

Measuring a Lexicographic Bias in Linear Conjoint Analysis Models

Joachim Giesen and Sören Laue and Katrin Nimczick

Institut für Informatik, Friedrich-Schiller-Universität Jena

Abstract. Conjoint analysis aims at measuring preferences on a set of options that are elements in the Cartesian product of parameter sets. Preferences are measured jointly, e.g., by comparing two options, where all parameters contribute to the outcome of the comparison. For optimizing the options and for deriving preference models from measurements it is helpful to know how much each parameter sets contributed to the observed result, or at least which parameters contributed more than others. In a strict lexicographic preference order one parameter decides the outcome and only if the same value for this parameter is present in the compared options, the second most important parameter becomes decisive. In practice a strict lexicographic preference model often may not be adequate, but still some parameters can be more important than others. We discuss methods to measure a lexicographic bias in linear conjoint analysis models, i.e., we quantify how close this model is to a strict lexicographic ordering in a metric sense. We apply our methods to data sets obtained in two conjoint measurement studies.

1 Introduction

Preferential choices by individuals on a multi-parameter set of options have been studied from two points of view, namely (1) behavioral, and (2) empirical. Under the behavioral point of view one investigates strategies that individuals may use when confronted with a (multi-parameter) choice task, whereas the empirical approach is to fit a preference model to a set of observed choices.

Behavioral strategies address the effort vs. accuracy trade-off faced by an individual confronted with a choice task, i.e., it may not always pay off for the individual—in terms of effort—to determine the best option. Thus behavioral strategies tend to be simple in terms of the required effort and are mostly of non-compensatory nature. Non-compensatory strategies basically rank the different aspects (parameter levels) of the choice options whereas compensatory strategies weigh them. A crucial difference is, lacking a highly ranked aspect cannot be compensated for by other aspects of an option, whereas a highly weighted aspect still can be outweighed by a combination of other aspects. Weighting aspects and combining the weights typically requires more effort than just ranking the aspects. Ranking the aspects essentially means employing a lexicographic strategy. On the other hand, the empirical models used by practitioners, e.g., in market research, are more general and mostly of compensatory nature.

Here we consider a special class of compensatory preference models, namely linear conjoint analysis models, that are frequently used in market research to represent preferences assessed in conjoint measurement studies. Linear conjoint analysis models estimate weights assigned to the different aspects of the choice options by the individuals whose preferences have been assessed. The weight of an option is then assumed to be the sum of the weights of its aspects. These models have turned out to be well suited for predicting future choices by the same or similar individuals, but it is not straightforward to determine the impact of the different parameter sets on these choices from a linear model, in contrast to the parameter levels. The impact of the parameter sets (or short parameters) can often be attributed to a lexicographic bias in the linear model. Knowing the unimportant parameters can be helpful information when optimizing the options. For example if the preferred level of an unimportant parameter is costly to realize, then one can stick with a lesser level since the effect of improving to the better level on the whole option is small. Similarly knowing the important parameters can be helpful.

We introduce four different methods to define and compute a lexicographic bias in a linear model and discuss their advantages and disadvantages. We apply the methods to data from two conjoint measurement studies that we have conducted earlier, and observe a strong lexicographic bias for the first study and a lack of a lexicographic bias in the second study. The ability of our methods to detect a lexicographic bias in data is confirmed by comparing the predictive power of linear conjoint analysis models and non-compensatory lexicographic models derived from linear conjoint analysis models on holdout data. For the first study—where we detect a strong lexicographic bias—the lexicographic models perform similarly as the underlying linear conjoint analysis model, whereas for the second study where we detected a lack of a strong lexicographic bias the corresponding lexicographic models perform, as expected, much worse than the underlying linear model.

2 Linear conjoint analysis models

Conjoint analysis comprises a family of techniques to assess and represent preferences on a multi-parameter set of options [6]. Here we want to denote the set of options as $A = A_1 \times \dots \times A_n$, where A_i is the set of values for the i 'th parameter. The assessment stage of conjoint analysis involves a conjoint measurement, i.e., a measurement on an element of A (jointly on all parameter values present in the element) or more generally on an element of A^k (the Cartesian product of k copies of A). That is, in a conjoint measurement the parameter values are considered jointly. Examples of conjoint measurements include among others the following: (1) Given an option $a \in A$, it is assessed if an individual likes or dislikes this option. Note that at this stage it is not obvious how the different parameter levels present in a contributed to the observed outcome of the measurement. (2) Given $k \geq 2$ options $a_1, \dots, a_k \in A$, it is assessed which of these options is most preferred by an individual. The latter example is known as choice based conjoint

analysis since the individual has to choose from k options. In the following we want restrict our discussion to choice based conjoint analysis with $k = 2$ choice options, i.e., binary choices. The techniques developed here, however also apply in the general case.

The goal of (choice based) conjoint analysis is to derive a model from a sequence of conjoint measurements. Most popular are linear models that assign a numerical value to every parameter level. These values are also called *partworth values* and denoted as $v_i(a_i)$ for $a_i \in A_i$. The linear model is then given as

$$v : A \rightarrow \mathbb{R}, a = (a_1, \dots, a_n) \mapsto \sum_{i=1}^n v_i(a_i).$$

The popularity of linear models has its roots in axiomatic conjoint measurement theory [10] that provides axioms guaranteeing that ordinal measurements as in choice based conjoint analysis result in an interval scale on the parameter sets A_i that can be combined linearly into an interval scale for A . The interval scale on A is invariant under transformations of the form $\alpha v_i(\cdot) + \beta_i$ with $\alpha > 0$ and $\beta_i \in \mathbb{R}$. Note that the scale factor α needs to be the same for all $i = 1, \dots, n$, also note already here that in practical situations the axioms of conjoint measurement are almost never met. We will comment more on this later.

A machine learning approach to estimate partworth values for finite parameter sets A_i has been introduced by Evgeniou et al. [3] who build on ideas of regularization. This approach is a special case of Joachims' Ranking SVM [8] and related to Herbrich et al.'s SVM approach to ordinal regression [7]. In [3] it has been suggested to compute the partworth values as the solution of the following convex quadratic program for l choice measurements

$$\begin{aligned} \min_{v_i(a), z_j} \quad & \sum_{i=1}^n \sum_{a \in A_i} v_i(a)^2 + c \sum_{j=1}^l z_j \\ \text{s.t.} \quad & v(a) - v(b) + z_j \geq 1, \\ & \text{if } a \succeq b \text{ in the } j\text{'th choice measurement.} \\ & z_j \geq 0, j = 1, \dots, l \end{aligned}$$

Note that the objective function describes the classical trade-off between model complexity (the regularization term $\sum_{i=1}^n \sum_{a \in A_i} v_i(a)^2$) and accuracy on the training data (the error term $\sum_{j=1}^l z_j$, where z_j is a slack variable for the j 'th choice measurement). It is also interesting to have a closer look at the constraints $v(a) - v(b) + z_j \geq 1$ with $a, b \in A$. If we disregard the slack variable z_j and shrink the margin from 1 to 0, then the constraint on the partworth values can be written as $\langle v, \chi_a - \chi_b \rangle \geq 0$, where $v \in \mathbb{R}^m$, with $m = \prod_{i=1}^n |A_i|$, is the vector of all partworth values, and $\chi_a \in \{0, 1\}^m$ is the characteristic vector of $a \in A$, i.e., χ_a has entry 1 whenever the corresponding parameter level is present in a and 0 otherwise. Constraints of the form $\langle v, \chi_a - \chi_b \rangle \geq 0$ describe halfspaces with normal vector $\chi_a - \chi_b$ whose bounding hyperplane

$$\langle v, \chi_a - \chi_b \rangle = 0 \tag{1}$$

contains the origin. Note that there are only finitely many normal vectors $\chi_a - \chi_b$ possible. The arrangement of the corresponding hyperplanes divides \mathbb{R}^m into a finite number of cells, and each of the cells corresponds to a different partial ordering (ranking) of the elements in A . Note however that not all (partial) orderings of A have a representation as a cell of A . We call the (partial) orderings that correspond to the cells of the arrangement *realizable*. A solution to the convex quadratic program from above is a point in one of the cells of the hyperplane arrangement and thus represents a realizable (partial) ordering of A . In general we cannot interpret this solution as an interval scale as can be seen from the following dimension analysis: an interval scale for A has $n + 1$ degrees of freedom, namely, the value of $\alpha > 0$ and the values for the $\beta_i \in \mathbb{R}$. Thus any cell of dimension larger than $n + 1$ must contain representatives for different interval scales. Any such cell is the intersection of halfspaces with normal vectors of the form $\chi_a - \chi_b$, and the ordinal measurements that correspond to these normal vectors cannot satisfy the axioms of the axiomatic theory of conjoint measurement. Nevertheless, the more practical approach towards conjoint analysis of computing a linear model (which not necessarily provides a unique interval scale) proved to be successful in practice.

3 Choices between multi-parameter options

3.1 Choice strategies (behavioral)

Psychologists have studied for a long time how individuals choose between multi-parameter options as we have discussed them in the context of choice based conjoint analysis. In the light of humans as actors of bounded rationality [12, 5, 11] it is natural to consider the effort-accuracy trade-off that individuals face when choosing between multi-parameter options, i.e., it might not always pay off to identify the better (best) option, especially if this requires some effort. One can distinguish two classes of strategies that individuals may apply in multi-parameter choice situations:

- (1) non-compensatory strategies, and
- (2) compensatory strategies.

One possible compensatory strategy is reflected in the additive models of conjoint analysis, namely an individual could weight the different aspects (parameter levels) of an option, sum up the weights (linearity assumption) and then choose the option with the highest weight. In this way a particularly bad aspect (level) of an important parameter of an option could be outweighed by several other good aspects of less important parameters. Non-compensatory strategies have been discussed more extensively in the psychology literature. One of the first studied non-compensatory strategies is Tversky's *elimination by aspects* [13], which besides being non-compensatory is also randomized. In the *elimination by aspects* strategy individuals associate weights with the parameter levels (aspects) and successively remove (unwanted) levels from the remaining ones at random,

where the probability for a level to be removed is proportional to its weight. The options are then ranked according to the random ranking of the aspects, i.e., the rank of an option is determined by its last eliminated level where ties are broken accordingly. Note that in that in *elimination by aspect* the ranking of the aspects will in general mix levels from different parameters. Hence a restricted deterministic form of *elimination by aspects* is first lexicographically ranking the parameters and then ranking the levels within the parameters. Here we want to focus on the latter lexicographic strategy and refer to it as *strict lexicographic ranking*.

3.2 Choice models (empirical)

Choice strategies have their counterparts in choice models. A choice strategy is a means of choosing between multi-parameter options whereas a choice model aims at explaining an observed sequence of observed (measured) choices. We have already discussed linear models of conjoint analysis in the previous section. Our goal now is to compare linear models with models that rank the options lexicographically. Kohli and Jedidi [9] observed that both strategies (*deterministic elimination by aspects* and *strict lexicographic ranking*, respectively, can for finite parameter sets always be represented by a linear model. That is, the class of linear models is not only richer than the class of lexicographic models, it even contains the latter.

Next we are investigating the following question: given a linear model, how close is it to a lexicographic model, and how can a close lexicographic model be computed.

4 Deriving lexicographic models from linear models

We want to derive from a given linear model, e.g. one computed from from choice measurements as described in Section 2, a “close” lexicographic model (which by the observation of Kohli and Jedidi [9] can be expressed again as a linear model).

Note that a linear model always specifies a ranking of the levels within every parameter set. Hence the task to determine a close strict lexicographic model reduces to determining a suited ranking of the parameter sets. There is no unique or obvious method to compute such a ranking on the parameter sets but certain requirements should be satisfied by any such method, namely

- (1) the method should actually compute a ranking, i.e., a (partial) order, and
- (2) the result of the method should be independent of the choice of representative for equivalent linear models.

Here we consider two linear models equivalent if and only if they encode the same (partial) order information. Remember that a linear model is represented by a point in \mathbb{R}^m (see Section 2), and observe that two linear models are equivalent if they (viewed as points in \mathbb{R}^m) fall into the same cell of the hyperplane arrangement given by Equations 1.

In the following we discuss four methods to compute a ranking of parameter sets from a linear model.

4.1 Max-min partworth difference

The leading conjoint analysis software company Sawtooth Software Inc. [1] is using the following technique to assess the relative importance of the parameter sets in their adaptive conjoint analysis software ACA: respondents in a conjoint study are asked to rate on scale with four levels how important it is for a given parameter set to improve from the worst to the best level. That is, essentially they are asked for a quantization (into four levels) of the difference of the partworth values of the best and the worst parameter level. Weighting the parameter sets with the latter differences immediately provides a (partial—in the presence of ties) order on the parameter sets, thus satisfying Requirement 1. The derived ranking also satisfies Requirement 2: assume to the contrary that it does not, i.e., there exist parameter sets A_i and A_j such that there are equivalent partworth value functions v_i, v_j and \hat{v}_i, \hat{v}_j satisfying the following relations for the best levels $b_i \in A_i, b_j \in A_j$ and the worst levels $w_i \in A_i, w_j \in A_j$:

$$v_i(b_i) \geq v_i(w_i), \quad v_j(b_j) \geq v_j(w_j), \quad \hat{v}_i(b_i) \geq \hat{v}_i(w_i), \quad \text{and} \quad \hat{v}_j(b_j) \geq \hat{v}_j(w_j),$$

which implies by using the assumption that Requirement 2 is violated

$$v_i(b_i) - v_i(w_i) \geq v_j(b_j) - v_j(w_j), \quad \text{and} \quad \hat{v}_j(b_j) - \hat{v}_j(w_j) \geq \hat{v}_i(b_i) - \hat{v}_i(w_i).$$

Then by the linearity of value functions

$$v(\dots, b_i, \dots, w_j, \dots) \geq v(\dots, w_i, \dots, b_j, \dots),$$

but

$$\hat{v}(\dots, w_i, \dots, b_j, \dots) \geq \hat{v}(\dots, b_i, \dots, w_j, \dots).$$

Hence v and \hat{v} cannot be equivalent in contradiction to the assumed equivalence of the partworth value functions v_i and \hat{v}_i , and v_j and \hat{v}_j , respectively.

4.2 Partworth variance

The max-min partworth difference method to rank parameter sets satisfies both of our requirements but is using only limited information from the underlying linear model, namely the partworth values of two levels for every parameter set. Weighting the parameter sets A_i by the variance

$$\text{Var}(A_i) = \frac{1}{|A_i| - 1} \sum_{a \in A_i} \left(v_i(a) - \frac{1}{|A_i|} \sum_{a \in A_i} v_i(a) \right)^2$$

of their partworth values takes more information into account and obviously also satisfies Requirement 1. Weighting the parameter sets by the variance of

their partworth values has been commonly used to rank parameter sets, but unfortunately Requirement 2 need not be satisfied by this ranking scheme as can be seen from the following example: given two parameter sets A_1 and A_2 with six levels each, i.e., $n = 2$ and $m_1 = m_2 = 3$. The following two partworth value vectors

$$\begin{aligned} v_1(A_1) &= (260, 50, -310) & \text{and} & & v_2(A_2) &= (310, -20, -290) \\ \hat{v}_1(A_1) &= (229, 112, -341) & \text{and} & & \hat{v}_2(A_2) &= (310, -20, -290) \end{aligned}$$

encode equivalent linear models, i.e., they induce the same order on the set $A = A_1 \times A_2$, but $\text{Var}(A_1) < \text{Var}(A_2)$ for the partworth vectors v_1, v_2 , and $\text{Var}(\hat{v}_1) > \text{Var}(\hat{v}_2)$ for the partworth vectors \hat{v}_1, \hat{v}_2 . A contradiction.

4.3 Trade-offs

In marketing applications conjoint analysis is mainly used to study how consumers trade-off different parameters (attributes) of a product against each other. For example consider a very simple model of cars with two attributes *maximum speed* and *fuel efficiency*. For technical reasons it is not possible to build a car that optimizes both attributes, i.e., fast cars cannot be fuel efficient. For a population of potential car buyers it is interesting to study how they trade-off the two parameters. This can be the basis for a market share prediction for, e.g., sport cars (speed wins the trade-off), family cars (compromise between speed and efficiency) and economic cars (efficiency wins the trade-off). Note that trade-offs are related to ranking the different parameters.

Given a linear model one could try to use trade-offs to rank the parameter sets. A linear model provides a ranking of the levels within each parameter set, i.e., just ranking the levels according to their partworth values. Consider now two parameter sets A_1 and A_2 whose levels $\{a_{11}, \dots, a_{1m_1}\}$ and $\{a_{21}, \dots, a_{2m_2}\}$ are ordered decreasingly according to their partworth values. A trade-off is a comparison of the form

$$(a_{1i}, a_{2j}) \quad \text{vs.} \quad (a_{1l}, a_{2k})$$

with $i > l$, $i, l \in \{1, \dots, m_1\}$ and $j < k$, $j, k \in \{1, \dots, m_2\}$. If

$$v_1(a_{1i}) + v_2(a_{2j}) \geq v_1(a_{1l}) + v_2(a_{2k}),$$

then we say that A_1 wins this trade-off against A_2 , and otherwise A_2 wins the trade-off against A_1 . If A_1 wins more trade-offs than A_2 , then A_1 is ranked higher as a set than A_2 , otherwise A_2 is ranked higher than A_1 . Obviously, this ranking satisfies Requirement 2, but it does not satisfy Requirement 1 in general. In fact, we claim that the trade-off relation is transitive for $A = A_1 \times \dots \times A_n$ with $|A_i| \leq 3$ for $1 \leq i \leq n$ but not transitive in general. Consider the following example. Let $A = A_1 \times A_2 \times A_3$ and let $A_1 = \{a_{11}, a_{12}, a_{13}\}$, $A_2 = \{a_{21}, a_{22}, a_{23}, a_{24}\}$ and $A_3 = \{a_{31}, a_{32}, a_{33}\}$. Let the partworth values be as follows:

$$\begin{aligned} v_1(a_{11}) &= 16, & v_1(a_{12}) &= 2, & \text{and} & v_1(a_{13}) &= 1, \\ v_2(a_{21}) &= 22, & v_2(a_{22}) &= 14, & v_2(a_{23}) &= 11, & \text{and} & v_2(a_{24}) &= 1, \\ v_3(a_{31}) &= 17, & v_3(a_{32}) &= 8, & \text{and} & v_3(a_{33}) &= 1. \end{aligned}$$

Then, it follows that A_1 wins 10 : 8 against A_2 , and A_2 wins 10 : 8 against A_3 but A_3 wins 5 : 4 against A_1 , hence proving that the trade-off relation is not transitive in the general case. A simple calculation shows that the trade-off relation for $|A_i| \leq 3$ for $1 \leq i \leq n$ coincides with the max-min relation and hence is transitive for this special case.

4.4 Metrically closest lexicographic model

The trade-off method is in contrast to the max-min and the variance methods of combinatorial nature. A more direct combinatorial method is to compute the closest lexicographic model in terms of a metric on the permutation group S_m , where $m = \prod_{i=1}^n |A_i|$. That is, we minimize the distance of the order on $A = A_1 \times \dots \times A_n$ induced by the linear model and the order given by a strict lexicographic model. To do so we need to test $n!$ strict lexicographic models—remember that the order within the parameter sets is already fixed by the linear model. Any such method automatically satisfies Requirements 1 and 2. There are several metrics on permutations group known and we will discuss four of them briefly. Let always π and σ be two permutations on S_m .

Hamming distance The Hamming distance is defined as

$$H(\pi, \sigma) = |\{i \in \{1, \dots, m\} \mid \pi(i) \neq \sigma(i)\}|.$$

The Hamming distance though it is a metric on S_m is not particularly well suited for our application: the Hamming distance of a permutation $\pi \in S_m$ to the one that is obtained from π by swapping the elements of rank i in π with the ones of rank $i + 1$, for $i = 2k - 1, k = 1, \dots, \lfloor m/2 \rfloor$ is maximal, but in terms of preferences both permutations are actually very close since the maximum rank difference of any element is one. This demonstrates that just being a metric on S_m is not sufficient. Metrics on S_m that are frequently used in the context of preference orderings are Spearman's footrule and rank correlation metrics, and Kendall's tau metric, see [2] for more details.

Spearman's footrule metric Spearman's footrule metric is defined as

$$F(\pi, \sigma) = \sum_{i=1}^m |\pi(i) - \sigma(i)|.$$

Its mean is $(m^2 - 1)/3$, its variance is $(m + 1)(2m^2 + 7)/45$, and the maximum possible distance is $m(m - 1)/2$. Spearman's footrule metric is closely related to Kendall's tau metric $K(\pi, \sigma)$, i.e., the minimum number of pairwise adjacent transpositions taking π^{-1} to σ^{-1} . Both metrics are related as follows, $K \leq F \leq 2K$.

Spearman's rank correlation metric Spearman's rank correlation metric is defined as

$$R(\pi, \sigma) = \sum_{i=1}^m (\pi(i) - \sigma(i))^2.$$

Its mean is $(m^3 - m)/6$, its variance is $m^2(m-1)(m+1)^2/36$, and the maximum possible distance is $(m^3 - m)/3$. Spearman's rank correlation metric is also often referred to as Spearman's rho.

4.5 Computational costs

The computational cost of deriving a lexicographical model from a linear model is:

1. $O(\sum_{i=1}^n |A_i| + n \log n)$ for the max-min partworth difference and the partworth variance methods.
2. $O((\max_{i=1, \dots, n} |A_i|^2) n \log n)$ for the trade-offs method.
3. $O(n!)$ for computing the metrically closest model (if we need to compare to every permutation of the attributes). Note though that in many cases n is fairly small (e.g., five in both of our examples, see Section 6).

5 Measuring a lexicographic bias

So far we have discussed methods to compute a strict lexicographic model from a linear model. Next we want to use such a derived strict lexicographic model to measure a lexicographic bias in the underlying linear model. To this end we also use the metrics on permutation groups introduced in Subsection 4.4 and measure the distance of the order induced by the linear model and the order given by the derived strict lexicographic model. Intuitively, there is a strong lexicographic bias if this distance is small. This leaves us with the problem to judge what it means for a distance to be small. A natural option is to compare the distance to either the maximum possible distance or to the mean of a metric (see Subsection 4.4 for these measures for the Spearman metrics). But these yard sticks (mean or maximum possible distance) are problematic in the sense that the permutations induced by linear models (realizable permutations/orderings) are only a subset of the permutation group, and so far we are lacking a good understanding of how these realizable permutations are distributed in the whole permutation group. Here we choose to take a practical approach to mitigate this problem and randomly generate instances of a linear model for which we derive a strict lexicographic model each using one of the four techniques from Section 4. The empirical distribution of these distances provides a better yard stick than the aforementioned measures, see Figures 1 to 13.

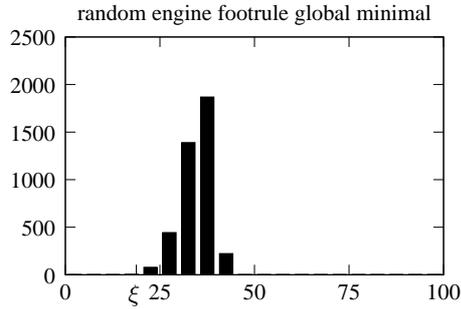


Fig. 1. Distance distribution to the optimal lexicographic model with respect to Spearman's footrule metric.

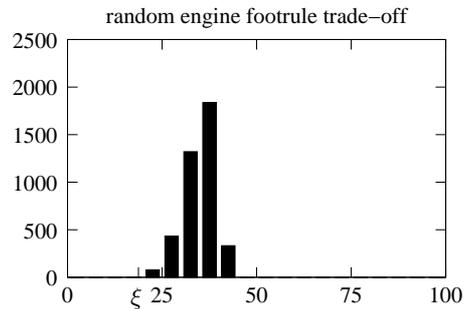


Fig. 2. Distance distribution to lexicographic model computed by the trade-off method.

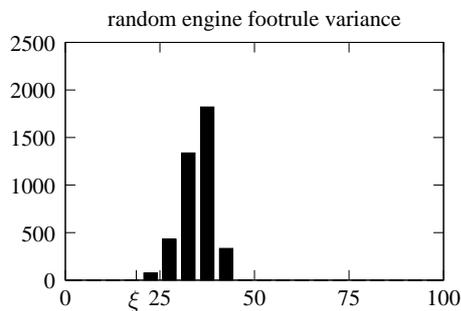


Fig. 3. Distance distribution to lexicographic model computed by the variance method.

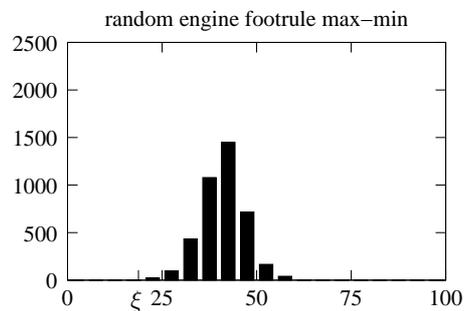


Fig. 4. Distance distribution to lexicographic model computed by the max-min method.

6 Results

We compare the different methods to compute a strict lexicographic model from a linear model on preference data that originated from two larger user studies. The first study measured the perceived quality in a visualization task [4] and the second study measured the importance of various skills for IT professionals. When computing the partworth values with the convex quadratic program from Section 2 we optimized the regularization parameter c using cross validation.

6.1 Visualization study

Let us first provide briefly the background of the visualization study: the purpose of volume visualization is to turn 3D volume data into images that allow a user to gain insight into the data. Turning volume data into images is a highly parameterized process among the many parameters are for example:

- (a) The choice of color scheme: often there is no natural color scheme for the data, but even when it exists it need not best

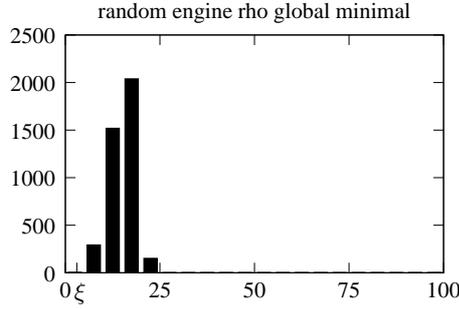


Fig. 5. Distance distribution to the optimal lexicographic model with respect to Spearman's rank correlation metric.

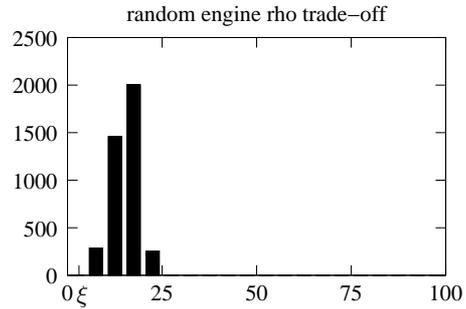


Fig. 6. Distance distribution to lexicographic model computed by the trade-off method.

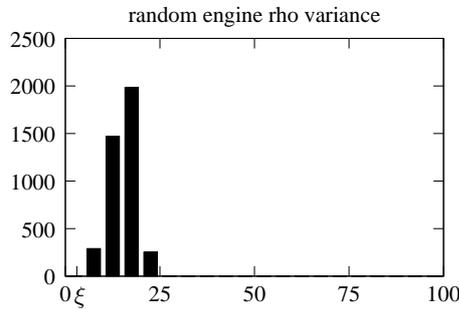


Fig. 7. Distance distribution to lexicographic model computed by the variance method.

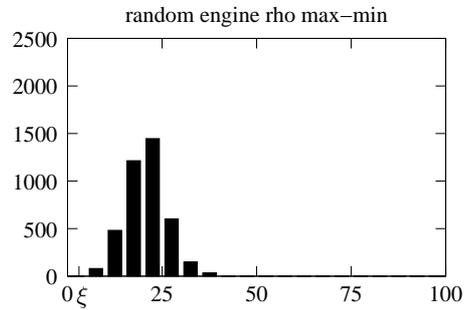


Fig. 8. Distance distribution to lexicographic model computed by the max-min method.

- (b) The viewpoint: an image is a 2D projection of the 3D data, but not all such projections are equally valuable in providing insights.
- (c) Other parameters like image resolution or shading schemes.

In our study [4] we were considering five parameters (with 2 to 15 levels each) for two data sets (foot and engine) giving rise to 2250 (foot) or 2700 (engine) options, respectively. Note that options here are images, i.e., different renderings of the data sets. On these data sets we were measuring preferences by either asking for the better liked image (aesthetics), or for the image that shows more detail (detail). That is, in total we conducted four studies (foot-detail, foot-aesthetics, engine-detail, and engine-aesthetics).

To quantify a lexicographic bias we compute linear conjoint analysis models for the four studies using the optimization method described in Section 2 from which we derive four strict lexicographic models, namely the ones computed by the max-min, variance, trade-off, and metric (Spearman footrule and rho, respectively) methods from Section 4.

In Figures 1 to 4 we show histograms of Spearman’s footrule distances between randomly generated instances of a linear model (having the same structure as the corresponding model in the visualization study) to a derived lexicographic model using one of our four methods from Section 4. The random instances are random partworth value vectors, where we make sure that we only take one vector for each cell of the arrangement generated by Inequalities 1, i.e., at most one for each realizable ordering. In these figures ξ denotes the distance of the linear model computed for the engine-aesthetics study data to the derived strict lexicographic model.¹ More specifically, each histogram shows the distance of 4000 randomly chosen linear models in % to its derived strict lexicographic model, where a distance of 0% means a strict lexicographic ordering and a distance of 100% is the theoretical upper bound among all orderings that however need not be realizable by a linear model as we pointed out earlier. Among these 4000 random instances less than 0.1% have a smaller distance to their derived strictly lexicographic ordering than the linear model computed for the engine-aesthetics study data has to its derived strict lexicographic models.

In similar histograms (Figures 5 to 8) of distances using Spearman’s rank correlation (ρ) metric (instead of the footrule metric) of randomly generated linear models to their derived strict lexicographic models we observe that the engine-aesthetics, engine-details results are even a bit more to the left, i.e., relatively, even closer to a strict lexicographic ordering.

All histograms indicate that the linear models for all four studies indeed have a strong lexicographic bias. We can test this further by comparing the predictive power of the linear models and their corresponding derived lexicographic models on hold out data. The average correct prediction of strict lexicographic models is shown in Table 1 and compared to the predictions of the linear model. Note the difference between the detail and the aesthetics studies: for the two detail studies the strict lexicographic models achieve a correct prediction rate that is only about 1.7% worse than the underlying linear model, and for the two aesthetics studies the lexicographic models achieve a prediction rate that is about 3.9% worse than the underlying linear model. That is, the lexicographic bias is stronger for the detail studies.

6.2 IT skills study

In our second study (IT-skills) we have assessed the importance of five different skill groups for IT professionals (soft skills, domain expertise, package implementation, application development, and architecture). For this study about 70 representatives of IT consulting firms were polled at the CeBit 2009 fair in Hannover. Every participant of the study was shown two skill profiles (each skill group represented by a score between one and four) of two potential job candidates. The participants had to choose the one with the better chances on the IT job market, see Figure 9.

¹ In similar histograms for the engine-detail study ξ is even a bit more to the left, so even closer to a strict lexicographic ordering.

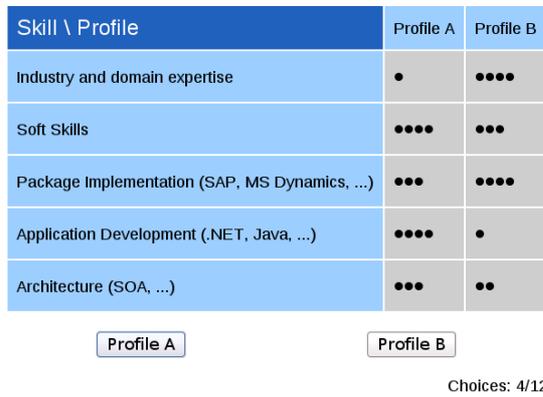


Fig. 9. Web interface for paired comparison of IT skills profiles. The two profiles A and B are presented visually using one to four stars for every skill level. A study participant had to indicate the profile with better job market chances by pressing the corresponding button.

As can be seen in the histograms (Figures 10 to 13) we detected a lack of a strong lexicographic bias (using Spearman’s footrule metric). Our observation is supported by the prediction rate of the derived lexicographic models which is at least 7.9% worse than for the linear model, see also Table 1.

	engine-details	engine-aesthetics	foot-details	foot-aesthetics	it-skills
linear	0.8337(21)	0.7517(7)	0.7085(15)	0.6590(9)	0.7957(41)
lex. 1	0.8177(10)	0.7190(3)	0.6917(10)	0.6105(23)	0.7167(15)
lex. 2	0.8191(6)	0.7191(1)	0.6919(9)	0.6100(23)	0.7170(31)
lex. 3	0.8060(184)	0.6765(245)	0.6800(67)	0.5830(128)	0.6912(144)

Table 1. Average percentage of correct predictions for the four visualization studies and the IT skills study. Shown is the mean for $k = 10$ strata and the estimated standard deviation in brackets. lex. 1 refers to the strict lexicographical model that minimizes Spearman’s footrule distance, lex. 2 refers to the strict lexicographical model that minimizes Spearman’s rho distance and lex. 3 refers to the lexicographical model that minimizes Kendall’s tau metric.

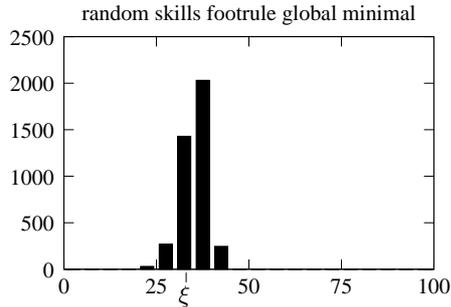


Fig. 10. Distance distribution to the optimal lexicographic model with respect to Spearman's footrule metric.

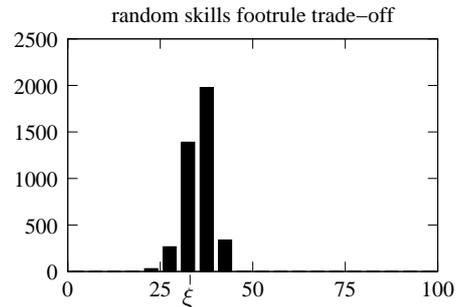


Fig. 11. Distance distribution to lexicographic model computed by the trade-off method.

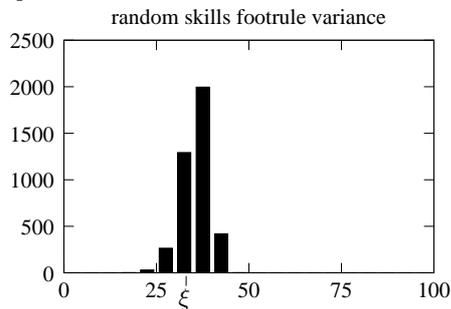


Fig. 12. Distance distribution to lexicographic model computed by the variance method.

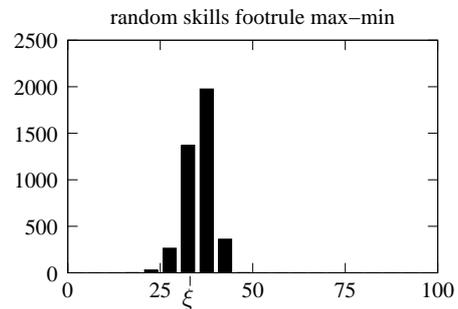


Fig. 13. Distance distribution to lexicographic model computed by the max-min method.

6.3 Discussion

In the visualization studies where we observe a strong lexicographic bias some of the parameters are consequently of lesser importance. One of the lesser parameters, especially for the foot dataset studies, is the screen resolution in pixels of the rendered image. Hence, it does not pay off in terms of perceived quality of rendered image to optimize (increase) resolution. This is a relevant insight since increasing the resolution is costly in terms of processing time and memory.

For the IT skills study we do not observe a strong lexicographic bias. This can be interpreted as none of the five different skills categories should be neglected, i.e., all these categories can be important for professional success or at least offer interesting professional opportunities.

7 Conclusion

We have developed methods to measure a lexicographic bias in linear conjoint analysis models and we also showed how to derive strict lexicographic preference

models from linear conjoint analysis models. There is some resemblance with general dimension reduction techniques, like principal component analysis. In a sense our methods detect “principal parameters”.

We proved the applicability of our methods on data from two user studies. In the data sets from our first study that we had assessed to measure perceived quality of volume rendering algorithms our methods predicted a strong lexicographic bias. This observation was confirmed by the fact that the strict lexicographic models fared not much worse than their underlying linear conjoint analysis model on these data sets. In our second study our methods observed a lack of lexicographic bias (all parameters need to be considered “principal”). This was confirmed by the poor prediction rate of the derived strict lexicographic models compared to the linear conjoint analysis model.

Acknowledgement.

This work has been supported by the DFG (grant GI-711/3-1).

References

1. <http://www.sawtoothsoftware.com>.
2. Persi Diaconis and Ron L. Graham. Spearman’s footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B.*, 39(2):262–268, 1977.
3. Theodoros Evgeniou, Constantinos Boussios, and Giorgos Zacharia. Generalized robust conjoint estimation. *Marketing Science*, 24(3):415–429, 2005.
4. Joachim Giesen, Klaus Mueller, Eva Schubert, Lujin Wang, and Peter Zolliker. Conjoint analysis to measure the perceived quality in volume rendering. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1664–1671, 2007.
5. Gerd Gigerenzer and Reinhard Selten. *Bounded Rationality: The Adaptive Toolbox*. MIT Press, 2002.
6. A. Gustafsson, A. Herrmann, and F. Huber. Conjoint analysis as an instrument of market research practice. In A. Gustafsson, A. Herrmann, and F. Huber, editors, *Conjoint Measurement. Methods and Applications*, pages 5–45. Springer, Berlin, 2000.
7. Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Support vector learning for ordinal regression. In *International Conference on Artificial Neural Networks*, pages 97–102, 1999.
8. T. Joachims. Optimizing search engines using clickthrough data. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142, 2002.
9. Rajeev Kohli and Kamel Jedidi. Representation and inference of lexicographic preference models and their variants. *Marketing Science*, 26(3):380–399, 2007.
10. David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of Measurement Volume I: Additive and Polynomial Representations*. Dover Publications, 2006.
11. Ariel Rubinstein. *Modeling Bounded Rationality*. MIT Press, 2003.
12. Herbert Simon. *Models of Man: Mathematical Essays on Rational Human Behavior in a Social Setting*. Wiley, 1957.
13. Amos Tversky. Elimination by aspects: A theory of choice. *Psychological Review*, 79:281–299, 1972.