levels. Note that because Models 3 and 4 are not nested, it is possible to assign models in the sequence 5, 3, 4. This can result in the reassignment of some subjects to Model 3. For these subjects, both the satisficing and binary lexicographic models do as well as the binary-satisficing model.

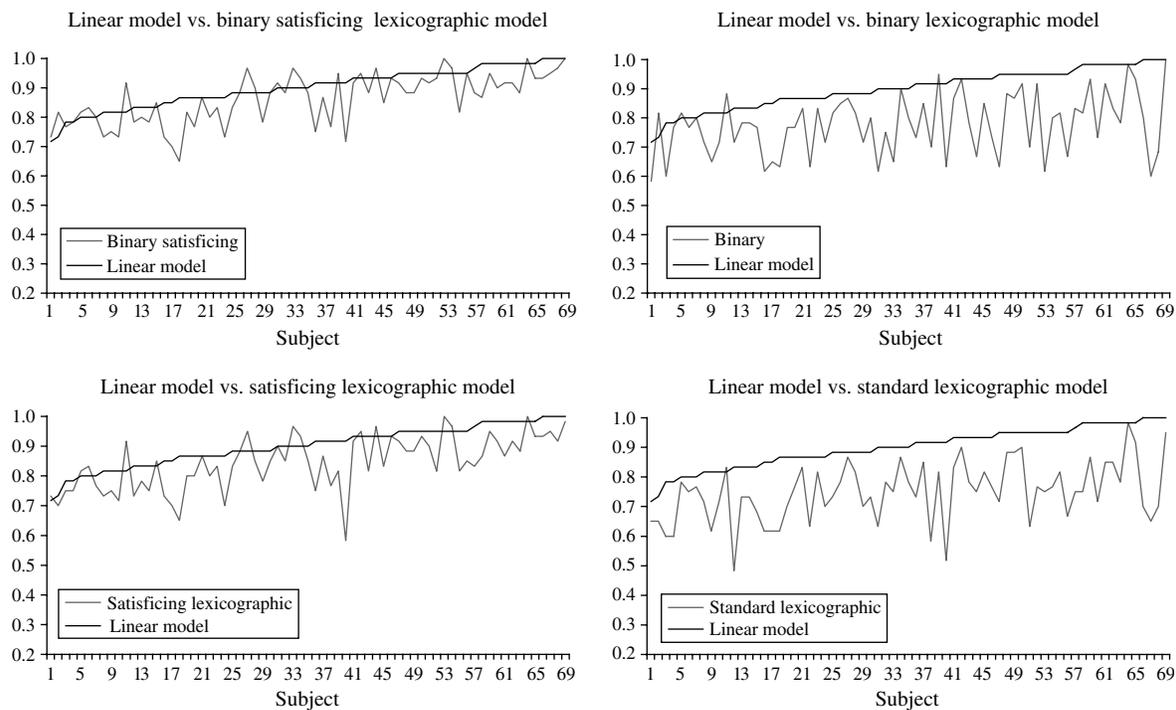
The above procedure gives unbiased assignments. To see why, consider again the relationship among the models shown in Figure 3. If there were no error in the data, and if the true rule were lexicographic, we would obtain exactly the same (perfect) fits with the general model and with a nested model (e.g., with a linear model and with the true lexicographic model). On the other hand, if the true model were linear and compensatory (i.e., a linear model with parameters not representing a lexicographic model), then the linear model would fit better than any lexicographic model. Thus, if there is no error in data, one should assign (i) a lexicographic model if there are no differences in the number of correct pairwise comparisons; and (ii) a linear model if it gives a higher number of correct pairwise comparisons. This is what the proposed assignment procedure does. The addition of error to data modifies the above argument in

the following way. Suppose we reverse a certain proportion p of pairwise comparisons. Suppose the true model is linear compensatory. Consider, for example, a problem with three attributes where the first (third) attribute is the most (least) important. Then, (i) a linear model, L , estimated using the data, has an expected error rate of $E[\hat{p}_{L|L}] = p$; and (ii) a lexicographic model, Lex (equivalently, a constrained linear model with $\beta_1 > \beta_2 + \beta_3$ and $\beta_2 > \beta_3$), has a higher expected error rate $E[\hat{p}_{Lex|L}] > p$, because it cannot represent the trade-off in the linear model. Now suppose the true model is lexicographic. Then a lexicographic model will have an expected error rate of $E[\hat{p}_{Lex|Lex}] = p$; and a linear model, estimated using the data, will also have an error rate of $E[\hat{p}_{L|Lex}] = p$, because it is equivalent to a lexicographic model when $\beta_1 > \beta_2 + \beta_3$ and $\beta_2 > \beta_3$. It follows that the following rule is unbiased: (1) Assign the linear model when it has a lower error rate than the lexicographic model; and (2) assign the lexicographic model when the two models have the same error rate. The remaining issue concerns the use of the rule using the limited amount of data one obtains from a respondent. In this case, we use the above test, but make an assessment of whether or not the linear model provides a better fit based on a test of statistical difference in fit. The variance of the difference in the two estimated proportions (of correct predictions obtained by a linear and lexicographic model) decreases with sample size, going to zero in the limit. We note that the sequential assignment of rules that we use is similar to a procedure by Hauser and Shugan (1980) for classifying respondents to different groups depending on how much information they are able to provide when answering conjoint questions. Although this is standard statistical practice (see, e.g., Broder 2000), we further assess, in §6, the performance of this procedure for assigning alternative preference models to consumers.

Conjoint Design and Data Collection Procedure. Each laptop computer was described using five attributes. The attributes and their levels are as follows: (1) price: \$1,999, \$2,599, or \$2,895; (2) memory: 32 MB, 64 MB, or 96 MB; (3) brand: Dell, Compaq, or “no brand;” (4) processor: Intel Pentium at 266 MHz or 300 MHz; and (5) hard drive: 3 GB or 4 GB. Sixteen product profiles were obtained as the treatments of an orthogonal fractional-factorial plan in which the attributes were used as the design factors. Each product profile was described on a separate card. The 16 cards describing the product profiles were shown in random order to each of 69 MBA students at an East Coast university in Spring 2001. All respondents rated each profile for liking on a 1–100 scale, assigning a higher rating to a preferred alternative. We use the ratings data to estimate individual-level part-worths estimates from LINMAP (Srinivasan

⁵ We thank an anonymous reviewer for bringing this point to our attention.

Figure 4 Comparison of Kendall's tau Value for Linear Model and Four Lexicographic Models



and Shocker 1973).⁶ We compare these estimates to the various lexicographic models below. We also discuss the adequacy of inferring lexicographic structures based on importance weights derived using a traditional conjoint approach.

Results. The overall results are in agreement with the partial, hierarchical description of the models in Figure 3. Figure 4 shows the Kendall's tau values for each lexicographic rule, sorted in ascending order across the 69 subjects. For comparison, we plot the Kendall's tau value for the linear model in each of these graphs. For the two nonnested models in Figure 3, the plots in Figure 4 suggest the superiority of the satisficing lexicographic model to the binary lexicographic model. There are some cases in which the Kendall's tau value is higher for a binary-satisficing lexicographic model than for the linear model. This is not possible if the two models are estimated using the same objective function: As we discussed in §4, each lexicographic model is a constrained version of a linear model. The explanation for the discrepancy

is that LINMAP maximizes a goodness-of-fit measure that is not directly related to Kendall's tau, whereas our estimation procedure, as noted in §4, is equivalent to maximizing Kendall's tau.

Referring to Figure 4, we observe that the fit of the standard lexicographic model is the poorest when compared to the fit of the linear model. On average, the value of Kendall's tau is 0.75 for the standard lexicographic model and 0.9 for the linear model. If we did not consider the other lexicographic models proposed in this paper, we would reject the hypothesis of no differences in fit between the two models. The binary lexicographic model has an average Kendall's tau value of 0.78 and therefore does slightly better than the standard lexicographic model. The satisficing lexicographic model performs substantially better, obtaining fits very similar to those achieved using the linear model (average Kendall's tau = 0.84). However, there is little additional improvement in fit from using a binary-satisficing model (average Kendall's tau = 0.86). Overall, if one were to select a single lexicographic model or a linear compensatory model to describe the data across subjects, the satisficing model would appear to be the most appropriate: It is the most constrained of the tested models, and it does about as well as a linear model.

A series of statistical comparisons support this conclusion. Across subjects, we find that the difference in the proportion of discordant pairs between the standard lexicographic and binary lexicographic models is insignificant ($p > 0.32$). The fits of both

⁶ We also use OLS regression to estimate the individual-level part worths. Full details of the OLS analysis are available from the authors. Here, we focus on LINMAP estimates for two reasons. First, like our proposed methods, LINMAP only considers the ordinal properties of the data for estimation. Second, LINMAP estimates typically give higher values for Kendall's tau, which is the measure we use to evaluate the fits for lexicographic models. This allows for a more conservative evaluation of the proposed lexicographic models than does a comparison with the results of the OLS regression.

Table 1 Average Kendall's Tau Values for Models Assigned to Subjects

Model	Number of subjects	Average tau	Average deviation of tau from linear model
Linear	23	0.94	0.00
Lexicographic			
Binary satisficing	1	0.95	0.00
Binary	3	0.89	−0.03
Satisficing	28	0.88	0.02
Standard	14	0.80	0.06

models, however, are significantly worse than those obtained from the satisficing lexicographic model and the binary-satisficing model ($p < 0.01$). In addition, a t -test comparing the proportion of discordant pairs between the linear model and the binary-satisficing model fails to reject the null hypothesis of no difference in fit between these two models ($p > 0.1$). Finally, the proportion of discordant pairs are not statistically different for the binary-satisficing and the satisficing lexicographic models ($p > 0.35$).

We now turn to rule assignment by allowing for the possibility that subjects differ in terms of which model they use. Table 1 shows the distribution of subjects across models when Model 4 is evaluated before Model 3. As expected, the assigned lexicographic models do not differ substantially in their fits from the more general linear model.⁷ The small negative value for the average difference in values of Kendall's tau for the binary lexicographic model does not mean that the empirical analysis does not support the theoretical claim that the linear model is more general than all lexicographic models discussed. As discussed in §2, there is always a linear model implied by a lexicographic rule, and for this parameterization, the linear model gives the same Kendall's tau value as the lexicographic model.

Table 2 reports the pooling of the attribute levels by the 28 subjects who are assigned to the satisficing lexicographic model. Forty-six percent of the subjects are indifferent between the \$1,999 and \$2,599 price levels;

⁷ As all four lexicographic models are special cases of the linear model, a lack of statistical difference in the fits of a linear model and a lexicographic model implies that the latter does as well as the former (it obviously cannot do any better). We interpret this as implying that the lexicographic model is adequate for representing preferences. Thus, when selected, a lexicographic model's fit should always be close (i.e., not significantly different) to that of a linear model. Hence, the near equality of fit between the selected lexicographic model and the linear model in Table 1 is to be expected. Note that for the 23 subjects to whom the linear model is assigned, there is indeed a significant difference in fit. The average Kendall's tau for the (selected) linear model is 0.94, whereas that of the binary-satisficing model (the most general lexicographic model) is only 0.83.

Table 2 Pooling of Attribute Levels for the 28 Subjects Assigned to the Satisficing Lexicographic Model

Attribute	Levels pooled	Frequency	Percent
Price	None	6	21.43
	\$1,899/\$2,599	13	46.43
	\$2,599/\$2,895	9	32.14
RAM	None	3	10.71
	32 MB/64 MB	8	28.57
	64 MB/96 MB	17	60.71
Brand	None	2	7.14
	Unbranded/Compaq	2	7.14
	Unbranded/Dell	2	7.14
	Compaq/Dell	22	78.57

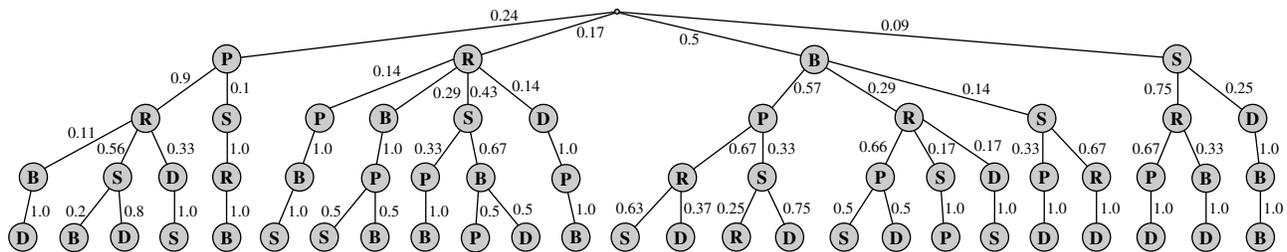
61% consider the 64 MB and 96 MB levels of RAM as being equivalent; and 79% assigned to the satisficing lexicographic model are indifferent between the two brands Dell and Compaq, treating them as a single branded category. For the 14 subjects assigned to the standard lexicographic model, where attribute pooling is not permitted, 86% prefer Dell.

As noted earlier, Models 3 and 4 are not nested, making it is possible to assign models in the sequence 5, 3, 4. This results in the reassignment of 15 subjects to Model 3. Thus, in all, there are 14 subjects for whom we cannot say whether Model 3 or Model 4 is the more appropriate. For these subjects, both the satisficing and binary lexicographic models do as well as the binary-satisficing model.

Adequacy of Inferring Lexicographic Structure Using Conjoint-Based Importance Weights. We examine the possibility that the parameter estimates from a part-worths model, estimated for each subject using the ratings data, might contain sufficient information to infer the ordering of attributes or attribute levels in a lexicographic model.⁸ For example, one might compute the relative importance of attributes, and then test a standard, or satisficing, lexicographic model in which the attributes are ordered by the magnitudes of these relative importance. How well will this method of inferring a lexicographic rule perform, compared to the greedy algorithm described in §4? The answer is: not too well. For example, consider the 28 subjects to whom we assign a satisficing lexicographic model. If we use a lexicographic ordering implied by the attribute importance in a regression (as measured by their relative part-worth ranges), the average (across subjects) Kendall's tau correlation between the actual and estimated preference orderings is 0.65; the comparable value of Kendall's tau correlation is 0.88 for the attribute ordering obtained

⁸ Following the suggestion of a reviewer, we use regression-based analysis of the ratings in the present discussion; in all other places in the paper, we use LINMAP to estimate the linear model.

Figure 5 Summary of Lexicographic-Preference Structure for 42 Subjects



Notes. Legend: P—Price; R—RAM; B—Brand name; S—Speed of processor; D—Hard drive capacity. The numbers on the edges are the proportion of subjects associated with the node into which an edge terminates. For example, the proportion of subjects who consider price to be the most important lexicographic criterion is 0.24. Among these subjects, the proportion who consider RAM to be the next most important attribute is 0.9. The proportion of subjects for whom price is the most important attribute and RAM is the second-most important attribute is $(0.24)(0.9) = 0.216$. Similar calculations give the proportion of all subjects associated with any node of the tree.

using the greedy algorithm (see Table 1). Across subjects, the range of Kendall’s tau values is 0.35 to 0.83 for the regression-based method; it is 0.73 to 1.00 for the proposed algorithm.

Overall, the regression-based method never provides a better fit than the proposed algorithm; and it does substantially worse than the proposed algorithm both on average and in the worst case. We have also checked the magnitude of correlation between the attribute ordering obtained using the regression and the greedy algorithm. Across subjects, the average Kendall’s tau correlation is 0.65, and it ranges from -0.40 to 1.00. We will further examine this issue in the simulation described in §6.

Constructing Hierarchical Clusters and Inferring Market Structure. We examine the decision rules for the 42 subjects to whom we assign the standard or satisficing lexicographic models. Figure 5 summarizes their decision rules. The sequence of edges from the root to a terminal node corresponds to the ordering of attributes for one or more of these subjects. For example, the leftmost sequence of edges corresponds to the decision process for the (one) individual in the sample for whom price is the most important attribute, followed by memory, brand name, hard disk, and speed. Multiplying the proportions along a particular path in the tree gives the proportion of subjects using the associated attribute ordering. For example, the attribute sequence: brand–price–memory–speed–hard disk is used by five subjects, comprising a $0.5 \times 0.57 \times 0.67 \times 0.63 = 12\%$ of the sample.

We can interpret Figure 5 as describing a hierarchical clustering of the subjects. Each terminal node of the tree is associated with one or more consumers. Proceeding upwards in the tree, we obtain a clustering of consumers in terms of the similarity of their decision rules. For example, the sequence of nodes labeled P-R-S groups consumers who all use price, memory, and speed as the three most important attributes in that order, but who differ in their ordering of the last attribute (B or D) shown in the figure.

One step higher, the sequence of nodes P-R groups consumers who all use price as the most important attribute and memory as the secondmost important attribute, but who differ in their ordering of the last two attributes (B, S or D). At the highest level, we obtain four segments that differ in terms of the most important attribute: brand (50% of the consumers), price (24%), memory (17%), and speed (9%). Similarly, we obtain 11 segments if we focus on the two most important attributes. For example, the brand-price (price-memory) segment represents $0.5 \times 0.57 = 28\%$ ($0.24 \times 0.9 = 22\%$) of the consumers. Using all the attributes for classifying consumers results in 25 segments. As in all hierarchical clustering methods, there are no definite rules for how many segments to retain, although there are rules of thumb that we can use. For example, we can compute an average value of Kendall’s tau across consumers at different levels of the tree (i.e., for varying numbers of the most important attributes used). The marginal improvement in the average Kendall’s tau value can then be used to guide the selection of the level at which to truncate the tree. In addition, there are managerial issues we need to consider. In a one-to-one marketing context, we are likely to retain most of the segments. However, if the purpose of the analysis is strategic guidance (e.g., positioning), then fewer levels are likely to suffice.

Figure 5 can also be interpreted as describing an aggregate, hierarchical market structure for the associated 42 subjects. Fifty percent of these subjects use a brand-first evaluation process; 24% use a price-first evaluation process; and the rest use a feature-first evaluation process. Put another way, Dell and Compaq have a corner on half the market, and the lowest-priced sellers are most favorably evaluated by nearly a quarter of the market. Among those consumers who first look for a branded computer, 57% use price as the next attribute. Among those who use price as the first screening criterion, 90% evaluate the screened alternatives based on memory (RAM). Brand name, if at all

relevant, is a lower-level consideration in their choice. One can make other similar deductions by further interpreting this figure as a hierarchical market structure. Note that the ordering of attributes is contingent on the range of values for the attribute levels. Thus, for example, if we were to include absurdly high prices, a consumer might use a binary lexicographic model in such a way that alternatives with the highest prices are all ranked below alternatives priced within an acceptable range. The consumer might then switch to another attribute, and return to price, in one or more later stages of a binary lexicographic rule. For nonprice attributes, the satisficing model also allows consumers a way to ignore excessively high (or very low) values of attribute levels.

Predictive Validity. We use the following cross-validation method to assess the predictive accuracy of the lexicographic models assigned to each of the subjects (Mosier 1951). First, we remove each product profile, one at a time, for each subject. Then we reestimate the (assigned) model for a subject using the remaining 15 profiles. We use the parameter estimate so obtained to make pairwise comparisons between the holdout profile and each profile used in the estimation. As each of the 16 product profiles is held out, one at a time, this gives $16 \times 15 = 240$ pairwise comparisons per subject. We use these 240 comparisons to compute a (predicted) value of Kendall's tau for each subject. For comparison, we compute the corresponding value of Kendall's tau for a linear model.

Figure 6 shows these Kendall's tau values for each of the 46 subjects who are assigned to a lexicographic model. Across these subjects, the mean value of Kendall's tau is 0.77 for the linear model and 0.78 for the lexicographic model. This suggests that there are no significant differences in holdout predictions obtained from a (more general) linear model and a lexicographic model assigned to a subject. The lack of significant difference in holdout predictions provides further support for the validity of our procedure for

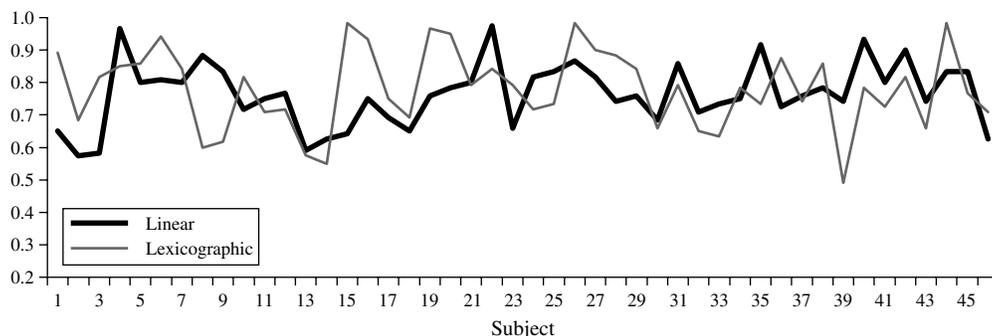
assigning models to subjects. Such a validity would be questionable if the linear model had better predictive validity.

In summary, the above analysis suggests the following answers to the questions noted at the beginning of this section. (1) The data are consistent with the use of a lexicographic rule by two-thirds of the subjects in this study. The lexicographic models assigned to each of these subjects perform as well as a part-worths model in terms of model fit and holdout prediction of preferences. For the remaining subjects, a part-worths model provides a better fit than does any lexicographic model. (2) The standard lexicographic model is satisfactory for representing the preferences for only 14 subjects. However, there are 32 subjects for whom an alternative lexicographic model is more appropriate. Thus, if we were only testing for the standard lexicographic model, we would find limited support for the use of lexicographic rules by the subjects in this study. In this sense, the proposed variants of lexicographic models are useful. (3) If we were to select a single lexicographic model for the 46 consumers for whom such a model is adequate, then the satisficing lexicographic model would appear to be the most appropriate in terms of both model fit and holdout prediction. Managerially, the lexicographic analysis can be useful for the hierarchical segmentation of consumers and for inferring market structure.

6. Simulated Testing of Lexicographic Models

We performed a Monte Carlo simulation experiment to assess: (i) the computational time needed to implement the greedy algorithm; (ii) the accuracy of the greedy algorithm in recovering a known lexicographic preference model when there are errors in the pairwise-comparisons data; (iii) the adequacy of the proposed method for assigning a linear model, or one of the lexicographic models, to subjects; and

Figure 6 Predictive Accuracy of Assigned Lexicographic Model and Linear Model



Notes. In this figure, the 46 subjects are sorted based on which lexicographic model they were assigned: Subject 1 (binary-satisficing model); Subjects 2, 3, 4 (binary lexicographic model); Subjects 5–32 (satisficing lexicographic model); and Subjects 33–46 (standard lexicographic model).

(iv) the suitability of inferring lexicographic orders using conjoint-based importance weights. We first describe the design of the simulation study, and then the results.

Simulation Design. We estimate lexicographic and linear models using a $(5 \times 2 \times 2 \times 2)$ factorial design with the following treatments: true preference model (standard, binary, satisficing, binary-satisficing, and linear compensatory); amount of error in pairwise-comparisons data (5% and 20% error); number of attributes ($m = 5$ and $m = 8$); and number of product profiles (24 and 32). In problems with five attributes, we generate the product profiles using a fractional factorial plan in which two attributes had four levels each, two attributes had three levels, and one attribute had two levels. Fractional factorial plans for problems with eight attributes use three additional two-level attributes. We perform 20 replications per treatment. Each replication uses a different set of values (attribute orderings or part worths). The data set for each replication is used to estimate the linear model and all four lexicographic structures.

Data Generation. We randomly select a lexicographic attribute ordering and use it to compute a true (ordinal) preference score for each product profile in the conjoint design.⁹ We then convert these rank orders into paired-comparison data. In each pair, the higher-ranked product profile is specified as the preferred alternative. There are 276 pairs in the 24-profile design ($24 \times 23/2$) and 496 pairs in the 32-profile design. We add error by randomly flipping this true preference order in 5% or 20% of the pairs, and use these data for estimating the parameters of the various models.

The procedure for generating data for the linear model is similar to the one used for the lexicographic model, but is different in terms of how the preference scores are computed. We randomly generate (from a uniform distribution) a part-worth value for each of the dummy indicators of the m product attributes. There are 11 (14) such indicators for the five-attribute (eight-attribute) conjoint design. We use these part-worth values as weights to compute a (true) preference score for each of the product profiles. We then follow the same procedure above to convert these preference scores into paired comparison data and to add error.

⁹To generate this attribute ordering, we draw m random numbers from a uniform distribution, and assign one number to each attribute. The attributes are assumed to be ordered in importance by the assigned numbers. For the binary lexicographic model, we convert the 5 (8) original attributes into 11 (14) dummy attributes. For the binary-satisficing model, we pool levels two and three from the product attributes with three and four levels, which reduces the number of dummy attributes by four. We used this same pooling of attribute levels for the satisficing model.

Measures. The performance criteria of interest are: (i) the recovery of the true attribute orderings for the lexicographic models and the true part-worth parameters for the linear model; and (ii) the robustness of the proposed method for assigning preference models to consumers. We use the Kendall's tau correlation between the true and estimated attribute orderings as a measure of bias. As the part-worth coefficients have metric properties, we use the Pearson correlation coefficient as a measure of recovery of the true parameters for the linear model. We also compute a hit rate (i.e., percent of correct predictions) as a measure of fit. For comparison, we report this measure for all the estimated models. We assess the accuracy of model assignment by the percent of replications in which it correctly assigns a true model to a consumer. We also report the CPU time (in seconds) taken by the greedy algorithm for estimating the lexicographic rules.

Results. A summary of the results appears in Table 3. Computationally, the greedy algorithm estimated each of the standard and binary lexicographic models in less than one second. On average, it estimates each satisficing model in 3.7 seconds, and each binary-satisficing model in 9.8 seconds.

The average (Kendall's tau) correlation coefficient between the true and estimated attribute orderings is 0.91 across all treatment conditions. The average hit rate is 0.86. The average tau correlation (hit rate) across the lexicographic models is 0.90 (0.86) and is similar to that obtained for the linear model 0.93 (0.83). Note that the hit rates corresponding to each of the five true models in Table 3 (highlighted in bold-face) are always greater than or equal to those of the remaining four competing preference models. For example, when the standard lexicographic model represents the true preference structure, all the hit rates from estimating this model are identical to those from the binary-satisficing model (see top panel in Table 3). This is expected, because the former model is a special case of the latter model.

Regardless of the preference structure, lower error is associated with more-accurate recovery of a true attribute ordering, and with a higher statistical fit for the true model. The average tau correlation (hit rate) is 0.97 (0.94) for 5% error, and is 0.85 (0.78) for 20% error. The number of attributes, and the number of profiles in the conjoint design, has substantially less effect on the recovery of the attribute orderings, and virtually no impact on model fit. The average tau correlation increases from 0.88 (0.88) if the conjoint design includes eight attributes (24 profiles) to 0.93 (0.94) when it has five attributes (32 profiles).

The LINMAP goodness of fit is poorer when it is estimated using data from a true lexicographic model than when it is estimated using a linear compensatory

Table 3 Simulation Results: Recovery of Attribute Ordering and Statistical Fit

True model	Error level (%)	Number of attributes	Number of profiles	CPU in seconds	Tau correlation	Fitted model's hit rate				
						Standard	Binary	Satisficing	Binary satisficing	Linear
Standard	5	5	24	<1	1 ^a	0.94^b	0.94	0.94	0.94	0.89
	5	5	32	<1	0.97	0.95	0.95	0.95	0.95	0.76
	5	8	24	<1	0.87	0.95	0.95	0.95	0.95	0.72
	5	8	32	<1	0.97	0.91	0.91	0.91	0.91	0.50
	20	5	24	<1	0.88	0.79	0.79	0.78	0.79	0.75
	20	5	32	<1	0.82	0.79	0.79	0.79	0.79	0.63
	20	8	24	<1	0.63	0.80	0.80	0.80	0.80	0.72
	20	8	32	<1	0.82	0.64	0.64	0.64	0.64	0.49
Binary	5	5	24	<1	0.98	0.80	0.96	0.90	0.96	0.82
	5	5	32	<1	1.00	0.80	0.95	0.89	0.95	0.65
	5	8	24	<1	1.00	0.80	0.96	0.91	0.96	0.83
	5	8	32	<1	1.00	0.79	0.95	0.90	0.95	0.85
	20	5	24	<1	0.83	0.69	0.80	0.76	0.80	0.66
	20	5	32	<1	0.88	0.68	0.79	0.74	0.79	0.60
	20	8	24	<1	0.83	0.67	0.81	0.79	0.81	0.65
	20	8	32	<1	0.95	0.67	0.81	0.78	0.81	0.60
Satisficing	5	5	24	1.05	0.99	0.86	0.85	0.94	0.95	0.72
	5	5	32	2.75	1.00	0.85	0.84	0.90	0.90	0.79
	5	8	24	4.00	0.84	0.87	0.85	0.95	0.95	0.81
	5	8	32	7.00	0.98	0.85	0.83	0.95	0.95	0.79
	20	5	24	1.05	0.91	0.73	0.73	0.78	0.79	0.62
	20	5	32	2.65	0.83	0.74	0.73	0.79	0.79	0.66
	20	8	24	4.00	0.79	0.72	0.75	0.80	0.80	0.71
	20	8	32	7.10	0.81	0.72	0.73	0.80	0.80	0.59
Binary satisficing	5	5	24	3.70	0.96	0.85	0.93	0.92	0.96	0.77
	5	5	32	7.00	1.00	0.75	0.87	0.84	0.94	0.60
	5	8	24	10.10	0.97	0.81	0.91	0.92	0.95	0.79
	5	8	32	18.45	1.00	0.78	0.89	0.90	0.95	0.66
	20	5	24	3.75	0.86	0.74	0.82	0.82	0.84	0.75
	20	5	32	7.00	1.00	0.53	0.65	0.61	0.70	0.50
	20	8	24	10.00	0.60	0.67	0.77	0.78	0.80	0.68
	20	8	32	18.20	0.87	0.67	0.77	0.78	0.80	0.71
Linear	5	5	24		0.95 ^c	0.82	0.85	0.83	0.85	0.89
	5	5	32		0.97	0.80	0.83	0.81	0.83	0.91
	5	8	24		0.95	0.82	0.83	0.85	0.85	0.90
	5	8	32		0.97	0.83	0.83	0.84	0.84	0.92
	20	5	24		0.89	0.71	0.73	0.73	0.74	0.75
	20	5	32		0.95	0.70	0.71	0.69	0.71	0.75
	20	8	24		0.86	0.72	0.71	0.74	0.74	0.74
	20	8	32		0.93	0.71	0.71	0.73	0.73	0.76

^aFor a Lexicographic model, this column reports the Kendall's tau correlations between the true and estimated attribute orderings.

^bTo be read: The percent of pairs of alternatives for which the actual and predicted preference orderings match is 94%.

^cBecause the parameters of a linear model are ratio scaled, we report the Pearson correlations between the true and estimated part worths for LINMAP. The corresponding Kendall's tau correlations are, respectively, 0.82, 0.90, 0.82, 0.87, 0.73, 0.84, 0.69, 0.78.

model. As discussed in §2, the reason for this is that LINMAP assumes an interval-scaled utility function, and this model is not identified with lexicographic preference data. Given pairwise comparisons from an error-free lexicographic model, LINMAP stops after identifying the first (most important) attribute, and sets all other part worths to zero values. There were 160 simulation runs in which we used LINMAP to estimate parameters after adding error to the input data generated using a lexicographic model. In 53%

of these cases, LINMAP again stopped after identifying one attribute; in an additional 17% of cases, it obtained nonzero estimates for a second attribute. No information was obtained in these cases about the ordering of the other attributes. This situation is encountered whether we use 24 or 32 product profiles, five or eight attributes, and low or high error, the latter condition reversing 20% of the true pairwise comparisons produced by a lexicographic model. These results suggest that it may not be appropri-

Table 4 Simulation Results: Model Assignment

True model	Error level (%)	Number of attributes	Number of profiles	Assigned model			
				Standard	Binary	Satisficing	Binary satisficing
Standard	5	5	24	0.95^a	0.00	0.00	0.00
	5	5	32	1.00	0.00	0.00	0.00
	5	8	24	1.00	0.00	0.00	0.00
	5	8	32	1.00	0.00	0.00	0.00
	20	5	24	0.95	0.05	0.00	0.00
	20	5	32	1.00	0.00	0.00	0.00
	20	8	24	1.00	0.00	0.00	0.00
Binary	5	5	24	0.00	1.00	0.00	0.00
	5	5	32	0.00	1.00	0.00	0.00
	5	8	24	0.00	1.00	0.00	0.00
	5	8	32	0.00	1.00	0.00	0.00
	20	5	24	0.00	0.90	0.05	0.00
	20	5	32	0.00	1.00	0.00	0.00
	20	8	24	0.00	1.00	0.00	0.00
Satisficing	5	5	24	0.00	0.00	1.00	0.00
	5	5	32	0.00	0.00	1.00	0.00
	5	8	24	0.00	0.00	1.00	0.00
	5	8	32	0.00	0.00	1.00	0.00
	20	5	24	0.00	0.00	1.00	0.00
	20	5	32	0.00	0.00	1.00	0.00
	20	8	24	0.00	0.00	1.00	0.00
Binary satisficing	5	5	24	0.00	0.00	0.00	1.00
	5	5	32	0.00	0.00	0.00	1.00
	5	8	24	0.00	0.00	0.05	0.95
	5	8	32	0.00	0.00	0.00	1.00
	20	5	24	0.00	0.25	0.35	0.40
	20	5	32	0.00	0.15	0.00	0.85
	20	8	24	0.00	0.30	0.55	0.15
Linear	5	5	24	0.05	0.05	0.05	0.00
	5	5	32	0.00	0.00	0.00	0.00
	5	8	24	0.05	0.00	0.10	0.00
	5	8	32	0.00	0.00	0.00	0.00
	20	5	24	0.30	0.10	0.05	0.00
	20	5	32	0.00	0.10	0.10	0.05
	20	8	24	0.45	0.05	0.15	0.00
20	8	32	0.25	0.05	0.05	0.00	

^aNumbers in bold correspond to correct model assignments.

ate to use monotone regression to infer lexicographic orders.

Table 4 summarizes the findings concerning the accuracy of the model assignment. The results suggest that the proposed testing procedure is quite robust. Overall, the method selects the correct preference model in about 90% of the cases. The accuracy of the assignment reduces from 98% to 81% when the magnitude of error increases from 5% to 20%; and it reduces from 94% to 84% when the number of profiles decreases from 32 to 24. Increasing the number of attributes has little effect on accuracy, which drops from 91% with five attributes to 88% with eight attributes. In general, model assign-

ment is more accurate for the more constrained standard, binary, and satisficing models, than it is for the binary-satisficing and linear models. The accuracy of the assignment is 98.75% for the former models, and drops to 75% for the latter models. As expected, the accuracy of the assignment increases with the number of product profiles and decreases with the amount of error in the pairwise comparisons.

There are four instances in Table 4 where the accuracy of the rule assignment is poor. These cases occur when the true model is linear or binary satisficing with 20% error and eight attributes. We expect the reason for this to be the small sample size. For confirmation, we performed additional simulation runs,

Table 5 Impact of Sample Size on Rule Assignment Accuracy

True model	Error level (%)	Number of attributes	Number of estimation profiles	% Correct assignment
Linear	20	8	24	35
	20	8	32	65
	20	8	48	90
	20	8	64	100
Binary satisficing	20	8	24	15
	20	8	32	70
	20	8	48	90
	20	8	64	90

the results of which are summarized in Table 5. As expected, the error in rule assignment steadily increases with increases in sample size.¹⁰

Finally, we reconsider the issue, discussed in §5, concerning the adequacy of inferring a lexicographic attribute ordering using traditional conjoint analysis. The added advantage of the simulation is that we can compare the ordering of the importance weights in a linear model to a true, lexicographic ordering of the attributes. We use the pairwise comparisons from a lexicographic model to estimate LINMAP, and arrange the attributes in decreasing order of their part-worth ranges. We find that this ordering does not accurately reproduce the true lexicographic ordering of the attributes. For example, consider the case of a true standard lexicographic model. Across experimental conditions, the average Kendall's tau correlation between the true and estimated attribute orderings is 0.48, its value decreasing from 0.61 to 0.34 as the error increases from 5% to 20%. The greedy algorithm obtains, on average, a corresponding tau value of 0.87 (see Table 3).

Summary. The simulation results suggest that the greedy algorithm does well in recovering the true lexicographic structure in the presence of error in the data. We do not recommend inferring an attribute ordering for a lexicographic rule using the results of a standard, part-worths model. The proposed greedy algorithm does well in recovering a lexicographic rule, and the assignment procedure we propose is quite accurate for discriminating among the alternative lexicographic models and for distinguishing

a lexicographic model from a linear compensatory model.

7. Conclusion

We describe two variants of lexicographic preference models. A consumer can use the variants alone or in combination with each other. We obtain necessary and sufficient conditions for a linear utility function to represent standard lexicographic models and their variants. We describe an algorithm for inferring the best-fitting lexicographic model from preference data over multiattribute alternatives. The algorithm is similar—in the sense that it does not have an error model—to the nonmetric methods for preference and choice scaling described by Toubia et al. (2004), Toubia et al. (2003), Srinivasan and Shocker (1973), and Kruskal (1965). The results of a simulation suggests that the greedy algorithm does well in recovering the true attribute ordering, and that the proposed method for assigning alternative preference models to consumers is robust. We find support for the use of lexicographic rules by two-thirds of the subjects in a study of consumer preferences for laptop computers. Among the different rules we test, a lexicographic rule allowing for indifference among attribute levels appears to be the one most appropriate for describing the preferences of a substantial proportion of the subjects. Finally, we illustrate how the lexicographic analysis can be used to obtain a hierarchical clustering of consumers, and how we can interpret the clustering as an aggregate market structure.

The assumption of finite divisibility of attributes plays a central role in the present research. For most marketing problems, there is little loss in assuming that attribute levels are finitely divisible. For example, we often make this assumption when estimating part-worth functions in conjoint analysis. In problems involving continuous attributes, it can often be convenient, but seldom necessary, to assume that such attributes as price, pick-up speed, gas mileage, weight, and length are real valued (or even denumerable) variables in the mathematical sense of the term. Although we think of such variables as being continuous, in reality prices are not discriminated beyond the smallest unit of currency; and features like time for acceleration to 60 mph, gas mileage, weight, and length are not measurable beyond some finite level of precision. Similarly, our assessment of such psychological constructs as attitudes and affect is subject to limitations of both human discrimination ability and those of our measuring instruments. It is reasonable to assume that these attributes are real-valued quantities if this assumption allows us to use results that extend to finite attributes as special cases. However, in such situations as concern us here, where a utility

¹⁰ As suggested by an anonymous reviewer, one can use holdout data for rule assignment. The results from additional simulation runs show good improvement in the percent of correct rule assignments when we use this assignment method. However, the improvement is substantially better when the holdout data is also used for estimation, which results in larger sample size. For example, suppose the true model is linear, the error rate is 20%, there are eight attributes, and we use 24 profiles for estimation and another 24 for holdout prediction. Then the accuracy of rule assignment increases from 35% when we use estimation data to 50% when we use holdout data. However, such accuracy reaches 90% when we use all 48 profiles for estimation.

function does not exist for real-valued attributes, it is appropriate to ask if something useful can be said by restricting attention to the case of finite attributes.¹¹

One useful area of future research is to examine an error model for lexicographic preferences, possibly in a Bayesian framework pooling data across respondents.¹² A second related extension is to consider probabilistic lexicographic preferences, where the probability of attribute orderings is inferred from the data using an error model. A third potentially fruitful area of research is the inference from preference data of mixed decision strategies, so that the overall judgments of a consumer represent lexicographic preferences over some attributes and trade-offs among other attributes. A fourth possible research area is to extend the present inference procedures to obtain segment-level estimates of lexicographic rules. A fifth useful area of research is to assess how consumers combine different rules for product consideration (e.g., Gilbride and Allenby 2004, Kohli and Jedidi 2005, Kohli et al. 2006) and then choice; how the preference structure changes over repeated measurements (DeSarbo et al. 2005); and how knowledge of preference functions and choice rules can guide user design of products (Randall et al. 2007). An example of this kind of research is the recent work by Yee et al. (2007), in which a two-step data collection procedure on the WWW is linked to a Greedoid method for inferring the structure of noncompensatory decision rules. Such work is useful if we are to develop models that not only predict well but also provide insight into why consumers make the decisions they do. Sixth, it may be useful to develop and examine the properties of adaptive conjoint models (see., e.g., Hauser and Toubia 2005) for inferring lexicographic (more generally, noncompensatory) preference structures. A final area of possible research concerns the trade-off between model complexity and fit when comparing multiple lexicographic models. The methods by Toubia et al. (2003, 2004) and Evgeniou et al. (2005) might be useful starting points in extending these considerations from standard conjoint models to methods for inferring lexicographic preference structures.

Appendix. Theorems and Proofs

We use the notation introduced in §2. In particular, we recall that $\min \Delta x_k = \min\{a_{nk}^k - a_{j-1}^k \mid j = 2, \dots, m\}$ denotes

¹¹ One can relax the assumption of finite divisibility for the case of two attributes. In this case, we can allow the less-important attribute to be defined over the real interval $[0, 1)$ and the more important attribute to be defined over the set of nonnegative integers $0, 1, \dots$. The resulting utility function is an “ordinary number,” the integer (fractional) part representing the utilities for the preferred (less-preferred) attribute.

¹² We note that Tversky’s (1972) elimination-by-aspects model is a probabilistic lexicographic model, not a model of lexicographic preference with error.

the smallest difference in the successive values of x_k ; and $\max \Delta x_k = a_{nk}^k - a_1^k$ denotes the difference between the largest and smallest values of x_k , for all $k = 1, \dots, m$. Theorem 1 characterizes the necessary and sufficient conditions under which a linear model represents lexicographic preferences.

THEOREM 1. Let

$$u(\mathbf{x}) = \beta_1 x_1 + \dots + \beta_m x_m, \quad (\text{A1})$$

where x_k is the k th-most important attribute, $k = 1, \dots, m$. Then $u(\mathbf{x})$ represents lexicographic preferences over the m attributes if, and only if,

$$\beta_k \min \Delta x_k > \sum_{j=k+1}^m \beta_j \max \Delta x_j, \quad \text{for all } k = 1, \dots, m. \quad (\text{A2})$$

PROOF. Consider two distinct alternatives, $\mathbf{x}, \mathbf{x}' \in M$. Let \mathbf{x} be lexicographically preferred to \mathbf{x}' . We first show that $u(\mathbf{x}) > u(\mathbf{x}')$. Let k denote the lowest-indexed attribute for which \mathbf{x} and \mathbf{x}' have different attribute levels; i.e., $x_j = x'_j$, $1 \leq j \leq k-1$, $x_k \neq x'_k$, $k = 1, \dots, m$. Then

$$u(\mathbf{x}) - u(\mathbf{x}') = \beta_k(x_k - x'_k) + \sum_{j=k+1}^m \beta_j(x_j - x'_j).$$

Now $x_k - x'_k \geq \min \Delta x_k$, and $x_j - x'_j \geq -\max \Delta x_j$ for all $j = k+1, \dots, m$. Thus,

$$u(\mathbf{x}) - u(\mathbf{x}') \geq \beta_k \min \Delta x_k - \sum_{j=k+1}^m \beta_j \max \Delta x_j > 0,$$

where the last inequality in the above expression follows from (A2).

We now show that if $u(\mathbf{x}) > u(\mathbf{x}')$, then $x_j > x'_j$ on the smallest index j for which $x_j \neq x'_j$. We prove the result by contradiction. Suppose $x_j < x'_j$ on the smallest index $j = k$ for which $x_j \neq x'_j$. Then

$$x_j = x'_j \quad \text{for all } j = 1, \dots, k-1, \quad \text{and} \quad x_k - x'_k \leq -\min \Delta x_k,$$

and so

$$\begin{aligned} u(\mathbf{x}) - u(\mathbf{x}') &= \beta_k(x_k - x'_k) + \sum_{j=k+1}^m \beta_j(x_j - x'_j) \\ &\leq -\beta_k \min \Delta x_k + \sum_{j=k+1}^m \beta_j(x_j - x'_j). \end{aligned}$$

The upper bound on the right-hand side of the inequality is attained when $x_j - x'_j = \max \Delta x_j$, for all $j = k+1, \dots, m$. Thus,

$$u(\mathbf{x}) - u(\mathbf{x}') \leq -\beta_k \min \Delta x_k + \sum_{j=k+1}^m \beta_j \max \Delta x_j.$$

It follows from (A2) that the right-hand side is always negative; i.e., $u(\mathbf{x}) < u(\mathbf{x}')$, which is the desired contradiction. \square

THEOREM 2. Suppose there exists an attribute ordering that perfectly reproduces an input rank ordering of product profiles. Then the greedy algorithm finds such an ordering.

PROOF. We prove the result by contradiction. Let $k = 1, \dots, m$ denote the attributes. If the greedy algorithm terminates in $t < m$ steps, we append the $m - t$ “unused” attributes in any arbitrary order to the end of the greedy solution. Without loss of generality, we assume that the greedy algorithm selects attribute k at step k , for all $k = 1, \dots, m$. Let \mathbf{G} denote the sequence of attributes $1, \dots, m$ selected by the greedy algorithm. We will sometimes refer to \mathbf{G} as the greedy sequence. Let $1, \dots, s, 1 \leq s \leq m$, denote the longest sequence of attributes in \mathbf{G} that is also a sequence in some optimal solution \mathbf{O} . That is, \mathbf{O} is the sequence

$$1, \dots, s, (s+1)^*, \dots, m^*, \quad \text{where } (s+1)^* \neq s+1.$$

If $s = m$, then the greedy solution is identical to an optimal solution. Also, $s = m - 1$ is not possible because then the last attribute in \mathbf{G} must also be optimal. We therefore consider $1 \leq s \leq m - 2$. We show that the sequence of attributes $1, \dots, s + 1$, appears in at least one optimal solution, which contradicts the claim that $1, \dots, s$, is the longest greedy sequence of attributes that appears in any optimal solution.

Let attribute $s + 1$ appear in the j th position of the sequence \mathbf{O} , where $s + 2 \leq j \leq m$. Suppose we remove attribute $s + 1$ from this position and insert it immediately before attribute $(s + 1)^*$ in the sequence \mathbf{O} . We denote the sequence thus obtained as \mathbf{C} :

$$1, \dots, s + 1, (s + 1)^*, \dots, (j - 1)^*, (j + 1)^*, \dots, m^*.$$

We use the following four observations to show that \mathbf{C} is also an optimal sequence of attributes.

(1) As the optimal solution has no associated reversals, there are also no reversals in any sequence obtained by truncating \mathbf{O} . The sequence $1, \dots, s, (s + 1)^*$, appears in \mathbf{O} and produces no reversals.

(2) Because attribute $(s + 1)^*$ is available but not selected at step $s + 1$ by the greedy algorithm, there are also no reversals in the partitions created by the attribute sequence $1, \dots, s + 1$, which appears in the greedy sequence \mathbf{G} .

(3) Let \mathbf{O}' denote the following subsequence in \mathbf{O} :

$$1, \dots, s, (s + 1)^*, \dots, (j - 1)^*.$$

Let \mathbf{C}' denote the following subsequence in \mathbf{C} :

$$1, \dots, s + 1, (s + 1)^*, \dots, (j - 1)^*.$$

We claim that \mathbf{C}' produces no reversals. This follows from three observations. First, from Observation (1) above, \mathbf{O}' produces no reversals. Second, from Observation (2) above, $1, \dots, s + 1$, also produces no reversals. Third, consider the sets of pairs over which incremental reversals are computed upon the introduction of

$$(s + 1)^*, \dots, (j - 1)^*,$$

in \mathbf{O}' and in \mathbf{C}' . The latter set of pairs comprise a subset of the former set of pairs, because $1, \dots, s + 1$, creates no fewer partitions than $1, \dots, s$. It follows that \mathbf{C}' can have no more reversals than \mathbf{O}' . However, \mathbf{O}' has no reversals, and so neither does \mathbf{C}' .

(4) The number of partitions created by any collection of attributes is independent of the ordering of these attributes. Thus, the incremental reversals introduced by the attribute sequence

$$(j + 1)^*, \dots, (m)^*,$$

which we denote \mathbf{T} , does not depend on the position of attributes that precede this sequence in a lexicographic ordering of the attributes. As this sequence of attributes produces no incremental reversals in the optimal sequence \mathbf{O} of the attributes, it follows that it also produces no incremental reversals in the candidate sequence \mathbf{C} , because the same attributes, $1, \dots, s + 1, (s + 1)^*$, precede \mathbf{T} in both \mathbf{O} and \mathbf{C} , albeit in a different sequence.

Observations (3) and (4) imply that \mathbf{C} is also an optimal lexicographic ordering of the attributes, which is the desired contradiction. \square

References

- Bridges, D. S. 1983. A numerical representation of intransitive preferences on a countable set. *J. Econom. Theory* **30** 213–217.
- Broder, A. 2000. Assessing the empirical validity of the “take-the-best” heuristic as a model of human probabilistic inference. *J. Experiment. Psych.: Learn., Memory, Cognition* **26**(5) 1332–1346.
- Chateaufneuf, A. 1987. Continuous representation of a preference relation on a connected topological space. *J. Math. Econom.* **16** 139–146.
- Colman, A. M., J. A. Stirk. 1999. Singleton bias and lexicographic preferences among equally valued alternatives. *J. Econom. Behav. Organ.* **40**(4) 337–351.
- Debreu, G. 1954. Representation of a preference ordering by a numerical function. R. M. Thrall, C. H. Coombs, R. L. Davis, eds. *Decision Processes*. John Wiley, New York, 159–165.
- DeSarbo, W. S., D. K. H. Fong, J. C. Liechty. 2005. Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Sci.* **24**(2) 285–293.
- Dhar, R., S. M. Nowlis. 1999. The effect of time pressure on consumer choice deferral. *J. Consumer Res.* **25**(4) 369–384.
- Drolet, A., M. F. Luce. 2004. Cognitive load and trade-off avoidance. *J. Consumer Res.* **31** 63–77.
- Evgeniou, T., C. Boussios, G. Zacharia. 2005. Generalized robust conjoint estimation. *Marketing Sci.* **24**(3) 415–429.
- Fishburn, P. C. 1974. Lexicographic orders, utilities and decision rules: A survey. *Management Sci.* **20** (11) 1442–1471.
- Fishburn, P. C. 1975. Axioms for lexicographic preferences. *Rev. Econom. Stud.* **42**(3) 415–419.
- Ford, J. K., N. Schmitt, S. L. Schechtman, B. M. Hults, M. L. Doherty. 1989. Process tracing methods: Contributions, problems, and neglected research questions. *Organ. Behav. Human Decision Processes* **43** 75–117.
- Garey, M. R., D. S. Johnson. 1979. *Computer and Intractability*. W. H. Freeman and Company, New York.
- Gigerenzer, G., U. Hoffrage, H. Kleinbolting. 1991. Probabilistic mental models: A Brunswikian theory of confidence. *Psych. Rev.* **98** 506–528.
- Gilbride, T. J., G. M. Allenby. 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Sci.* **23**(3) 391–406.
- Gonzalez-Vallejo, C., A. Bonazzi, A. J. Shapiro. 1996. Effects of vague probabilities and of vague payoffs on preference: A model comparison analysis. *J. Math. Psych.* **40** 130–140.
- Grover, R., V. Srinivasan. 1987. A simultaneous approach to market segmentation and market structure. *J. Marketing Res.* **24**(May) 139–153.

- Hauser, J. R., S. M. Shugan. 1980. Intensity measures of consumer preference. *Oper. Res.* **28**(2) 278–320.
- Hauser, J. R., O. Toubia. 2005. The impact of utility balance and endogeneity in conjoint analysis. *Marketing Sci.* **24**(3) 498–507.
- Jedidi, K., R. Kohli. 2005. Probabilistic subset-conjunctive models for heterogeneous consumers. *J. Marketing Res.* **42**(November) 483–494.
- Kahn, B., J. Baron. 1995. An exploratory study of choice rules favored for high-stakes decisions. *J. Consumer Psych.* **4**(4) 305–328.
- Kannan, P., G. Wright. 1991. Modeling and testing structured markets: A nested logit approach. *Marketing Sci.* **10** (Winter) 58–82.
- Knoblauch, V. 2000. Lexicographic orders and preference representation. *J. Math. Econom.* **34**(2) 255–267.
- Kohli, R. 1999. Lexicographic systems. *Complexity* **4**(4) 15–25.
- Kohli, R., K. Jedidi. 2005. Probabilistic subset conjunction. *Psychometrika* **70**(4) 737–757.
- Kohli, R., R. Krishnamurti, K. Jedidi. 2006. Subset conjunctive rules for breast-cancer diagnosis. *Discrete Appl. Math.* **154**(7) 1100–1132.
- Kruskal, J. B. 1965. Analysis of factorial experiments by estimating monotone transformations of the data. *J. Roy. Statist. Soc., Ser. B (Methodological)* **27**(2) 251–263.
- Manrai, A. K., P. K. Sinha. 1989. Elimination-by-cutoffs. *Marketing Sci.* **8**(2) 133–152.
- Martignon, L., U. Hoffrage. 1999. Why does one-reason decision making work? A case study in ecological rationality. G. Gigerenzer, P. M. Todd, and the ABC Research Group, eds. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York, 119–140.
- Martignon, L., U. Hoffrage. 2002. Fast, frugal and fit: Simple heuristics for paired comparison. *Theory Decision* **52** 29–71.
- Martignon, L., M. Schmitt. 1999. Simplicity and robustness of fast and frugal heuristics. *Minds Machines* **9** 565–593.
- Mosier, C. I. 1951. Problems and designs of cross-validation. *Educational Psych. Measurement* **11** 5–11.
- Payne, J. W., J. R. Bettman, E. L. Johnson. 1988. Adaptive strategy selection in decision making. *J. Experiment. Psych.: Learn., Memory, Cognition* **14** 534–552.
- Randall, T., C. Terwiesch, K. T. Ulrich. 2007. User design of customized products. *Marketing Sci.* **26**(2) 268–280.
- Roedder-John, D. 1999. Consumer socialization of children: A retrospective look at twenty-five years of research. *J. Consumer Res.* **26**(3) 183–213.
- Russell, G., W. Kamakura. 1994. Understanding brand competition using micro and macro scanner data. *J. Marketing Res.* **31**(May) 289–303.
- Schmitt, M., L. Martignon. 1999. Complexity of lexicographic strategies on binary cues. Working paper, Lehrstuhl Mathematik und Informatik, Ruhr Universität, D-44780, Bochum, Germany.
- Simon, H. A. 1956. Rational choice and the structure of the environment. *Psych. Rev.* **63** 129–138.
- Slovic, P. 1975. Choice between equally valued alternatives. *J. Experiment. Psych.: Human Perception Performance* **1** 280–287.
- Srinivasan, V., A. D. Shocker. 1973. Estimating the weights for multiple attributes in a composite criterion using pairwise judgments. *Psychometrika* **38**(4) 473–493.
- Toubia, O., J. R. Hauser, D. I. Simester. 2004. Polyhedral methods for adaptive choice-based conjoint analysis. *J. Marketing Res.* **41**(1) 116–131.
- Toubia, O., J. R. Hauser, D. I. Simester, E. Dahan. 2003. Fast polyhedral adaptive conjoint estimation. *Marketing Sci.* **22**(3) 273–303.
- Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psych. Rev.* **79** 281–299.
- Tversky, A., S. Sattah, P. Slovic. 1988. Contingent weighting in judgment and choice. *Psych. Rev.* **95** 371–384.
- Urban, G. L., P. L. Johnson, J. R. Hauser. 1984. Testing competitive market structures. *Marketing Sci.* **3**(2) 83–112.
- Varian, H. R. 1984. *Microeconomic Analysis*. W. W. Norton, New York.
- Wakker, P. 1988. Continuity of preference relations for separable topologies. *Internat. Econom. Rev.* **29** (1) 105–110.
- Westenberg, M. R. M., P. Koele. 1994. Multi-attribute evaluation processes: Methodological and conceptual issues. *Acta Psychologica* **87** 65–84.
- Yee, M., E. Dahan, J. R. Hauser, J. Orlin. 2007. Greedoid-based noncompensatory consideration-then-choice inference. *Marketing Sci.* Forthcoming.