VALIDITY OF CUSTOMIZED AND ADAPTIVE HYBRID CONJOINT ANALYSIS

S. Hensel-Börner, H. Sattler

Unilever Chair of Marketing and Management Science,
Friedrich-Schiller-University of Jena, Carl-Zeiß-Str. 3, D-07743 Jena, Germany

Abstract: In this paper we compare the validity of a new type of customized hybrid conjoint analysis called Customized Computerized Conjoint Analysis (CCC) with the Adaptive Conjoint Analysis (ACA) and two self-explicated models. CCC combines self-explicated preference structure measurement with individually designed full-profile conjoint analysis in a fully computerized interview. For a sample of almost 500 German potential customers of refrigerators we find (similar to Srinivasan, Park 1997)) surprisingly robust results of the self-explicated approaches compared to CCC as well as ACA.

1 Introduction

Over the past two decades, conjoint measurement has been a popular method for measuring customers' preference structures. As early as the mid 80s, Wittink and Cattin (1989) estimated the number of commercial applications at 400 per year only in the US. In Europe, nearly 1000 conjoint studies were carried out by market research companies between 1986 and 1991 (Wittink et al. (1994)). In the meantime, more than 1000 conjoint analyses are performed each year in practical applications world-wide.

While traditional full-profile conjoint analysis with a small number of attributes (up to 6) typically leads to valid results, using a large number of attributes may cause problems of validity due to an information-overload of the respondents (Green, Srinivasan (1990)). In order to address this problem several types of hybrid conjoint analysis have been developed (Green (1984); Johnson (1987); Green, Krieger (1996); Baier, Säuberlich (1997)). Hybrid conjoint approaches combine compositional methods such as self-explicated approaches with decompositional models like full-profile conjoint analysis. Most promising for applications with many attributes seem to be customized and adaptive hybrid approaches because these approaches are able to select a research design in order to take advantage of the information obtained from respondent's answers in preceding steps of data collection. Today, an adaptive hybrid type of conjoint analysis called Adaptive Conjoint Analysis, ACA (Johnson (1987)) is the most popular method for measuring customers' preference structures (Wittink et al. (1994); Green et al. (1991)).

Despite its popularity, ACA entails several problems such as graded paired-comparison (Green et al. (1991)). Empirical studies that compare ACA with traditional full-profile conjoint analysis and self-explicated approaches respectively, find a slightly poorer or at best the same internal validity of ACA (Green, Srinivasan (1990); Green et al. (1991); Agarwal, Green (1991); Green et al. (1993); Huber et al. (1993)). In view of these results, the question arises whether alternative methods lead to more valid results than ACA. Possible alternatives for applications with large numbers of attributes (more than 6) are self-explicated approaches and, in particular, as pointed out, adaptive or customized hybrid methods (Green, Srinivasan 1990).

Our study aims at testing the validity of a new customized hybrid method called *Customized Computerized Conjoint Analysis* (CCC) and comparing it to ACA and to two self-explicated models.

2 Customized Computerized Conjoint Analysis (CCC)

Our new method called *Customized Computerized Conjoint Analysis* (CCC) is an extension of *Customized Conjoint Analysis* (CCA) proposed by Srinivasan and Park (1997). Therefore, we will first briefly describe CCA in the following section. After this, we will point out the main extensions of CCC (for details concerning CCA see Srinivasan, Park (1997)).

CCA can be divided into three main stages. First**,** similar to ACA, it starts with the possibility for respondents to identify completely unacceptable levels for all attributes investigated. For each remaining attribute respondents determine the most-preferred and least-preferred levels. Setting these desirability-ratings to 10 and 0 respectively, the other attribute levels are evaluated using this 11-point rating scale. After rating the single attribute levels, the overall attribute importances are evaluated on the same scale using the most important attribute as an anchor. From these data the first stage self-explicated partworths P_{ijk}^{SE} can be computed:

$$(1) \qquad P_{ijk}^{SE} = I_{ij}(D_{ijk} / 10)$$

where:

P_{ijk}^{SE}: respondent i`s self-explicated partworth for attribute j`s k-th level,

I_{ij}: respondent i`s importance (0-10) for attribute j, and

D_{ijk}: respondent i`s desirability rating (0-10) for attribute j`s k-th level.

Second, based on the results of the first stage, the so-called ”core” attributes are identified for each respondent individually. Core attributes are the most important attributes (not more than 6) which are included in a full-profile conjoint analysis. Full-profile stimuli for the conjoint stage are designed according to a fractional factorial design. As the core attributes differ from person to person the set of full-profile stimuli (e.g. 8 to 16 stimuli) is customized to each person (Customized Conjoint Analysis). Based on the preferences for the customized profiles normalized conjoint partworths P_{ijk}^{CA} can be estimated for each respondent.

Third, in a calibration phase the optimal combination of the estimated self-explicated partworths P_{ijk}^{SE} and conjoint partworths P_{ijk}^{CA} for the core attributes is determined. To do so respondents are again asked to rate or rank full-profile stimuli described by their core attributes. This task is similar to stage 2. However, different combinations of attribute levels are presented. The calibration phase might also be placed between the self-explicated and the full-profile main task to minimize biasing the result in favor of either procedure (Srinivasan, Park (1997)).

From this data we obtain the weighted partworths P_{ijk}^{w}:

$$(2) \qquad P_{ijk}^{w} = w_i P_{ijk}^{CA} + (1 - w_i) P_{ijk}^{SE}$$

where:

P_{ijk}^{w} : respondent i`s weighted partworth for attribute j`s k-th level,

w_i : respondent i`s weight, and

P_{ijk}^{CA} : respondent i`s conjoint partworth for attribute j`s k-th level.

Preferences for the stimuli in the calibration stage are predicted for different weights (0, 0.01, 0.02, ... , 1) to identify the optimal weight w_i* that produces the highest correspondence (i.e. cross validity) between actual and predicted preferences. With the optimal w* the predicted preference of respondent i for a new stimulus with level k_j for attribute j is given by:

$$(3) \qquad U_i = \sum_{j \in C_i} \left[w_i^* P_{ijk_j}^{CA} + (1 - w_i^*) P_{ijk_j}^{SE} \right] + \sum_{j \in NC_i} P_{ijk_j}^{SE}$$

where:

U_i : respondent i`s predicted preference,

w_i^* : respondent i`s optimal weight,

C_i : respondent i`s core attributes, and

NC_i : respondent i`s non-core attributes.

CCC contains basically two extensions of CCA. First, CCC uses a fully computer-based interview where all data are collected during one interview session. In contrast, Srinivasan and Park (1997) reported a time lag of two weeks between the first and second data collection stage. This time lag in CCA is necessary because the full-profile conjoint-stimuli in stage 2 have to be customized according to the results of stage 1. Apart from possible confound due to the time lag, the method is quite time and cost consuming. Combining all three stages in a single interview overcomes these main problems.

As a second extension of CCA we modify equation (3) concerning the computation of the preferences toward a new product or stimulus. The estimation procedure proposed by Srinivasan and Park (1997) allows for attribute importance weights of a core attribute to be (substantially) lower than the importance weight of a non-core attribute. In a large scale pretest with 72 respondents using coffee as stimuli we found that this was true for 86% of the respondents. This effect occurs particularly if some attributes within the selected core-attributes become dominant in the conjoint task or when there are non-core attributes with almost equal importance weights in the self-explicated part. To overcome this problem we propose a new approach to calculate weighted partworths:

$$U_i = \sum_{j \in C_i} \left[w_i^* P_{ijk_j}^{CA} + (1 - w_i^*) P_{ijk_j}^{SE} \right] + \sum_{j \in NC_i} (1 - w_i^*) P_{ijk_j}^{SE} \qquad (4)$$

Compared to CCA the probability that importance weights for core-attributes will be higher than those for the non-core-attributes is substantially higher for CCC due to the new weighting scheme employed by equation (4). The new model does not guarantee that all importance weights for core-attributes are higher than for the non-core attributes. This would imply that the distinction between core- and non-core attributes as well as the estimation of the self-explicated partworths can be done error-free. Whether formula (3) or (4) fits better to the actual preference structure of a respondent is mainly influenced by the distribution of the importance weights for each respondent and will be tested empirically.

3 Research Design

The purpose of our empirical study is to compare the validity of CCC with ACA as well as two different self-explicated models. With ACA we chose a widely used adaptive hybrid conjoint approach which has almost become the standard software for conjoint analysis. A comparison to self-explicated models is also of particular interest, because data collection as well as data analysis is much easier for these procedures. Furthermore, several empirical studies report robust validity of the self-explicated approach (Srinivasan, Park (1997); Green, Srinivasan (1990); Sattler, Hensel-Börner (1999)). In our study, we use the self-explicated models implemented in CCC and ACA labeled SE-CCC and SE-ACA respectively. Moreover, in order to compare the effectiveness of the new approach to calculate weighted partworths implemented by equation (4) with the approach proposed by Srinivasan and Park (1997) (see equation (3)) we apply both methods – called CCC-new and CCC-old respectively - in our study. Thus, in total we compare 5 methods: CCC-new, CCC-old, ACA, SE-CCC and SE-ACA.

In co-operation with the German market research company IPSOS, Germany data were collected from 242 respondents with CCC as well as from 249 different respondents with ACA in three German cities (Hamburg, Frankfurt/Main and Munich). In both subsamples a representative quota sampling procedure was used. A χ^2–test of homogeneity between the two independent samples with respect to several demographic data such as sex, age and education shows non significant differences ($p>0.10$). Therefore, we may conclude that the two samples are comparable and differences in the results might not be influenced by systematic deviations in the samples.

We use preferences with respect to refrigerators as context for the empirical application. 8 attributes of refrigerators with 2 to 4 levels were chosen (Tab. 1), which turned out to be relevant in a pretest.

Tab. 1: Refrigerator attributes and levels

Attribute:	**Levels:**
Brand name	Siemens, Liebherr, Electrolux, Quelle
Price	DM 449; DM 549; DM 649
Energy Consumption	kWh per year: 150; 200; 250
Warranty	12 months; 18 months; 24 months
Capacity	Liter: 115; 130; 145
Interior Decoration	fixed; partly variable; completely variable
Defrost	manual; automatic
Freezer	-6°C; -18°C

First, the respondents evaluated all 8 attributes and their levels in the self-explicated stage of CCC and ACA respectively. Next, for CCC the respondents were asked for their preferences with respect to 9 full-profile cards in the main task as well as in the validation task, each described by 4 customized core attributes. In order to design the “main” and “calibration” profiles we used a fractional factorial design. For ACA the self-explicated stage was followed by 19 paired-comparisons of stimuli first described by 2, and later by 3 attributes (according to the default settings of ACA). As a final step in the ACA-interview the respondents had to evaluate 6 calibration stimuli, each described by 5 attributes. At the end of the interview the respondents were required to answer demographic questions and to give information on past brand purchases. In order to validate the different approaches we used a holdout sample of 5 refrigerators described by 5 different attributes with respect to existing offers in the real market place.

4 Preliminary Results

In order to compare the 5 approaches described we use several measures of convergent and predictive validity. At this point the authors would like to emphasize that the following results are preliminary and we are working on several extensions, especially with respect to further validation criteria.

4.1 Convergent Validity

Because the 5 methods under consideration use different procedures for collecting data and calculating partworths, we first test whether the importance weights obtained from the methods are the same (convergent validity). If the importance weights are identical (i. e., not significantly different), we can expect equal predictive validity for all methods, because all further computations are based on the estimated partworths. Otherwise, there might be differences in the preference structures identified by the methods. To measure the convergent validity, we first use each method to estimate the importance weights at the individual level. Second, we make pairwise comparisons of the 5 methods with respect to the distributions of individual attribute importance weights calculated for each method. In total, 80 paired comparisons are made ((n * (n-1) / 2) * 8 attributes = 80, where n is the number of methods). A Kolmogorov-Smirnov-Test shows non-significant differences between the distributions of individual attribute importance weights ($p>0.10$) for only 3 out of the 80 paired comparisons . Thus, there is almost no convergent validity, i. e., the methods result in different preference structures. However, it is not possible to say which method comes closest to the “true” preference structure.

4.2 Predictive Validity: First-Hit-Rates

In order to test for predictive validity we calculate the first-hit-rates for each method. To do so, we use the individual partworths for each method estimated with the calibration sample to predict choice behavior in the holdout sample. We then calculate the percentage of choices correctly predicted in the holdout sample based on the first choice rule (Green, Krieger (1988)). Thus, the first-hit-rate indicates in which percentage of observations in the holdout sample the refrigerator with the highest predicted utility was actually chosen.

Tab. 2: Predictive Validity with Respect to First-Hit-Rate

	ACA	**CCC-old**	**CCC-new**	**SE-ACA**	**SE-CCC**
All respondents	249	242	242	249	242
Percentage of choices correctly predicted in the holdout sample	45.4	44.2	41.3	40.2	39.7
Percentage improvement over random model a)	31.7 ***	30.3 ***	26.7 ***	25.2 ***	24.6 ***

a) (100*[(percent correctly predicted - percent correctly predicted by random model)/(100 - percent correctly predicted by random model)]) (Srinivasan, Park (1997))

The results in Tab. 2 show a higher first-hit-rate for all three hybrid conjoint methods than for the two self-explicated approaches. The highest first-hit-rate can be observed for ACA and CCC-old. However, there are no significant differences between the methods. Pairwise comparisons of all methods show non-significant differences ($p>0.1$) between any of the pairs. All methods predict significantly better than the random model ($p<0.01$), but the size of the improvements over a random model is quite similar for all methods. Thus, an important result is that, although the approaches lead to significantly different distributions of the attribute importance weights, their predictive validity measured by the first-hit-rate is very similar, even for the two self-explicated approaches.

It comes as a surprise that CCC-new does not result in higher predictive validity than CCC-old. As discussed in chapter 2, CCC-new is expected to do better than CCC-old if some attributes among the selected core-attributes become dominant in the conjoint task or if there are non-core attributes with almost equal importance weights in the self-explicated part. Therefore, one explanation for our result of non-significant differences between CCC-new and CCC-old might be that the two conditions mentioned are of little relevance for the data under consideration.

4.3 Predictive Validity: Simulated Market Shares

Based on the data obtained in the holdout sample we can calculate ”market shares” for the 5 brands used in the holdout task. In order to test for predictive validity we also predict market shares for the same 5 brands using the estimated partworths of each method applying the First-Choice-Model as well as the Bradley-Terry-Luce-Model (BTL-Model) (Green, Krieger (1988)). For both models we compute correlations between market shares observed in the holdout task and the predicted market shares for each of the 5 approaches under consideration (Tab. 3). Furthermore, we compute the mean absolute deviation between observed and predicted market shares for all 5 approaches (Tab. 3).

Tab.3:Predictive Validity with Respect to Market Shares (Holdout Sample)

	First-Choice- Model		**BTL-Model**	
Approach	Correlation Coefficient Pearson (Kendall)	Mean absolute deviation	Correlation Coefficient Pearson (Kendall)	mean absolute deviation
ACA	0.43 (0.11)	12.05	0.65 (0.53)	6.71
CCC-old	0.67 (0.4)	10.75	0.79 (0.6)	6.00
CCC-new	0.63 (0.4)	8.02	0.77 (0.8)	6.32
SE-ACA	0.43 (0.32)	18.32	0.44 (0.53)	7.20
SE-CCC	0.67 (0.53)	15.37	0.73 (0.6)	6.13

The results for the BTL-Model are again very similar across methods (with SE-ACA being a minor exception): Correlation coefficients as well as mean absolute deviations between actual and predicted market shares are almost the same for all approaches. A comparison of the results of the BTL-Model with those of the first-choice-model reveals substantially greater mean absolute deviations for the later model for all 5 approaches. Thus, it can be concluded that for our data the first-choice-model is less appropriate for predicting ”real” choice behavior. Also, for the first-choice-model predictions are substantially worse with ACA and SE-ACA than with the other three approaches.

5 Summary and Implications

With customized computerized conjoint analysis (CCC) we propose a new hybrid conjoint method for measuring preferences. CCC is an extension of the customized conjoint analysis (CCA), recently published by Srinivasan and Park (1997). Since our method uses a fully computerized interview, it overcomes a main disadvantage of CCA. In addition, CCC uses a modified estimation procedure in order to compute the preferences for a stimulus (e.g. a new product). In an empirical application we compare the validity of CCC with that of ACA (the most often used type of adaptive hybrid conjoint analysis) and two self-explicated models. While we find non-convergent results of the five methods with respect to attribute importance weights estimated at the individual level, there are no significant differences between the 5 approaches in terms of choice behavior, i.e. both measures of predictive validity are very similar for the hybrid conjoint analysis as well as for the self-explicated models.

In the light of these results, we can first conclude that the proposed method CCC works as well as the widely established ACA. However, similar to the results reported by Srinivasan and Park (1997), we can not show a significant improvement for hybrid conjoint methods over self-explicated methods. In terms of first-hit-rates the conjoint methods are slightly (but not significantly) better than the self-explicated approaches, and with respect to simulated market shares no clear advantages can be shown. Therefore, the robustness of the self-explicated methods constitutes a second important result of our empirical analysis. Given this result, future applications of hybrid conjoint analysis for measuring customers' preference structures seem to be at least questionable because of the advantages of self-explicated approaches in terms of ease, time effort and costs. Nevertheless, besides additional tests of validity (e.g. ability of the approaches under consideration to predict real market shares) further empirical work is needed to expand our conclusions to a broader set of circumstances.

References

AGARWAL, M.K. and GREEN, P.E. (1991): Adaptive Conjoint Analysis versus Self-Explicated Models: Some Empirical Results. International Journal of Research in Marketing 8, 141-146.

BAIER, D. and SÄUBERLICH, F. (1997): Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse. Zeitschrift für betriebswirtschaftliche Forschung und Praxis 49, 951-972.

GREEN, P.E. (1984): Hybrid Models for Conjoint Analysis: An Expository Review. Journal of Marketing Research 21, 155-169.

GREEN, P.E. and KRIEGER A.M. (1988): Choice Rules and Sensitivity Analysis in Conjoint Simulators; Journal of the Academy of Marketing Science; 16 (Spring), 114-127.

GREEN, P.E. and KRIEGER, A.M. (1996): Individualized Hybrid Models for Conjoint Analysis; Management Science 42 (6), 850-867.

GREEN, P.E. and SRINIVASAN, V. (1990): Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice; Journal of Marketing 54 (October), 3-19.

GREEN, P.E., KRIEGER, A.M. and AGARWAL, M. (1991): Adaptive Conjoint Analysis: Some Caveats and Suggestions. Journal of Marketing Research 28 (May), 215-222.

GREEN, P.E., KRIEGER, A.M. and AGARWAL, M. (1993): A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. Marketing Letters 4:4, 369-380.

HUBER, J.C., WITTINK, D.R., FIEDLER, J.A and MILLER, R. (1993): The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. Journal of Marketing Research 30 (February), 105-114.

JOHNSON, R.M. (1987): Adaptive Conjoint Analysis. Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, Ketchum, ID: Sawtooth Software, Inc., 253-265.

SATTLER, H. and S. HENSEL-BÖRNER (1999): A Comparison of Conjoint Measurement with Self-Explicated Approaches. A. Gustafsson, A. Herrmann und F. Huber (Hrsg.): Conjoint Measurement: Methods and Applications.

SRINIVASAN, V. and PARK, C.S. (1997): Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement. Journal of Marketing Research 34 (May), 286-291.

WITTINK, D.R. and CATTIN, P. (1989): Commercial Use of Conjoint Analysis: An Update. Journal of Marketing 53, 91-96.

WITTINK, D.R., VRIENS M. and BURHENNE, W. (1994): Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections. International Journal of Research in Marketing 11 (1), 41-52.