Choice Based Conjoint Analysis: Discrete Choice Models vs. Direct Regression*

Bilyana Taneva[1], Joachim Giesen[2], Peter Zolliker[3], and Klaus Mueller[4]

[1]Max-Planck Institut für Informatik, Germany
[2]Friedrich-Schiller Universität Jena, Germany
[3]EMPA Dübendorf, Switzerland
[4]Stony Brook University, USA

Abstract. Conjoint analysis is family of techniques that originated in psychology and later became popular in market research. The main objective of conjoint analysis is to measure an individual's or a population's preferences on a class options that can be described by parameters and their levels. We consider preference data obtained in choice based conjoint analysis studies, where one observes test persons' choices on small subsets of the options. There are many ways to analyze choice based conjoint analysis data. Here we want to compare two approaches, one based on statistical assumptions (discrete choice models) and a direct regression approach. Our comparison on real and synthetic data indicates that the direct regression approach outperforms the discrete choice models.

1 Introduction

Conjoint analysis is a popular family of techniques mostly used in market research to assess consumers' preferences, see [4] for an overview and recent developments. Preferences are assessed on a set of options that are specified by multiple parameters and their levels. In general conjoint analysis comprises two tasks: (a) preference data assessment, and (b) analysis of the assessed data. Common to all conjoint analysis methods is that preferences are estimated from conjoint measurements, i.e., measurements taken on all parameters simultaneously.

Choice based conjoint analysis is a sub-family of conjoint analysis techniques named after the employed data assessment/measurement method, namely a sequence of choice experiments. In a choice experiment a test person is confronted with a small number of options sampled from a parametrized space, and has to choose his preferred option. The measurement is then just the observation of the test person's choice. Choice based conjoint analysis techniques can differ in the analysis stage. Common to all methods is that they aim to compute a scale on the options from the assessed choice data. On an ordinal scale that is a ranking of all the options, but more popular is to compute an interval scale where a numerical value, i.e., a scale value, is assigned to every option. The interpretation is, that an option that gets assigned a larger scale value is more preferred.

* Joachim Giesen and Peter Zolliker were partially supported by the Hasler Stiftung (Proj. #2200).

Differences of scale values have a meaning, but there is no natural zero. That is, an interval scale is invariant under translation and re-scaling by a positive factor.

The purpose of our paper is to compare two analysis approaches on measured and synthetic data. Both approaches compute an interval scale from choice data. The first approach builds on probabilistic modeling and can be seen as an extension of the popular discrete choice methods, see for example [7], to the case of conjoint measurements. The second approach is direct regression introduced by Evgeniou, Boussios and Zacharia [2] that builds on ideas from maximum margin classification aka support vector machines (though applicability of the kernel trick, which mostly contributed to the popularity of support vector machines, seems not so important for conjoint analysis). We also study a geometrically inspired direct regression approach based on computing the largest ball that can be inscribed into a (constraint) polytope. Both approaches, i.e., discrete choice models and direct regression, can be used to compute the scale for either a population of test persons from choice data assessed on the population, or for an individual solely from his choice data. There is also a third option, namely to compute the scale for an individual from his choice data and the population data weighted properly. Here we are not going to discuss this third option.

2 Notation

Formally, the options in the choice experiments are elements in the Cartesian product $A = A_1 \times \ldots \times A_n$ of parameter sets A_i, which in general can be either discrete or continuous—here we assume that they are finite. The choice data are of the form $a \succeq b$, where $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in A$ and a was preferred over b by some test person in some choice experiment. Our goal is to compute an interval scale $v : A \to \mathbb{R}$ on A from a set of choice data.

Often it is assumed that the scale v is linear, i.e., that it can be decomposed as

$$v(a) = v\big((a_1, \ldots, a_n)\big) = \sum_{i=1}^{n} v_i(a_i),$$

where $v_i : A_i \to \mathbb{R}$. In the case of continuous parameters A_i the linearity of the scale is justified when the parameters are preferentially independent, for details see [5]. For finite parameter sets linearity still implies preferential independence, but the reverse is in general not true anymore. Nevertheless, in practice linearity is almost always assumed. Also the two methods that we are going to discuss here both assume linearity of the scale[1]. The discrete choice models approach first estimates the scales v_i from the choice data individually first and then combines them in a second step. Note that the choice data are obtained from conjoint measurements, i.e., choices among options in A and not in A_i. The direct regression (maximum margin or largest inscribed ball) approach estimates the

[1] The linearity assumption can be mitigated by combining dependent parameters into a single one, see [3] for a practical example.

scales v_i simultaneously from the choice data. Note that both approaches have to estimate the same number of parameters, namely all the values $v_i(a), a \in A_i, i = 1, \dots, n$.

3 Discrete Choice Models

Discrete choice models deal with the special case of a single parameter, i.e., in a sense the non-conjoint case. Let the finite set A denote this parameter set. Choice data are now of the form $a \succeq b$ with $a, b \in A$ and the goal is to compute $v : A \to \mathbb{R}$ or equivalently $\{v_a = v(a) \,|\, a \in A\}$. Discrete choice models make the assumption that the observed choices are outcomes of random trials: confronted with the two options $a, b \in A$ a test person assigns values $u_a = v_a + \epsilon_a$ and $u_b = v_b + \epsilon_b$, respectively, to the options, where (the error terms) ϵ_a and ϵ_b are drawn independently from the same distribution, and chooses the option with larger value. Hence the probability p_{ab} that a is chosen over b is given as

$$p_{ab} = Pr[u_a \geq u_b] = Pr[v_a + \epsilon_a \geq v_b + \epsilon_b] = Pr[v_a - v_b \geq \epsilon_b - \epsilon_a]$$

Discrete choice models can essentially be distinguished by the choice of distribution for the ϵ_a. Popular choices are normal distributions (Thurstone's (probit) model [6]) or extreme value distributions (Bradley-Terry's (logit) model [1]), see also [7]. The values v_a can be computed for both models either via the difference $v_a - v_b$ from the probability p_{ab} which can be estimated by the frequency f_{ab} that a was preferred over b in the choice experiments, or computationally more involved by a maximum likelihood estimator. Here we introduce a least squares approach using the frequency estimates for the p_{ab}.

3.1 Thurstone's Model (probit)

In Thurstone's model [6] the error terms ϵ_a are drawn from a normal distribution $N(0, \sigma^2)$ with expectation 0 and variance σ^2. Hence the difference $\epsilon_b - \epsilon_a$ is also normally distributed with expectation 0 and variance $2\sigma^2$ and hence

$$\begin{aligned} p_{ab} &= Pr[u_a \geq u_b] = Pr[\epsilon_b - \epsilon_a \leq v_a - v_b] \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \int_{-\infty}^{v_a - v_b} e^{-\frac{x^2}{4\sigma^2}} dx = \Phi\left(\frac{v_a - v_b}{\sqrt{2}\sigma}\right), \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

This is equivalent to

$$v_a - v_b = \sqrt{2}\sigma\Phi^{-1}(p_{ab}).$$

Using the frequency f_{ab} that a was preferred over b (number of times a was preferred over b divided by the number that a and b have been compared) we set

$$v_{ab} = \sqrt{2}\sigma\Phi^{-1}(f_{ab}).$$

3.2 Bradley-Terry's Model (logit)

In Bradley-Terry's model [1] the error terms ϵ_a are drawn from a standard Gumbel distribution, i.e., the distribution with location parameter $\mu = 0$ and scale parameter $\beta = 1$. Since the difference of two independent Gumbel distributed random variables is logistically distributed we have

$$\begin{aligned} p_{ab} &= Pr[u_a \geq u_b] = Pr[\epsilon_b - \epsilon_a \leq v_a - v_b] \\ &= \frac{1}{1 + e^{-(v_a - v_b)}} = \frac{e^{v_a - v_b}}{1 + e^{v_a - v_b}} = \frac{e^{v_a}}{e^{v_a} + e^{v_b}}. \end{aligned}$$

This implies

$$\frac{e^{v_a}}{e^{v_b}} = \frac{p_{ab}}{1 - p_{ab}},$$

which is equivalent to

$$v_a - v_b = \ln\left(\frac{p_{ab}}{1 - p_{ab}}\right).$$

Analogously to what we did for Thurstone's model we set

$$v_{ab} = \ln\left(\frac{f_{ab}}{1 - f_{ab}}\right).$$

3.3 Computing Scale Values

From both Thurstone's and Bradley-Terry's model we get an estimate v_{ab} for the difference of the scale values v_a and v_b. Our goal is to estimate the v_a's (and not only their differences). This can be done by computing v_a's that best approximate the v_{ab}'s (all equally weighted) in a least squares sense. That is, we want to minimize the residual

$$r(v_a | a \in A) = \sum_{a,b \in A; b \neq a}^{n} (v_a - v_b - v_{ab})^2.$$

A necessary condition for the minimum of the residual is that all partial derivatives vanish, which gives

$$\frac{\partial r}{\partial v_a} = 2 \sum_{b \in A; b \neq a} (v_a - v_b - v_{ab}) = 0.$$

Hence

$$|A| v_a = \sum_{b \in A} v_b + \sum_{b \in A; b \neq a} v_{ab}.$$

Since we aim for an interval scale we can assume that $\sum_{b \in A} v_b = 0$. Then the values that minimize the residual are given as

$$v_a = \frac{1}{|A|} \sum_{b \in A; b \neq a} v_{ab}.$$

We can specialize this now to the discrete choice models and get for Thurstone's model

$$v_a = \frac{\sqrt{2}\sigma}{|A|} \sum_{b\in A; b\neq a} \Phi^{-1}(f_{ab}),$$

and for Bradley-Terry's model

$$v_a = \frac{1}{|A|} \sum_{b\in A; b\neq a} \ln\left(\frac{f_{ab}}{1-f_{ab}}\right).$$

3.4 Multi-parameter (conjoint) case

Now we turn to the multi-parameter case where the options are elements in $A = A_1 \times \ldots \times A_n$. We assume a linear model and describe a compositional approach to compute the scales for the parameters A_i. In a first step we compute scales v_i using a discrete choice model for the one parameter case, and then in a second step compute re-scale values w_i to make the scales v_i comparable. Our final scale for A is then given as $v = \sum_{i=1}^n w_i v_i$, i.e.,

$$v\big((a_1,\ldots,a_n)\big) = \sum_{i=1}^n w_i v_i(a_i).$$

To compute the scales v_i we make one further assumption: if $a = (a_1,\ldots,a_n) \in A$ is preferred over $b = (b_1,\ldots,b_n) \in A$ in a choice experiment we interpret this as a_i is preferred over b_i to compute the frequencies $f_{a_i b_i}$. If the parameter levels in the choice experiments are all chosen independently at random, then the frequencies $f_{a_i b_i}$ should converge (in the limit of infinitely many choice experiments) to the frequencies that one obtains in experiments involving only a single parameter A_i.

To compute the re-scale values w_i we use a maximum margin approach (essentially the same approach that was introduced by Evgeniou et. al to compute a scale $v : A \to \mathbb{R}$ by direct regression). The approach makes the usual trade-off between controlling the model complexity (maximizing the margin) and accuracy of the model (penalizing outliers). The trade-off is controlled by a parameter $c > 0$ and we assume that we have data from m choice experiments available.

$$\begin{aligned}
&\min_{w_i, z_j} \ \sum_{i=1}^n w_i^2 + c\sum_{j=1}^m z_j \\
&\text{s.t.} \quad \sum_{i=1}^n w_i\big(v_i(a_i) - v_i(b_i)\big) + z_j \geq 1, \\
&\qquad\quad \text{if } (a_1,\ldots,a_n) \succeq (b_1,\ldots,b_n) \text{ in the } j\text{'th choice experiment.} \\
&\qquad\quad z_j \geq 0, j = 1,\ldots,m
\end{aligned}$$

4 Direct Regression

The regression approach that we described to compute the re-scale values for discrete choice models can be also applied directly to compute scale values. We start our discussion again with the special case of a single parameter, where we have to estimate $v_a = v(a)$ for all $a \in A$.

4.1 Single Parameter Case

The naive approach to direct regression would be to compute scale values $v_a \in \mathbb{R}$ that satisfy constraints of the form $v_a - v_b \geq 0$ if $a \in A$ was preferred over $b \in A$ in a choice experiment. The geometric interpretation of this approach is to pick a point in the constraint polytope, i.e., the subset of $[-l, l]^{|A|}$ for sufficiently large l that satisfies all constraints. There are many such points that all encode a ranking of the options in A that complies with the constraints. Since we have only combinatorial information, namely choices, there is no way to distinguish among the points in the constraint polytope—except we have contradictory information, i.e., choices of the form $a \succeq b$ and $b \succeq a$ which render the constraint polytope empty. We will have contradictory information, especially when we assess preferences on a population, but also individuals can be inconsistent in their choices. It is essentially the contradictory information that makes the problem interesting and justifies the computation of an interval scale instead of an ordinal scale (i.e., a ranking or an enumeration of all partial rankings compliant with the choices) from choice information. The choice information now can no longer be considered purely combinatorial since also the frequency of $a \succeq b$ for all comparisons of a and b will be important. To avoid an empty constraint polytope we introduce a non-negative slack variable z_j for every choice, i.e., $v_a - v_b + z_j \geq 0, z_j \geq 0$ if a was preferred over b in the j'th choice experiment. Now the constraint polytope will always be non-empty and it is natural to aim for minimal total slack, i.e., $\sum_{j=1}^{k} z_j$ if we have information from m choice experiments. But since v_a = constant for all $a \in A$ is always feasible we get the optimal solution

$$v_a = \text{ constant, and } \sum_{j=1}^{m} z_j = 0.$$

To mitigate this problem we demand that the constraints $v_a - v_b + z_j \geq 0$ if $a \succeq b$ in the j'th choice experiment a should be satisfied with some confidence margin, i.e., the constraints get strengthened to $v_a - v_b + z_j \geq 1$. Finally, as we did when computing re-scale values for discrete choice models we control the model complexity by maximizing the margin. That is, we end up with the following optimization problem for direct regression of scale values:

$$\begin{array}{ll} \min_{v_a, z_j} \sum_{a \in A} v_a^2 + c \sum_{j=1}^{m} z_j & \\ \\ \text{s.t.} & v_a - v_b + z_j \geq 1, \\ & \quad \text{if } a \succeq b \text{ in the } j\text{'th choice experiment.} \\ & z_j \geq 0, j = 1, \ldots, m \end{array}$$

4.2 Conjoint Case

Now we assume again $A = A_1 \times \ldots \times A_n$. Of course we could proceed as for the discrete choice models and re-scale scales computed with direct regression for the different parameters A_i, but we can also use direct regression to compute all scale values simultaneously. With a similar reasoning as in for the single parameter case we obtain the following optimization problem:

$$\begin{aligned}
&\min_{v_i(a),z_j} \textstyle\sum_{i=1}^n \sum_{a\in A_i} v_i(a)^2 + c\sum_{j=1}^m z_j \\
&\text{s.t.} \qquad \textstyle\sum_{i=1}^n v_i(a_i) - v_i(b_i) + z_j \geq 1, \\
&\qquad\qquad \text{if } (a_1,\ldots,a_n) \succeq (b_1,\ldots,b_n) \text{ in the } j\text{'th choice experiment.} \\
&\qquad\qquad z_j \geq 0, j = 1,\ldots,m
\end{aligned}$$

4.3 Largest Inscribed Ball

For the conjoint case we also study a geometrically inspired direct regression approach, namely computing the largest ball inscribed into the polytope defined by the constraints

$$\sum_{i=1}^n v_i(a_i) - v_i(b_i) \geq 0, \text{ if } (a_1,\ldots,a_n) \succeq (b_1,\ldots,b_n) \text{ in a choice experiment.}$$

We want to estimate the $v_i(a)$ for $i = 1,\ldots,n$ and all $a \in A_i$. That is, we want to estimate the entries of a vector v with $m = \sum_{i=1}^n m_i$ components, where $m_i = \|A_i\|$. A choice experiment is defined by the characteristic vectors $\chi_a \in \{0,1\}^m$, whose i'th component is 1 if the corresponding parameter level is present in the option a, and 0 otherwise. The constraint polytope can now be re-written as

$$v^t(\chi_a - \chi_b) \geq 0, \text{ if } a \succeq b \text{ in a choice experiment,}$$

or equivalently $v^t x_{ab} \geq 0$, where $x_{ab} = (\chi_a - \chi_b)/\|\chi_a - \chi_b\|$.

The distance of a point $v \in \mathbb{R}^m$ to the hyperplane (subspace) $\{v \in \mathbb{R}^m \mid v^t x_{ab} = 0\}$ is given by $v^t x_{ab}$. The largest inscribed ball problem now becomes when using the standard trade-off between model complexity and quality of fit on the observed data

$$\begin{aligned}
&\max_{v,r,z}\ r + c\textstyle\sum_{j=1}^k z_j \\
&\text{s.t.} \qquad v^t x_{ab} \geq r - z_k, \text{ if } a \succeq b \text{ in the } j\text{'th choice experiment.} \\
&\qquad\qquad z_j \geq 0, j = 1,\ldots,k
\end{aligned}$$

where r is the radius of the ball and c is the trade-off parameter. This is a linear program, in contrast to the direct regression approach based on maximizing the margin which results in a convex quadratic program.

The largest inscribed ball approach does not work directly. To see this observe that the line $v_1 = v_2 = \ldots = v_m =$ constant is always feasible. If the feasible region contains only this line (which often is the case), then the optimal solution of our problem would be on this line. A solution $v_1 = v_2 = \ldots = v_m =$ constant however does not give us meaningful scale values. To make deviations from the line $v_i =$ constant possible we add a small constant $\epsilon > 0$ to the left hand side of all the comparison constraints. In our experiments we chose $\epsilon = 0.1$.

4.4 Cross Validation

The direct regression approaches (and also our re-scaling approach) have the free parameter $c > 0$ that controls the trade-off between model complexity and model accuracy. The standard way to choose this parameter is via k-fold cross validation. For k-fold cross-validation the set of choice data is partitioned into k partitions (aka strata) of equal size. Then $k - 1$ of the strata are used to compute the scale values, which can be validated on the left out stratum. For the validation we use the scale value to predict outcome in the choice experiments in the left-out stratum. Given $v(a)$ and $v(b)$ for $a, b \in A$ such that a and b have been compared in the left-out stratum, to predict the outcome one can either predict the option with the higher scale value, or one can make a randomized prediction, e.g., by using the Bradley-Terry model: predict a with probability $e^{v(a)}/\big(e^{v(a)} + e^{v(b)}\big)$. The validation score can then either be the percentage of correct predictions or the average success probability for the predictions. For simplicity we decided to use the percentage of correct predictions.

5 Data Sets

We compared the different approaches to analyze choice based conjoint analysis data on two different types of data sets: (a) data that we assessed in a larger user study to measure the perceived quality for a visualization task [3], and (b) synthetic data that we generated from a statistical model of test persons. So let us describe the visualization study first.

5.1 Visualization Study

The purpose of volume visualization is to turn 3D volume data into images that allow a user to gain as much insight into the data as possible. A prominent example of a volume visualization application is MRI (magnetic resonance imaging). Turning volume data into images is a highly parametrized process among the many parameters there are for example

(1) The choice of color scheme: often there is no natural color scheme for the data, but even when it exists it need not best suited to provide insight.
(2) The viewpoint: an image is a 2D projection of the 3D data, but not all such projections are equally valuable in providing insights.
(3) Other parameters like image resolution or shading schemes.

In our study [3] we were considering six parameters (with two to six 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). We had 317 test persons for the two details question studies and 366 test persons for the aesthetics studies, respectively. In each study the test persons were shown two images from the same category, i.e., either foot or engine, rendered with different parameter settings and asked which of the two images they prefer (with respect to either the aesthetics or the details question). Hence in each choice experiment there were only two options, see Figures 1 and 2 for examples.

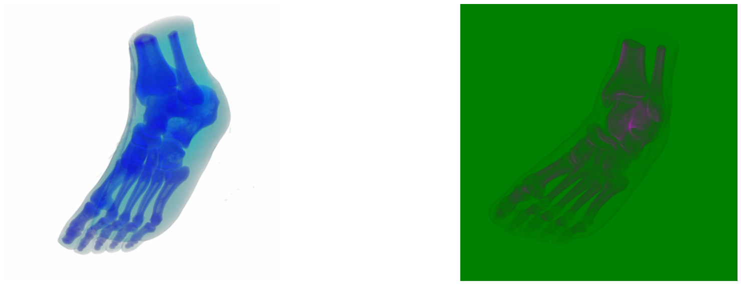


Fig. 1. Data set foot: Which rendering do you like (left or right)?

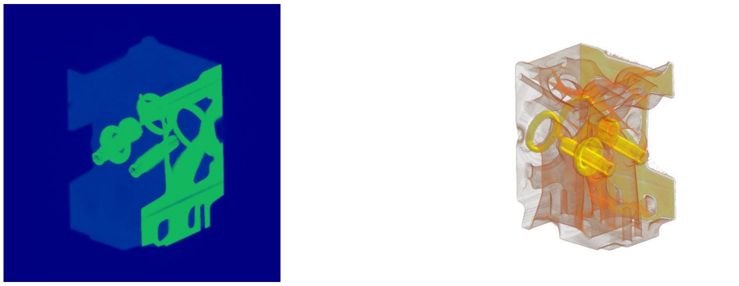


Fig. 2. Data set engine: Which rendering shows more detail (left or right)?

In [3] we evaluated the choice data using the Thurstone discrete choice model for the whole population of test persons. There we used a different method to re-scale the values from the first stage than described here. The method we used is based on the normal distribution assumption and thus not as general as the method described here.

5.2 Synthetic Data

We were also interested to see how well the different methods perform when we only use information provided by a single person. Unfortunately the information provided individually by the test persons in the visualization studies is very sparse (only 20 comparisons per person). Thus we also generated synthetic data as follows:

(1) We simulated a study with five parameters and five levels each.

(2) We generated 200 synthetic test persons represented by a scale value for every parameter level. The scale values for the levels of each parameter were chosen from normal distributions with mean $-2, -1, 0, 1$ and 2, respectively. The normal distributions always had the same standard deviation, which we choose to be $2, 5$ or 8 (for three different studies). Varying the standard deviation was used to model different degrees of heterogeneity in the population.

(3) The synthetic test persons provided answers to 200 binary choice problems following the Bradley-Terry model. That is, given two options a and b and test person dependent scale values $v(a)$ and $v(b)$, respectively, the test person prefers a over b with probability $p_{ab} = e^{v(a)}/\big(e^{v(a)} + e^{v(b)}\big)$. To simulate the choices we generated random numbers uniformly in $[0, 1]$ and compared them to the p_{ab}'s.

6 Results and Discussion

As pointed in Section 3 when assigning scale values to parameter levels in the discrete choice approach, we estimate the probability that level a is preferred over level b by the relative frequency f_{ab} that a was preferred over b. For sparse data (only very few comparisons per test person) the frequency matrix can also be sparse even in the sense of missing entries, i.e., levels a and b that never have been compared. To deal with sparseness we exploit a "transitivity of preferences" assumption. Whenever $a \succ b$ and $b \succ c$ we interpret this also as a (weak) vote for $a \succ c$. We implemented this idea as follows: we initialized f_{ab} with either the measured relative frequency, or when this is not available with $1/2$. Then we updated f_{ab} iteratively until convergence using the following formula (with some constant $c \in (0, 1)$, we obtained good results for $c = 0.3$):

$$f_{ab} = (1-c)f_{ab} + \frac{c}{n-2}\sum_{d \neq a,b} \frac{f_{ad}f_{db}}{f_{ad}f_{db} + f_{da}f_{bd}}$$

That is, we smoothed the frequencies using all the information available.

6.1 Visualization Studies

Let us start with a summary of the performance of our analysis methods on the data of the four visualization studies. The summary is given in Table 1. We consider the Thurstone and Bradley-Terry discrete choice models, the regression approach either compositional, i.e., with two stages as for the discrete choice models, or direct, and the largest inscribed ball regression approach. Here we report k-fold cross validation values, i.e., the percentage of correct predictions (on the left out strata). We were using 20 random partitions into $k = 10$ strata (also for everything that follows) and report the mean and estimated standard deviation of the percentage of correct predictions (i.e., for every correct prediction percentage that we report we had 200 data points).

	engine-aesthetics	engine-detail	foot-aesthetics	foot-detail
Thurstone	0.7535(8)	0.8265(8)	0.6640(10)	0.7388(5)
Bradley-Terry	0.7536(9)	0.8267(5)	0.6640(1)	0.7387(10)
Compositional regression	0.7397(10)	0.8280(10)	0.6539(20)	0.7069(20)
Direct regression	0.7529(9)	0.8401(20)	0.6638(10)	0.7411(10)
Largest ball	0.7530(9)	0.8414(7)	0.6638(16)	0.7405(10)

Table 1. Average percentage of correct predictions for the four visualization studies. Shown is the mean for $k = 10$ strata and the estimated standard deviation in brackets. See also Figure 3.

For the data that we report in Table 2 we consider only data provided by a test person to compute personal scale values for this person. The presented results are the mean percentage of correct predictions on the left out strata also averaged over all test persons that participated in the study. The standard deviation is computed with respect to the left out strata and the different test persons.

	engine-aesthetics	engine-detail	foot-aesthetics	foot-detail
Thurstone	0.636(2)	0.670(2)	0.597(2)	0.596(3)
Bradley-Terry	0.640(3)	0.675(2)	0.598(3)	0.599(3)
Compositional regression	0.636(3)	0.688(2)	0.598(2)	0.601(3)
Direct regression	0.618(3)	0.651(2)	0.585(2)	0.589(3)
Largest ball	0.616(5)	0.631(4)	0.578(3)	0.580(4)

Table 2. Average percentage of correct predictions for the four visualization studies for individual test persons. For the prediction only data provided by the individual test person were used. See also Figure 3.

While the direct regression approaches performed slightly but not statistically significant better than the compositional approaches (including the discrete choice approaches) in the non-personalized analysis they perform significantly worse if one uses the approach for personalized prediction on the same data sets. That is one reason why we also generated synthetic data. We wanted to study the behavior of the approaches when we have more data per test person available. Note that in the visualization studies we were restricted to only very few choice experiments per test person since these test persons volunteered to take part in the studies during an exhibition at the computer science department of ETH Zürich. Our conjecture was that the direct regression approach outperforms the other approaches once we have more data per test person.

Another interesting finding regarding the visualization studies is that we were not able to detect a statistically significant difference between Thurstone's model and Bradley-Terry's model. We expected that Bradley-Terry's model to perform slightly better, because of the more realistic fat tail assumption of the underlying distribution. Actually, it performs slightly better in the personalized analysis, but the advantage is not statistically relevant.

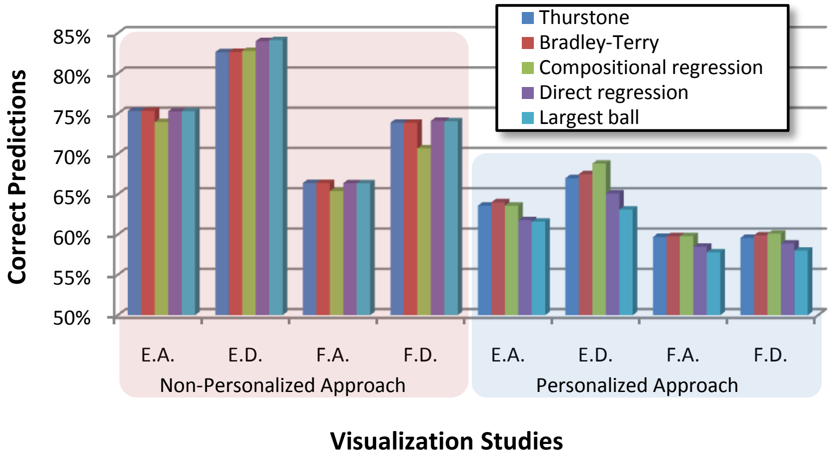


Fig. 3. Summarizing Tables 1 and 2.

Finally, in Table 3 we report on the dependence of the direct regression approach on the regularization parameter c that controls the trade-off between model complexity and training error. Here we report only on a non-personalized analysis, since the behavior in the other settings is similar.

Interestingly, the direct regression approach only marginally dependent on the choice of the regularization parameter c. All the results that we report here are at least in the range of the other approaches.

6.2 Synthetic Data

As we have mentioned earlier our main motivation to generate synthetic data was to study the effect of the number of choice experiments per test person on the performance of the different approaches.

	engine-aesthetics	engine-detail	foot-aesthetics	foot-detail
100	0.7525(6)	0.8401(20)	0.6635(10)	0.7402(10)
10	0.7529(10)	0.8396(20)	0.6636(10)	0.7401(10)
1	0.7529(10)	0.8341(20)	0.6638(10)	0.7411(10)
0.01	0.7405(10)	0.8313(10)	0.6585(10)	0.7167(10)

Table 3. Average percentage of correct predictions for the four visualization studies analyzed with direct regression using the values $c = 100, 10, 1$ and 0.01 for the trade-off parameter.

To check the validity of our model from which we generated the artificial data we first show in Table 4 the correct prediction percentages for the direct regression and the Bradley-Terry discrete choice model for varying standard deviation (which is one of the parameters we can control when generating the synthetic data).

Standard Deviation	Direct reg.	pers.	Bradley-Terry	pers.
2	0.716(1)	0.765(3)	0.723(1)	0.749(4)
5	0.606(1)	0.783(3)	0.619(1)	0.745(2)
8	0.574(1)	0.784(3)	0.565(2)	0.749(4)

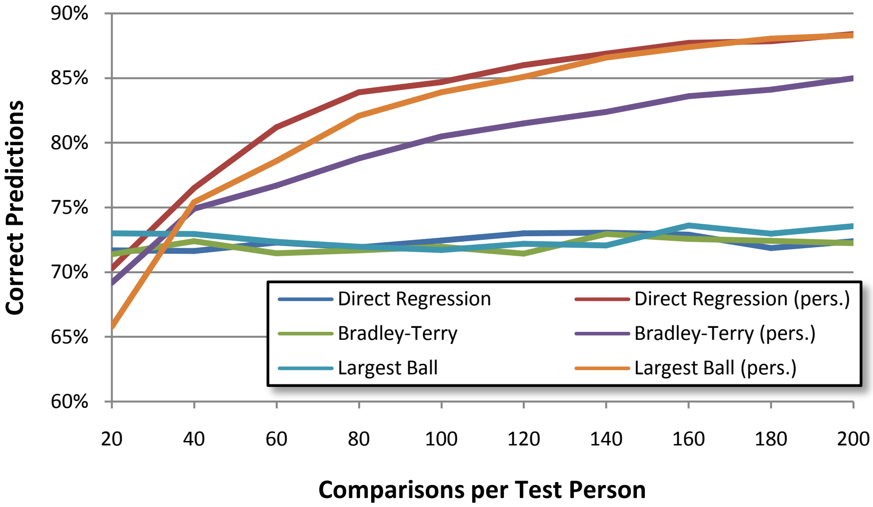
Table 4. Comparison of the average percentages of correct predictions on the artificial data with 40 choice experiments per person. Shown are prediction percentages for the direct regression and the Bradley-Terry discrete choice model approaches (non-personalized and personalized).

As expected it becomes more difficult to predict the outcome of a choice experiment on the population level when we increase the standard deviation (which is meant to model population heterogeneity), whereas in the personalized setting (where we consider only data provided by a test person to compute personal scale values for this test person) the prediction accuracy does not depend on the variance. The heterogeneity is actually large enough that it pays off (in terms of prediction accuracy) to personalize even for standard deviation 2.

In Table 5 we summarize the dependence on the number of choice experiments (comparisons) for the different approaches in the personalized setting. The results for the personalized setting show that—as we expected—the percentage of correct predictions hardly improves with growing number of choice experiments per test person. At the same time—which is also expected—the variance of correct prediction percentages goes down. But we do not only observe that the percentage of correct prediction increases with growing number of choice experiments per test persons, but also that the direct regression approaches (which includes the largest inscribed ball approach) outperform the Bradley-

# Comparisons	Direct reg.	non-pers.	Bradley-Terry	non-pers.	Largest ball	non-pers.
20	0.703(4)	0.717(2)	0.692(4)	0.714(2)	0.658(3)	0.730(1)
40	0.765(3)	0.7164(8)	0.749(4)	0.723(1)	0.754(5)	0.7296(8)
60	0.812(2)	0.7231(6)	0.767(2)	0.7147(5)	0.786(3)	0.7234(4)
80	0.839(2)	0.7192(5)	0.788(1)	0.7169(5)	0.821(4)	0.7197(4)
100	0.847(2)	0.7246(4)	0.805(2)	0.7198(5)	0.839(2)	0.7173(3)
120	0.860(1)	0.7302(3)	0.815(1)	0.7143(5)	0.851(1)	0.7220(3)
140	0.869(1)	0.7306(2)	0.824(1)	0.7295(3)	0.866(1)	0.7207(3)
160	0.8773(9)	0.7291(3)	0.836(1)	0.7259(4)	0.8741(7)	0.7361(3)
180	0.8785(8)	0.7188(3)	0.841(1)	0.7242(3)	0.8806(7)	0.7298(3)
200	0.8841(8)	0.7239(2)	0.850(1)	0.7226(2)	0.8832(4)	0.7357(2)

Table 5. Average percentage of correct predictions for the synthetic data set using 20 to 200 choice experiments per (artificial) test person. Shown are results for the direct regression (maximum margin), Bradley-Terry, and maximum inscribed ball regression approach. See also the figure below.



Terry model (which essentially behaves the same as the Thurstone model). Thus here we observe statistically significant what we already conjectured for the visualization studies, namely, that direct regression outperforms all other methods once enough data are available. Actually, the results show that the amount of data need not be very large before direct regression outperforms the other approaches.

Let us also briefly comment on the results in the non-personalized setting, where the percentage of correct predictions hardly improves with increasing number of comparisons per person. This means that if our goal is to compute scale values for a population of respondents, it can be enough for the test persons to participate in few choice experiments, e.g., 20 for our model (but probably more for conjoint studies with more parameters). This is good news for studies in which the respondents cannot (or are not willing to) participate in many choice experiments.

7 Conclusion

We compared two discrete choice approaches, namely Thurstone's model and Bradley-Terry's model, with a direct regression approach for choice based conjoint data analysis. We also introduced a new direct regression approach based on inscribing the largest ball into a constraint polytope. At least our personalized results on a synthetic data set suggest that both direct regression approaches outperform the discrete choice models—provided there are enough data per test person available.

Our main interest is in the use of conjoint analysis techniques to measure users' preferences for visualization and imaging algorithms. In the conjoint studies that we perform to this end we typically only get test persons to participate in a small number of choice experiments (about 20 choice experiments per test person — which takes roughly three minutes). In this range of numbers of choice experiments discrete choice models even seem to have a small advantage over direct regression (at least in non-personalized analysis). In the future we plan to conduct more user studies to figure out the best analysis approach for varying numbers of choice experiments and objectives (e.g., non-personalized vs. personalized).

References

1. R.A. Bradley and M.E. Terry. Rank analysis of incomplete block designs, i. the method of paired comparisons. *Biometrika*, 39:324–345, 1952.
2. Theodorus Evgeniou, Constantinos Boussios, and Giorgos Zacharia. Generalized robust conjoint estimation. *Marketing Science*, 24(3):415–429, 2005.
3. Joachim Giesen, Klaus Mueller, Eva Schuberth, 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.
4. 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.
5. R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs.* Cambridge University Press, Cambridge, 1993.
6. L. L. Thurstone. A law of comparative judgement. *Psychological Review*, 34:273–286, 1927.
7. K. Train. *Discrete Choice Methods with Simulation.* Cambridge University Press, 2003.