Developments in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 19, Ljubljana: FDV, 2003

An IRT-Approach for Conjoint Analysis

Gerhard Neubauer[1]

Abstract

Conjoint Analysis is a popular method in marketing research that is mainly used for product development. In a quasi-experimental setting product features (attributes) like brand, price or packaging are varied and the resulting synthetical products are evaluated by a sample of subjects. The major aim is to estimate (i) the attractivity of the products, (ii) the impact of the features for the product attractivity, and (iii) the market share. Estimating the attractivity is equivalent to scaling the products, and therefore we refer to (i) as the Scaling Problem and to (ii) as the Impact Problem. In most cases ANOVA-type of analyses are performed on both an individual and an aggregate level. The major disadvantages of this approach are, that the dependencies between the responses are not considered, and further that homogeneity of the subjects is assumed. For binary data from pairwise comparisons we propose to make use of a class of models from Item-Response-Theory (IRT), namely the Rasch Models. The Rasch Model (RM) was developed for the analysis of data from mental tests. It is a model for multivariate data that allows to model subject heterogeneity by the specification of corresponding parameters. The Linear Logistic Test Model (LLTM) allows to restrict the parameters of the RM, and here it is used for estimating the Scaling and the Impact parameters. The approach is applied to data from research on the objective determinants of subjective perception of car engine noise. The data analyses show that conventional Conjoint Analysis and the IRT-approach differ substantially in the estimates for product attractivity and market share. Modelling with and without subject variability makes a difference. The estimates from the models containing subject parameters are assumedly more reliable and consequently make a better basis for marketing decisions.

1 Introduction

Conjoint Analysis is a widely used methodology in marketing research that mainly addresses the following problems:

1. The Scaling Problem - measuring the attractivity of products.

[1] Institute of Applied Statistics, Joanneum Research Forschungsgesellschaft mbH, Graz, Austria; gerhard.neubauer@joanneum.at

2. The Impact Problem - determining the impact of product attributes (features like price, brand, package, color etc.) on the attractivity.

3. The Share Problem - estimating market shares.

The data for Conjoint Analyses stem from evaluations of products - or in a more general setting - evaluations of objects. Depending on the measurement procedure binary, ranking, rating or even metric data are collected. For metric data scaling is simply done by data collection, whereas for the other kinds of data a separate analysis has to be performed. In all cases some kind of a regression model is used for the Impact Problem and the estimated coefficients are called "part-worths". Thus a successfull Conjoint Analysis ends up with understanding which attributes drive the consumers' choices from a set of products.

The paired comparison procedure is a classical method for product evaluation, where binary data are collected when subjects choose one of two presented products. It provides meaningful data when only a small number of products is investigated. In order to handle situations with a moderate number of products the classical procedure has been modified in many ways (e.g. Choice Based Conjoint Analysis) leading to complex measurement situations. Furthermore often subjects nowadays have to express how likely it is that they would choose one of a set of products. This is certainly not a realistic shopping situation - products are either bought or not and thus binary data should be collected for product evaluation.

Usually data in Conjoint Analysis are analysed both on individual and aggregate level, i.e. all responses from one subject are first used to estimate individual coefficients and then generalization is either done by averaging the individual results or estimating a second model for all data. This procedure has two obvious disadvantages: (i) making multiple use of the data for modelling, (ii) modelling multivariate data as independent observations. The estimation on individual level indicates that there is some interest in subject variability, although the treatment by separate models is not appropriate. Only recently some effort has been made to model subject variability in the context of GLMs, generalized mixed linear models (GLMM) and hierarchical Bayes modelling (Dittrich, Hatzinger & Katzenbeisser, 1998; Frühwirth-Schnatter & Otter, 1999; Green, 2000).

In this paper a well-known family of models from Item-Response-Theory (IRT), namely the Rasch Models is used to meet the requirements of multivariate analysis in the presence of subject variability. The focus is on paired comparisons with binary reactions from the subjects. The Rasch Model (RM) and Linear Logistic Test Model (LLTM) are proposed for the Scaling, the Impact and the Share Problem. The theory is outlined and an application to data from investigations on car engine noise is presented to illustrate the use of the approach.

2 IRT Models for data from paired comparisons

2.1 Paired Comparisons

To arrive at a more formal presentation let me introduce the following terminology:

Table 1: Preference matrix for subject S_i.

					O_h				
		1	2	$\cdots$	h	$\cdots$	$m-1$	m	
	1	$-$	Y_{i12}	$\cdots$	Y_{i1h}	$\cdots$	$Y_{i1,m-1}$	Y_{i1m}	$Y_{i1.}$
	2	Y_{i21}	$-$	$\cdots$	Y_{i2h}	$\cdots$	$Y_{i2,m-1}$	Y_{i2m}	$Y_{i2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
O_g	g	Y_{ig1}	Y_{ig2}	$\cdots$	Y_{igh}	$\cdots$	$Y_{ig,m-1}$	Y_{igm}	$Y_{ig.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$m-1$	$Y_{i,m-1,1}$	$Y_{i,m-1,2}$	$\cdots$	$Y_{i,m-1,h}$	$\cdots$	$-$	$Y_{i,m-1,m}$	$Y_{i,m-1,.}$
	m	Y_{im1}	Y_{im2}	$\cdots$	Y_{imh}	$\cdots$	$Y_{im,m-1}$	$-$	$Y_{im.}$
		$Y_{i.1}$	$Y_{i.2}$	$\cdots$	$Y_{i.h}$	$\cdots$	$Y_{i.,m-1}$	$Y_{i.m}$	

1. F_s, $s = 1, \ldots, t$ are factors like brand or color with C_s nominal categories.

2. O_j, $j = 1, \ldots, m$ are objects, where $m = \prod_s C_s$ for full factorial designs.

3. (O_g, O_h) is a subset of $(O_1, \ldots, O_m)$, with $g > h$, and consequently $k = \binom{m}{2}$, the number of possible subsets, i.e. pairs in the paired comparison procedure.

4. S_i, $i = 1, \ldots, n$ is a sample of subjects.

5. Y_{igh}, is the evaluation of (O_g, O_h) by S_i.

6. $X_j = (X_{j1}, \ldots, X_{jp})$, $r = 1, \ldots, p$, is the attribute vector of O_j, consisting of metric and/or binary information.

In marketing research the F_s are used to generate the O_j. Many F_s and/or many C_s per F_s lead to large m, and huge k. As a consequence paired comparison and full factorial designs become completely impractical. As a remedy fractional designs are applied often in combination with response surface methods. For other applications there are no F_s and the O_j are characterized by a set of metric variables X_j. An example is the investigation of car engine noise, where the sound recordings of a set of cars are the O_j, and the X_j are sound characteristics expressed as functions of the frequency spectrum.

In the paired comparison procedure all possible k pairs (O_g, O_h) from a set of m objects $(O_j, j = 1, \ldots, m)$ are presented to n subjects. Each person makes a choice with respect to his/her preference. The data can be coded into a binary variable $Y_{gh} = 1$ if O_g is preferred $(O_g > O_h)$, and $Y_{gh} = 0$ if O_h is preferred $(O_h > O_g)$. Of course the reverse coding yields the redundant variable $Y_{hg} = 1 - Y_{gh}$. Table 1 gives the preference matrix for one subject S_i, which is the redundant representation of the data.

The lower and the upper triangle have the same information and parameter estimation is based on one of them only. The row and column sums are related by $Y_{ig.} = (m-1) - Y_{i.h}$.

2.2 The Bradley-Terry-Luce Model - BTLM

To model Y_{igh} let

$$Y_{igh} = \left\{ \begin{array}{ll} 1 & \text{if object } O_g \text{ is preferred, and} \\ 0 & \text{if object } O_h \text{ is preferred} \end{array} \right\} \sim Bin(1, p_{gh}).$$

Bradley & Terry (1952) and Luce (1959) proposed (1) as a model for p_{gh}, also known as Bradley-Terry-Luce model (BTLM):

$$p(O_g > O_h) = p(Y_{igh} = 1) = p_{gh} = \frac{\tau_g}{\tau_g + \tau_h} \qquad 0 < \tau_g, \tau_h < \infty. \tag{1}$$

The probability to prefer object O_g over O_h is modelled as a function of the parameters τ_g and τ_h that reflect the attractivity of the objects. For $\tau_g > \tau_h$ we have $p_{gh} > 0.5$, i.e. it is more likely that O_g is preferred. For $\tau_g < \tau_h$ we have the reverse situation. Thus the two objects are modelled in a concurring manner and one could say that "O_g beats O_h" whenever $\tau_g > \tau_h$. The BTLM can be written equivalently as

$$p_{gh} = \frac{\tau_g/\tau_h}{1 + \tau_g/\tau_h}$$

and introducing the new parameters $\delta_{gh} = \tau_g/\tau_h$, letting $\alpha_{gh} = \ln \delta_{gh}$ model (1) is finally reformulated as

$$p_{gh} = \frac{\exp(\alpha_{gh})}{1 + \exp(\alpha_{gh})}. \tag{2}$$

3 IRT Models for paired comparisons

3.1 The Rasch Model - RM

In deriving model (1) we skipped the index i, i.e. the BTLM assumes that subjects are homogeneous, which is in many situations unrealistic. In order to account for subject variability Rasch (1960) proposed (3) as a model for data from mental tests. Several extensions of the model allow to handle categorical, ranking and rating data, as well as the treatment of experimental designs within the model. Fischer and Molenaar (1995) is recommended for further reading on theory and applications of various types of Rasch Models.

The RM can be viewed as a BTLM where instead of the comparison of two objects, we have the comparison of one object (item) and one subject. In the BTLM the response is modelled as function of the characteristics of two objects, whereas in the RM it is the characteristics of a subject and an object. Consequently we now have

$$Y_{ij} = \left\{ \begin{array}{ll} 1 & \text{if subject } S_i \text{ "beats" Item } I_j \text{, and} \\ 0 & \text{if Item } I_j \text{ "beats" subject } S_i \end{array} \right\} \sim Bin(1, p_{ij}).$$

The expression "beats" is used to stress the analogy with the BTLM, where O_g "beats" O_h if O_g is preferred. In mental testing of course it is better to use "S_i

does or does not solve I_j". Following the notation in (2) we have the probability for subject S_i solving Item I_j as

$$p_{ij} = p(S_i > I_j) = \frac{\exp(\lambda_i + \kappa_j)}{1 + \exp(\lambda_i + \kappa_j)}, \tag{3}$$

where λ_i is a parameter describing the subject's ability to solve the item, and κ_j is the easiness of item I_j.

For complete paired comparisons we have each of the k combinations of objects presented to the n subjects. Viewing these k combinations as items, leads immediately to a RM for these data.

Let

$$Y_{igh} = \left\{ \begin{array}{ll} 1 & \text{if object } O_g \text{ is preferred by subject } S_i\text{, and} \\ 0 & \text{if object } O_h \text{ is preferred by subject } S_i \end{array} \right\} \sim Bin(1, p_{igh}).$$

Now a Rasch model for paired comparisons is simply derived by replacing κ_j in (3) with α_{gh} from (2):

$$p_{igh} = p(O_g > O_h | S_i) = \frac{\exp(\lambda_i + \alpha_{gh})}{1 + \exp(\lambda_i + \alpha_{gh})}. \tag{4}$$

The λ_i account for subject variability, while the α_{gh} reflect the easiness of the combination, i.e. how easy it is to prefer object O_g in the combination (O_g, O_h). Of course the α_{gh} are only a technical quantities that have no interpretation in Conjoint Analysis.

3.2 The Linear Logistic Test Model - LLTM

For scaling the objects and determining the impact of the product attributes we need to know the object parameters $\boldsymbol{\beta}$ and the attribute parameters $\boldsymbol{\gamma}$. Fischer (1972, 1983, 1995) developed a technique called the Linear Logistic Test Model (LLTM) that allows to impose a set of linear restrictions on the item parameters of the RM. The idea is to re-estimate the RM while $\boldsymbol{\kappa} = \boldsymbol{Q}\boldsymbol{\eta}$ with a known $(k \times q)$ design matrix $\boldsymbol{Q}$, and $q < k$. Thus we have sort of a regression of the k item parameters on so-called basic parameters $\boldsymbol{\eta}$. It turns out that the Scaling and the Impact problem can be solved by using the LLTM. As the restrictions apply to the item parameters only, the LLTM has the same subject-specific properties as the RM.

3.2.1 The Scaling LLTM

For the Scaling LLTM we have the linear restrictions

$$\boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{\beta} \tag{5}$$

incorporated into (4), where $\boldsymbol{C}$ is a $(k \times m)$ matrix of known coefficients (Table 2). The structure of $\boldsymbol{C}$ is obtained from the BTLM. Remember that $\delta_{gh} = \tau_g/\tau_h$ and $\alpha_{gh} = \ln \delta_{gh} = \ln \tau_g - \ln \tau_h$. Replacing $\ln \tau_j$ by β_j and writing in matrix notation we

Table 2: Design matrix for the scaling LLTM.

$$
\boldsymbol{C} = \left(\begin{array}{cccccccc}
1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & \dots & 0 & 0 & 0 \\
1 & 0 & 0 & -1 & \dots & 0 & 0 & 0 \\
 & & \vdots & & \vdots & & & \\
1 & 0 & 0 & 0 & \dots & 0 & 0 & -1 \\
\hline
0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & \dots & 0 & 0 & 0 \\
 & & \vdots & & \vdots & & & \\
0 & 1 & 0 & 0 & \dots & 0 & 0 & -1 \\
\hline
 & & \vdots & & \vdots & & & \\
 & & & & & & & \\
 & & \vdots & & \vdots & & & \\
\hline
0 & 0 & 0 & 0 & \dots & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & \dots & 1 & 0 & -1 \\
\hline
0 & 0 & 0 & 0 & \dots & 0 & 1 & -1
\end{array}\right)
$$

arrive at $\boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{\beta}$. Thus the Scaling LLTM can be seen as a BTLM with subject parameters. As $Rank(\boldsymbol{C}) = m - 1$ one of the β_j must be set equal to zero for estimation.

The Scaling LLTM can be written explicitly as

$$p_{igh} = p(O_g > O_h|S_i) = \frac{\exp[\lambda_i + (\beta_g - \beta_h)]}{1 + \exp[\lambda_i + (\beta_g - \beta_h)]}. \tag{6}$$

3.2.2 The Impact LLTM

To determine the impact of the product attributes we need to know the attribute parameters $\boldsymbol{\gamma}$. Applying the same approach as above we impose linear restrictions on the object parameters:

$$\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\gamma} \tag{7}$$

As a matter of fact there is no direct method to estimated the attribute parameters $\boldsymbol{\gamma}$, but replacing $\boldsymbol{\beta}$ in (5) with (7) we derive a different set of restrictions on $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\gamma} = \boldsymbol{D}\boldsymbol{\gamma} \tag{8}$$

with $\boldsymbol{C}$ as before, $\boldsymbol{X}$ $(m \times p)$, and $p < k$. In addition to the Scaling LLTM we need a matrix $\boldsymbol{X}$ containing the objects attributes. The p attribute parameters can be estimated whenever $p < k$ and $Rank(\boldsymbol{X}) = p$. Thus for paired comparisons p the number of attributes can exceed m the number of objects. This is of special interest when the objects are not generated by the attributes. The Impact LLTM is obtained by replacing α_{gh} in (4) with $\sum_r d_{ghr}\gamma_r$

$$p_{igh} = p(O_g > O_h|S_i) = \frac{\exp(\lambda_i + \sum_r d_{ghr}\gamma_r)}{1 + \exp(\lambda_i + \sum_r d_{ghr}\gamma_r)}, \tag{9}$$

where $((d_{ghr})) = \boldsymbol{D}$ and $r = 1, \ldots, p$.

3.3 Estimating market shares

Model (6) is taylored to extract the object attractivity parameters β_j of data from paired comparisons. For market shares we need to know $p(O_j|S_i) = p_j$, subject to $\sum_j p_j = 1$, rather than $0 \leq p(O_g > O_h|S_i) \leq 1$ which is what we get from model (6). Procedures of data collection where subjects choose out of m objects are better for share estimation. The resulting categorical data are usually analysed by discrete choice models. Rasch models for such categorical data are well developed, but unfortunately not estimable in this situation where $k = 1$. The estimation of market shares from the parameters of model (6) is therefore not straightforward. The naive way to obtain market shares from paired comparisons is to count the times an object O_g has been preferred and divide it by the total number of responses, i.e. $\hat{p}_g = \sum_i \sum_{h \neq g} Y_{igh}/n(k-1)$ (see Table 1). To incorporate the information from the model one can simply replace the data Y_{igh} by the model predictions $\widehat{Y}_{igh}$. Applying this approach leads to $\hat{p}_g = \sum_i \sum_{h \neq g} \widehat{Y}_{igh}/n(k-1)$, where $\widehat{Y}_{igh}$ can be obtained (i) for the sample of subjects, (ii) for some ideal person, e.g. $\hat{\lambda}_i = 0$, or (iii) for some distribution of $\hat{\lambda}_i$. The proposed methodology is similar to the estimation of the share of choice (preference) in conventional Conjoint Analysis.

3.4 Parameter Estimation and Model Testing for the RM and the LLTM

For estimating the parameters of the RM several techniques have been proposed, applied and implemented, e.g. Maximum Likelihood (ML), Marginal ML and Conditional ML. The latter is implemented in the software LPCM-Win 1.0 (available from ProGAMMA, Groningen, NL) that is used for model estimation throughout this paper. The Conditional ML method allows to estimate the item parameters independently of the subject parameters. Thus we have a model that accounts for subject variability, while the estimation technique allows to neglect it for item parameter estimation. One could say that the item parameters are estimated in the presence of subject parameters, but without having to make any assumption about them. Thus estimation of the item parameters is not affected by assumptions about the subject parameters.

The RM was designed to handle the measurement problem when using mental tests. (As a metter of fact this problem is present whenever questionaires are used.) Therefore the assumptions of the RM refer to measurement problems and statistical problems. The most important measurement assumption is *Unidimensionality*, i.e. all items of a questionaire cover one and the same behavioural aspect. The most important statistical assumption is *(Local Stochastic) Independence*, i.e. a subject's response to one item does not affect its response to another item. Other assumptions concern item and subject raw scores and the so-called item characteristic curves. Taking all assumptions together the RM turns out to be a very restrictive model.

Various procedures for detailed testing of the assumptions of the RM have been developed. For using the RM in Conjoint Analysis only the overall validity of the model is assessed. The overall validity of the RM is usually tested by a Likelihood Ratio Test (LRT). The test statistic is

$$T = 2(\ln L_1 - \ln L_0) \qquad \sim \chi^2 \text{ with } df = p_1 - p_0,$$

with $p_1 = k - 1$ and $p_0 = S(k-1)$, the number of estimated parameters.[2] L_1 is the likelihood of the model applied to the whole data set. $L_0 = \prod_s L_s$, where the L_s are obtained when the model is applied to say S groups of subjects obtained from the sample by segmenting it according to the values of criteria like raw-score or gender. If the model is correct then the true item parameters should be the same in each group and for the data $L_1 \approx L_0$ should hold. The LRT is a general method for hypotheses testing and can be applied to other problems too. Let H_0 and H_1 denote hypotheses about the data, L_0 and L_1 the corresponding likelihoods, and p_0 and p_1 not necessarily depending on k the number of items. As an example consider the LLTM with $q < k$. Testing the LLTM against the RM we have $p_0 = q$, $p_1 = k - 1$ and $df \geq 1$. Therefore the LLTM is always a restriction to the RM that can be seen as a H_1 and tested by the LRT. Analogously different LLTMs can be tested against each other as long as $p_0 \neq p_1$. As the LLTM imposes restrictions on the item parameters of the RM, we can obtain the reproduced item parameters $\hat{\alpha}_{gh}^{rep}$ from the estimated parameters $\hat{\beta}_j$ and $\hat{\gamma}_r$ by using (5) and (8). In case of significant T it may be interesting to get more information about the possible causes. Scatterplots of $\hat{\alpha}_{gh}$, the estimates of the RM, and $\hat{\alpha}_{gh}^{rep}$ give a detailed picture of the discrepancy between the two models. Items that are badly reproduced can be identified and possibly excluded from the analysis.

4 Application to data from sound design

The proposed model is applied to data from research in car engine noise.[3] Here we have the sound recording of a set of m car engines as objects, and several characteristics of the sound as $p > m$ attributes. The research task was to identify and quantify the objective sound characteristics that are driving the subjective judgements. For $m = 12$ car engine noises we had a group of $n = 52$ subjects making judgements using the paired comparison method. The data were collected in a acoustic laboratory. The complete combination yields $k = 66$ pairs of noises. For 12 combinations almost all subjects preferred the same object. As a consequence we obtain $|\hat{\beta}| \to \infty$ and also $var(\hat{\beta}) \to \infty$ for these items, which makes them worthless for further analysis. These items contain no information about the parameters and they were therefore excluded from the analysis. The final set consists of $k = 54$ informative items. In fact any subset of the items that guarantees a minimum of connectedness in the data will suffice to estimate the β_j, although the precision of the estimates may not be satisfying.

[2] For identifiability the parameters have to be constrained, e.g. by letting $\kappa_1 = 0$ or $\sum_j \kappa_j = 0$. Therefore only $k - 1$ parameters are estimated.

[3] The author wishes to thank Franz Brandl and Wolfgang Stücklschwaiger from AVL List GesmbH for making the data available.

Table 3: Some details of the estimated models.

Model	n	Estimated Parameters	Number of Parameters	$\ln L$
RM	52	α_{gh}	53	-1449.64
RM_1	29	α_{gh}	53	- 879.69
RM_2	23	α_{gh}	53	- 537.78
Scaling LLTM	52	β_j	11	-1477.26
Impact LLTM	52	γ_r	9	-1478.90

Table 4: Likelihood Ratio Tests.

Test	T	df	χ^2_{95}
RM vs. (RM_1,RM_2)	64.3	53	70.9
RM vs. Scaling LLTM	55.2	42	58.1
RM vs. Impact LLTM	58.5	44	60.4

4.1 Model estimation

Table 3 gives the log-likelihood and some more details of the estimated models. The abbreviations RM_1 and RM_2 refer to models estimated for subjects with low (1) and high (2) raw scores.

Table 4 gives the results of the model tests. For the first test the likelihoods were obtained for a split by the mean of the subject raw scores. This is the standard procedure to test the validity of the RM. Of course all raw scores for the RM in this application are only of technical importance. In the consequent lines we see the results for the tests of the Scaling and the Impact LLTM. For the Impact LLTM there were a total of 39 attributes available, here variables related to the frequency spectrum of the noise. It turned out that some were strictly linear dependend, while other were of limited interest, so that finally we had 9 variables in X.[4]

Comparing the test statistic T with the critical value χ^2_{95} for significance level $\alpha = 0.05$ we find that all LRT in Table 4 are not significant. Therefore the RM, the Scaling LLTM and the Impact LLTM are considered valid models for the data.

Figure 1 and Figure 2 show scatterplots of $\hat{\alpha}_{gh}$ and $\hat{\alpha}^{rep}_{gh}$ for the Scaling and the Impact LLTM. The reproduced item parameters are quite close to the estimates from the RM.

[4]A technical remark: In general the design matrix Q of an LLTM can take on real values. For the use with the software LPCM-Win 1.0 it is strongly recommended to transform the columns of Q to the unit interval $[0, 1]$. When using the data just as they were generated, the program did not converge, even though it stated "Weight matrix has full column rank" and "Data are well-conditioned", while convergence was no problem after transformation to the unit interval.

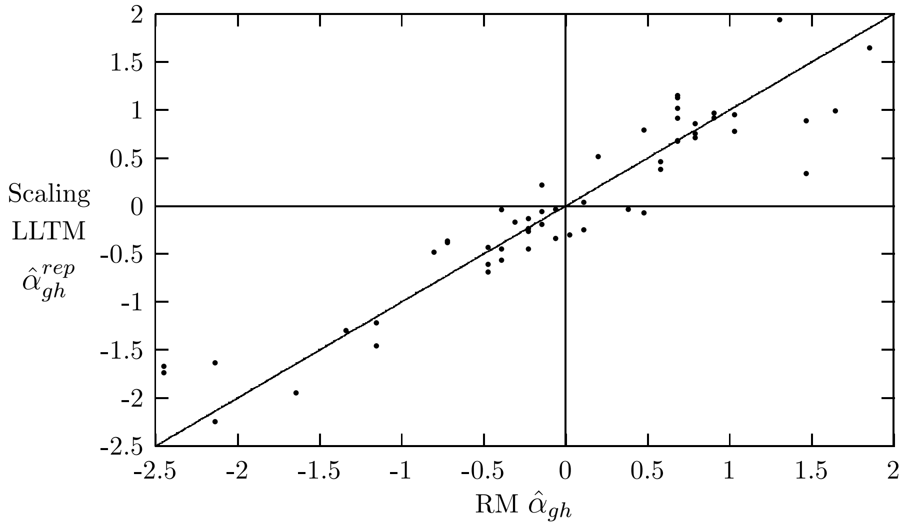

Figure 1: Scatterplot of item parameters from the RM and the Scaling LLTM.

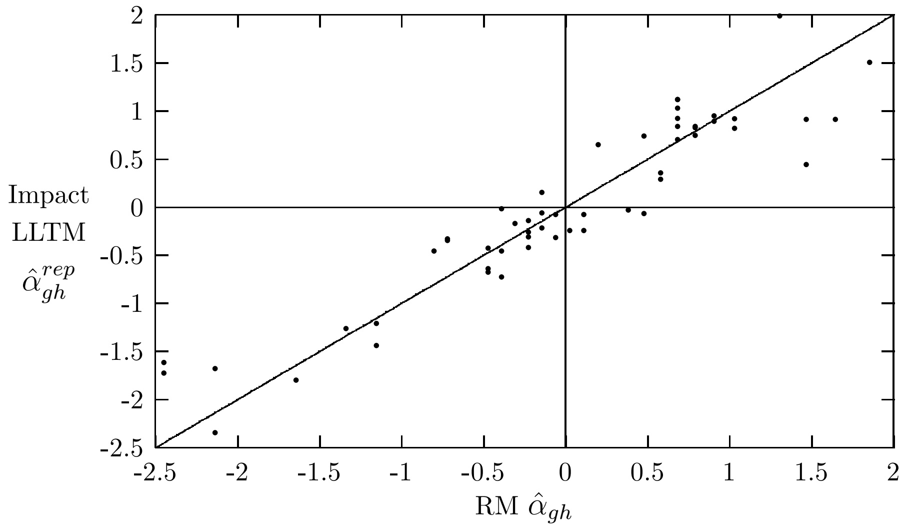

Figure 2: Scatterplot of item parameters from the RM and the Impact LLTM.

4.2 Model application

The results of the model estimation are here used to get the preference scale and the market shares of the products. The estimation of the market shares has no interpretation for the data used in this application and is only performed to illustrate the methodology. To make the effect of modelling subject variability visible results from conventional Conjoint Analysis are also included for comparison. In standard software packages like SPSS, there is no built-in handling of binary data within Conjoint Analysis. Taking the conventional approach to the situation with binary data leads to a logistic regression model (LogReg). For the estimation of the object parameters we have a design matrix X with dummy variables only. For the estimation of the attribute parameters there may be metric variables. Figure 3 and Figure 4 show the scaling of the 12 car engines as obtained from RM and LogReg.

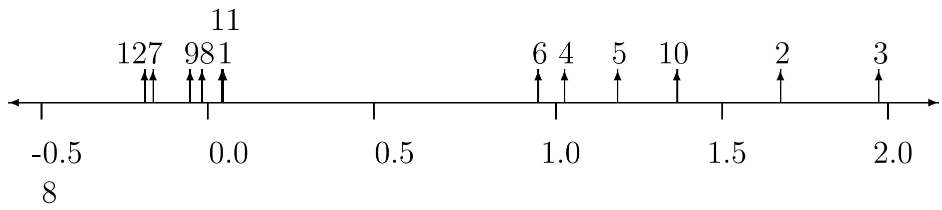

Figure 3: The preference scale for the 12 car engines as estimated by the Scaling LLTM.

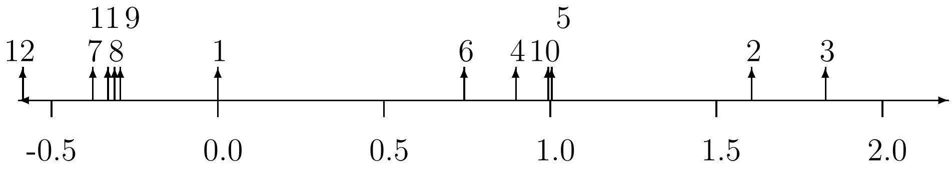

Figure 4: The preference scale for the 12 car engines as estimated by Logistic Regression.

On a rather coarse grid there is some similarity in the two scales, but a closer look reveals substantial differences in the distances between the objects as well as in the ordering of the objects. Object 10 is scaled just below object 5 by the Logistic Regression, while it is far above object 5 on the scale from the IRT model. Object 12 is clearly the least attractive on the LogReg scale while it is close to object 7 on the other scale. The objects 1, 7, 8, 9, 11 and 12 close together on the IRT scale while objects 1 and 12 are clearly separated from the rest of the group on the LogReg scale. Moreover within this group the ordering of the objects 8, 9 and 11 is reversed.

As some of the items have been skipped from the analysis, there is less information for the objects 7-11. This has a substantial effect on the share estimates and therefore the outlined approach is not applicable to this reduced data set. To get share estimates all items that involved objects 7-11 were discarded and the analysis was repeated with the remaining $k = 21$ items. The share estimates for $\lambda_i = 0$ are given in Table 5. We have parameter estimates from two models[5] (RM and Scaling LLTM) and also two share estimates. We see that the RM-based and the LLTM-based estimates are almost identical. Also contained in Table 5 are the naive estimates from the frequencies and the estimates from Logistic Regression. The naive estimates and the LogReg estimates have also some similarity, but give quite a different picture of the "market situation" when compared to the RM and the LLTM estimates. Of special concern are the objects 1, 5, 6, and 12. The most striking results are those for object 12. Assume that objects 1-6 were products

[5]For the design matrix of the Impact LLTM the software LPCM-Win 1.0 did not converge.

Table 5: Estimated market shares for 7 objects, see text.

Car engine	LLTM	RM	Naive	LogReg
1	17.77%	16.88%	9.62%	9.87%
2	23.69%	23.38%	20.05%	19.95%
3	21.12%	21.49%	20.60%	21.38%
4	15.00%	15.05%	15.11%	10.28%
5	12.72%	12.90%	16.03%	16.78%
6	7.82%	8.26%	12.55%	14.96%
12	1.79%	2.03%	6.04%	6.78%

already present on the market and that the launch performance of object 12 was investigated. The estimated share of 6-7% from conventional Conjoint Analysis would clearly indicate a positive performance of object 12, while the shares of about 2% from the IRT approach would certainly lead to a redesign or even dropping of object 12.

The obvious explanation for the observed differences between conventional Conjoint Analysis and the IRT-approach is, that accounting for subject variability can have a substantial effect on the results of the data analysis and as a matter of fact on the decisions based on such results.

4.3 Conclusion

The proposed IRT-based methodology for Conjoint Analysis with binary data from paired comparisons has the advantage of accounting for subject variability. The conditional ML approach allows to estimated the item parameters without having to make assumptions about the subject parameters. This is a crucial difference to other approaches like hierarchical Bayes models or GLMM.

Despite the fact that the RM is a very restrictive model for data we find that product evaluation is a promising area for applying this model. Especially the assumption of unidimensionality seems to be appropriate for product evaluation. When compared to solving mental tests it is obvious that product evaluation is a much simpler task and therefore the evoked mental processes are likely to have a simple structure that may well be covered by the unidimensionality assumption. Furthermore unidimensionality means that the items refer to one and the same behavioural aspect. This is also easier to fulfill in product evaluation than in mental testing.

In the application of the model to data from car engine noise investigation substantial differences compared to conventional Conjoint Analysis were observed. The preference scale of the objects and the market share estimates differed markedly. It is believed that the estimates from the IRT models are more reliable than those from conventional CA, and that the proposed approach makes a better basis for marketing decisions. As a matter of fact this believe cannot be examined within the present paper but this should be done in future marketing research.

Acknowledgements

The author thanks the anonymous referees and the editors for very helpful comments and suggestions.

References

[1] Bradley, R.A. and Terry, M.E. (1952): Rank analysis of incomplete block designs, I. the method of paired comparisons. *Biometrika*, **39**, 324-345.

[2] Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (1998): Modelling the effect of subject-specific covariates in paired comparison studies with an application to university ranks. *Applied Statistics*, **47**, 511-525.

[3] Fischer, G.H.(1972): A measurement model for the effect of mass-media. *Acta Psychologica*, **36**, 207-220.

[4] Fischer, G.H. (1983): Logistic latent trait models with linear constraints. *Psychometrika*, **48**, 3-26.

[5] Fischer, G.H. (1995): The linear logistic test model. In G.H. Fischer and I.W. Molenaar (Eds.): *Rasch Models. Foundations, Recent developments, and Applications.* New York: Springer.

[6] Fischer, G.H. and Molenaar, I.W. (Eds.) (1995): *Rasch Models. Foundations, Recent developments, and Applications.* Springer: New York,

[7] Frühwirth-Schnatter, S. and Otter, T. (1999): Conjoint analysis using mixed effect models. *Proceedings of the 14th International Workshop on Statistical Modelling, Graz, July 19-23, 1999*, 181-191.

[8] Green, M. (2000): Statistical models for conjoint analysis. mixed effect models. *Proceedings of the 15th International Workshop on Statistical Modelling, Bilbao, July 17-21, 2000*, 216-222.

[9] Luce, R.D. (1959): *Individual Choice Behaviours: A Theoretical Analysis.* New York: J. Wiley.

[10] Rasch, G. (1960): *Probabilistic Models for some Intelligence and Attainment Tests.* Copenhagen: The Danish Institute of Educational Research.