Measuring Consumer Preferences for Complex Products: A Compositional Approach Based on Paired Comparisons

Sören W. Scholz, Martin Meissner, and Reinhold Decker

WEB APPENDIX

DERIVATION OF ATTRIBUTE IMPORTANCES AND LEVEL DESIRABILITIES

We define the overall utility or desirability U of stimulus m as a function of the relative importances $\beta_1,\ldots,\beta_H$ of attributes $c_1,\ldots,c_H$ and the preference weights or desirability scores $w^h(c_{mh})$ the respondents assign to the attribute levels featured by stimulus vector x_m:

$$\text{(W1)} \quad U(x_m)=\sum_{h=1}^{H}\beta_h\sum_{l_h=1}^{L_h}w(c_{hl_h})I_{hl_h}^{m} \qquad \forall m \qquad \text{with} \qquad I_{hl_h}^{m}=\begin{cases}1, & \text{if } x_{mh}=c_{hl_h}\\ 0, & \text{otherwise}\end{cases} \qquad \forall h,m$$

In PCPM stated paired comparisons a_{ij}^h are used to reveal the attribute importances β_h and the weights $w(c_{mh})$, where $a_{ij}^h>1$ indicates a respondent's preference of element i over element j in sub-problem h. Accordingly, $a_{ji}^h=1/a_{ij}^h$ indicates the preference of element j over element i. If a respondent is indifferent regarding both elements, then $a_{ij}^h=1$ results.

Two scales are used in AHP to measure the ratios a_{ij}^h (Saaty 1980). First, the respondent has to mark on a dichotomous scale which element in the present paired comparison he or she prefers. Then, a 9-point rating scale is used to transform his or her preference strength, expressed by verbal judgments, into numerical preference values ranging from 1 ("i and j are equal") to 9 ("i is absolutely better than j"). In consumer surveys, however, we often noticed that people do not understand the artificial separation of the direction and the strength of preferences, and thus, the use of two scales. Therefore, a bipolar scale is used in PCPM which measures both the direction and the strength of preference simultaneously. The 9-point bipolar scale is depicted in Table A1.

Scale level q	**Verbal statement**	**Preference value** s_q
1	i is absolutely preferred to j	9.00
2	i is strongly preferred to j	5.20
3	i is considerably preferred to j	3.00
4	i is weakly preferred to j	1.73
5	i and j are equal	1.00
6	j is weakly preferred to i	1/1.73
7	j is considerably preferred to i	1/3.00
8	j is strongly preferred to i	1/5.20
9	j is absolutely preferred to i	1/9.00

Table A1: NEW SCALE FOR MEASURING CONSUMER PREFERENCES IN PCPM

This scale provides a geometric increase/decrease in the measured preferences between the adjacent scale levels, i.e. $\mathrm{s}_{\mathrm{q}+1} / \mathrm{s}_{\mathrm{q}} = \text{const.}$ for $q = 1, \ldots, 8$. In contrast to that, the ratios $\mathrm{s}_{\mathrm{q}+1} / \mathrm{s}_{\mathrm{q}}$ are unevenly dispersed when using Saaty's AHP scale (Salo and Hämäläinen 1997).

The stated paired comparisons $\mathrm{a}_{\mathrm{ij}}^{\mathrm{h}}$ belonging to sub-problem h make up a ratio preference network and can be represented by the pairwise comparison matrix (termed "ratio preference matrix" by Hartvigsen 2005)

$$\text{(W2)} \quad \mathrm{A}_{\mathrm{h}} = (\mathrm{a}_{\mathrm{ij}}^{\mathrm{h}})_{\mathrm{i,j}=1,\ldots,\mathrm{n_h}} = \begin{pmatrix} 1 & \cdots & \mathrm{a}_{1\mathrm{n_h}}^{\mathrm{h}} \\ \vdots & \ddots & \vdots \\ \mathrm{a}_{\mathrm{n_h}1}^{\mathrm{h}} & \cdots & 1 \end{pmatrix} \qquad \forall \mathrm{h},$$

where n_h denotes the number of elements to be considered in sub-problem h (with $n_h := L_h \ \forall h > 0$). To derive the unknown preference weights w_i^h from the paired comparisons, the following eigenvalue problem has to be solved (Saaty 2003):

(W3) $\quad A_h w_h = \lambda_h^{max} w_h \qquad \forall h$,

where $w_h = (w_1^h, \ldots, w_{n_h}^h)^T$ is the principal right eigenvector belonging to the largest eigenvalue λ_h^{max} of A_h.

According to Saaty (1980), the extent to which w_h reflects the stated paired comparisons a_{ij}^h can be captured by the so-called consistency index:

(W4) $\quad CI_h = (\lambda_h^{max} - n_h) / (n_h - 1) \quad \forall h$

This term measures the relative deviation of the collected judgments from the consistent approximation resulting from the eigenvalue problem (A3). In perfectly consistent ratio preference networks the principal eigenvalue λ_h^{max} of A_h is equal to n_h, otherwise it is greater than n_h. In order to get a notion of the consistency of A_h for varying matrix sizes, CI_h is related to the average consistency index RI of random reciprocal matrices of the same size (Saaty 1980). The resulting measure is called the consistency ratio CR_h, with $CR_h = CI_h / RI \ \ \forall h$, and captures the quality of the paired comparisons and thus provides a means to assess the data quality of the resulting importances and preference weights. Normalizing vector w_h ensures that the weights $w_1^h, \ldots, w_{n_h}^h$ sum to 1 for each sub-problem h. In doing so, the weights $w_1^0, \ldots, w_H^0$ of sub-problem 0 can be interpreted as the relative attribute importance measures and are therefore referred to as $\beta_1, \ldots, \beta_H$.

TAKING INTO ACCOUNT THE NUMBER-OF-LEVEL EFFECT IN PCPM

The normalization of w_h implies that the average preference weight $\bar{w}^h = {}^1\!/_{n_h} \sum_{i=1}^{n_h} w_i^h$ decreases when further elements are included in sub-problem h. Since the respondents can be assumed to be unaware of this fact, the eigenvalue approach has to be aligned to the number of attribute levels such that on average, each level yields the same weight. Therefore, we simply multiply the preference weights w_i^h with the number of elements n_h in the sub-problem they belong to. With this in mind, we define the aligned preference weights:

(W5) $\quad \tilde{w}(c_{mh}) = n_h w(c_{mh}) \quad \forall h > 0$.

Thus, the basic function for computing the overall utility of stimulus x_m reads as follows:

(W6) $\quad U(x_m) = \sum_{h=1}^{H} \beta_h \sum_{l_h=1}^{L_h} \tilde{w}(c_{hl_h}) I_{hl_h}^m \quad \forall m \quad \text{with} \quad I_{hl_h}^m = \begin{cases} 1, & \text{if } x_{mh} = c_{hl_h} \\ 0, & \text{otherwise} \end{cases} \quad \forall h, m$

DESIGNING EFFICIENT INCOMPLETE RATIO PREFERENCE NETWORKS

In large sub-problems, the number of paired comparisons increases exponentially with the number of the elements n_h. Miyake et al. (2003) propose a simple and effective method to select a small subset of these paired comparisons which is called two-cyclic design.

Let

(W7) $\quad Q_z = \left\{ \left(k_1^t, k_2^t\right) \middle| \left(1, z+1\right), \left(2, z+2\right), \ldots, \left(n_h, z+n_h\right) \right\}$ with $z \in \mathbb{N}$.

This set of pairs makes up a one-cyclic design defining the pairs to be compared by the respondent $(\tilde{k}_1^t, \tilde{k}_2^t)$ with

(W8) $\tilde{k}_i^t = \begin{cases} k_i^t, & \text{if } k_i^t \le n_h \\ \left(k_i^t \bmod n_h\right)+1, & \text{else} \end{cases}, \quad \forall t; i=1,2.$

A one-cyclic design is sufficient to estimate the resulting preference weights if, and only if, each element is included in exactly two paired comparisons, i.e., the corresponding pairwise comparison matrix is irreducible. The solid line in Figure A1 displays such a one-cyclic design satisfying this criterion. Considering sub-problems with $n_h \ge 5$, there exist $n_h/2$ different one-cyclic designs for sub-problems comprising even number of elements. For odd numbers of elements, $(n_h - 1)/2$ different one-cyclic designs exist (Miyake et al. 2003).

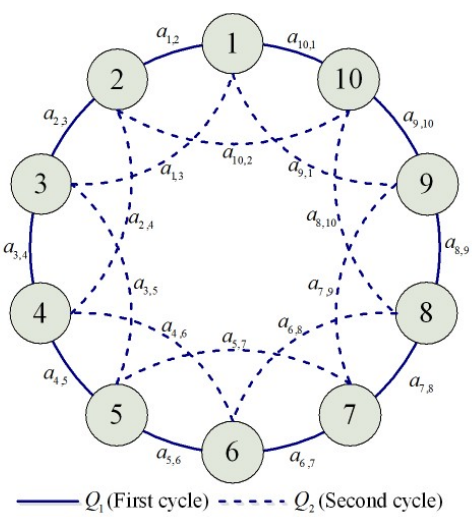

Figure A1: PAIRED COMPARISONS OF THE TWO-CYCLIC DESIGN FOR A SUB-PROBLEM INCLUDING 10 ELEMENTS

Let a two-cyclic design be defined as the combination of two different one-cyclic designs, where the two sets of paired comparisons Q_z and $Q_{z'}$ are disjoint. We then seek for two one-cyclic designs which comprise altogether $2n_h$ unique comparisons. The first two cyclic designs (i.e. $z=1$ and $z=2$) meet this criterion for all sub-problems with $n_h \ge 5$. Figure A1 shows the corresponding paired comparisons of a sub-problem including 10 elements. Here, the paired comparisons of the first one-cyclic design (with $z = 1$) are depicted by the solid line while the dotted line displays the second one-cyclic design (with $z = 2$). Two-cyclic designs for sub-problems including varying numbers of elements can be developed analogously.

Please note that the inclusion of incomplete comparisons affects the eigenstructure of the resulting pairwise comparison matrix. Usually, the consistency index decreases with increasing numbers of missing paired comparisons. This effect can be taken into account by calculating the average consistency index RI of random reciprocal matrices of the same size and the same missing paired comparisons when measuring the consistency ratio CR_h.

USING PREFERENCE DIFFERENCE NETWORKS TO ESTIMATE PREFERENCE WEIGHTS

Interval scales are used instead of ratio scales to measure the respondents' preference weights in the interval-scale PCPM approach. Accordingly, a paired comparison a_{ij}^h measures the stated preference difference between two elements, i.e. $a_{ij}^h = w_i^h - w_j^h$. Since we assume a symmetric relationship between each two elements, $a_{ji}^h = w_j^h - w_i^h$ applies. The following scale values are used to transform the verbal statements depicted in Table A1 into numerical values: $s_1 = 4; s_2 = 3; s_3 = 2; s_4 = 1; s_5 = 0; s_6 = -1; s_7 = -2; s_8 = -3; s_9 = -4$. Similar to the geometric ratio scale used in original PCPM, this scale provides an arithmetic increase/decrease in the measured preferences between adjacent scale levels. If a respondent is indifferent regarding two elements then $a_{ij}^h = 0$ results. Positive values indicate a respondent's preference of element i over element j in sub-problem h and vice versa.

If the paired comparisons are perfectly consistent, then $a_{ij}^h = a_{ik}^h + a_{kj}^h \quad \forall i, j, k = 1, \ldots, n_h$ holds for all paired comparisons. In this case, the unknown preference weights w_{ij}^h are derived as follows: Without loss of generality set $w_1^h = 0$. Then the remaining preference weights are calculated as $w_j^h = w_1^h - a_{1j}^h = -a_{1j}^h$.

In sub-problems comprising many elements, the missing comparisons $\underline{a}_{ij}$ can be calculated in a preference difference network. Analogously to the ratio preference network, a missing paired comparison is calculated by means of the arithmetic mean of the connecting paths between the respective elements i and j:

(W9) $\quad \hat{\underline{a}}_{ij}^h = \dfrac{1}{R_{ij}^h} \sum_{r=1}^{R_{ij}^h} CP_{ij}^{hr} \;\; \forall i, j$,

where

(W10) $\; CP_{ij}^{hr} = a_{ik_1}^h + a_{k_1 k_2}^h + \ldots + a_{k_{t-1}k_t}^h + \ldots + a_{k_T j}^h$ with $\; k_1, \ldots, k_T \in \{1, \ldots, n_h\} \setminus i, j$.

The above procedure works as long as at least one connecting path CP_{ij}^{hr} for each missing element $\underline{a}_{ij}^h$ exists.

Usually, the stated paired comparisons are not fully consistent. In this case, it is advisable not to use the directly stated paired comparison a_{ij}^h but to use the complete information of all elementary paths CP_{ij}^{hr} between two elements *i* and *j* to calculate the preference weights. Accordingly, we first calculate $\hat{\underline{a}}_{1j}^h$ by the arithmetic mean of all connecting paths $\forall j = 2, ..., n_h$, irrespective whether the element is missing or not (see Equations A6 and A7). These estimates are then used for the calculation of the preference weights as outlined above.

THE IMPACT OF ERRONEOUS PAIRED COMPARISONS IN INCOMPLETE RATIO PREFERENCE NETWORKS – A MONTE CARLO SIMULATION STUDY

Due to the multiplicity of connecting paths, complete ratio preference networks can be used for the elicitation of robust preferences, even in the case of substantial errors in the paired comparisons (Scholz, Meißner, and Wagner 2006). In incomplete ratio preference networks however, the number of connecting paths between two elements is reduced which can harm the accuracy and robustness of the elicited preferences.

Monte Carlo simulations have been carried out to assess the impact of these measurement errors for varying numbers of collected paired comparisons. Let $w^{true} = (w_1^{true}, w_2^{true}, \ldots, w_{10}^{true})$ be the true preference weights of a fictitious respondent. In accordance with the basic assumptions of ratio theory (Hauser and Shugan 1980), the simulations are based on the assumption of a log-normally distributed error ε_{ij} (with mean equal to 1 and standard deviation equal to 0.732) which affects the stated preference ratios, i.e., $a_{ij} = \frac{w_i^{true}}{w_j^{true}} \varepsilon_{ij}$. In doing so, each "stated" paired comparison, on average, deviates by one scale value from the actual preference ratio w_i^{true} / w_j^{true} (*cf.* Table A1). Since the applied 9-point scale only allows discrete values, the stated preferences a_{ij} are always rounded to the adjacent scale value.

The design of the Monte Carlo simulations was as follows: First, 100 fully consistent ratio preference networks, including ten elements each, were constructed to minimize the impact of varying true preference weights w^{true}. Then, the above-sketched distortion was applied to these ratio preference networks and 100 runs were carried out for each initial pairwise comparison matrix. We systematically varied the number of paired comparisons from the minimum ($n-1=9$) to the maximum ($n(n-1)/2=45$) by randomly deleting the unused preference ratios.

Finally, the mean absolute error (MAE) between the true preference weights w^{true} and the weights resulting from the eigenvalue approach was computed for each run to assess the impact of distortion. The results are displayed in Figure A2 and show a significant decrease in measurement error when increasing the number of paired comparisons. In fact, with each additional paired comparison, further elementary paths are added to the ratio preference network, which leads to an increasingly robust and accurate estimation of the preferences values. Obviously, high numbers of different elementary paths between any two elements of the ratio preference network are leveling out even substantial noise in the preference judgments. Although the simulations do not identify a univocal solution regarding the optimal number of paired comparisons, diminishing returns for each additional paired comparison can be substantiated.

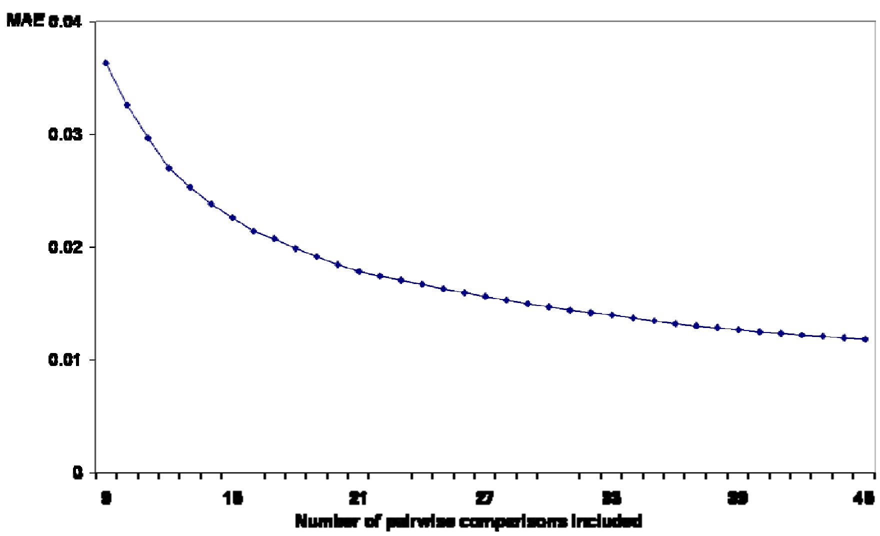

Figure A2: MEAN ABSOLUTE ERROR FOR DIFFERENT NUMBERS OF PAIRED COMPARISONS

As outlined above, we use the two-cyclic design by Miyake *et al*. (2003) to reduce the set of paired comparisons to an efficient subset including only $2n$ comparisons. Here, each element is included in four paired comparisons (see Figure A1). If there is an erroneous paired comparison, 3 further comparisons can smooth this effect. Noticably, this number of paired comparisons can only be gathered in sub-problems that include at least $n = 5$ elements (since $n(n-1)/2 < 2n = 5$ holds for $n \leq 4$). Therefore, we recommend collecting the maximum number of paired comparisons in smaller sub-problems.