The Validity of Conjoint Analysis: An Investigation of Commercial Studies Over Time

Sebastian Selka, Daniel Baier[1] and Peter Kurz[2]

[1] Chair of Marketing and Innovation Management,
Institute of Business Administration and Economics,
Brandenburg University of Technology Cottbus, Postbox 101344, 03013 Cottbus, Germany
{selkaseb, baier}@tu-cottbus.de

[2] TNS Infratest Forschung GmbH
Landsberger Str. 284, 80687 Munich, Germany peter.kurz@tns-infratest.com

Abstract. Due to more and more online questionnaires and possible distraction – e.g. by mails, social network messages, or news reading during the processing in an uncontrolled environment – one can assume that the (internal and external) validity of conjoint analyses lowers. We test this assumption by comparing the (internal and external) validity of commercial conjoint analyses over the last years. Research base are (disguised) recent commercial conjoint analyses of a leading international marketing research company in this field with about 1.000 conjoint analyses per year. The validity information is analyzed w.r.t. research objective, product type, period, incentives, and other categories, also w.r.t. other outcomes like interview length and response rates. The results show some interesting changes in the validity of these conjoint analyses. Additionally, new procedures to deal with this setting will be shown.

1 Introduction

Conjoint Analysis (CA) is a wide used and well established method for measuring consumer preferences (see Green and Srinivasan (1990), Sattler and Hartmann (2008)). Even today, after more than 40 years of research in the context of marketing science, CA is still an object of investigation in this field of research (e.g., Green and Rao 1971, Green et al. (2001), Meissner and Decker (2009), Pelz (2012), Sänn and Baier (2012), Selka et al. (2012), Toubia et al. (2012)). Therefore, an extensive view on conjoint analyses (CAs) seems to be productive. Nowadays, with a lot of distraction potentials due to the ubiquitous availability of information in the internet in general (e.g., online newspapers, Twitter) or social network services (e.g., Facebook, LinkedIn) a lower validity for computer-assisted self interviews (CASI) on the web could be expected. Further the dynamic nature of consumers' preferences, measurement errors due to fatigue and boredom as well as the often stated learning effects of respondents can have negative impacts on CA' validity (see McCullough and Best (1979), DeSarbo et al. (2005), Netzer et al. (2008)). Information overload is also stated as potential source of lower internal and external validity (e.g., Jacoby (1984), Chen et al. (2009)). In contrast to this background, some intuitive presumptions, based on methodological and technological development in the past, will be given and completed with an analysis of the mean validity and the validity variance of conjoint analysis over time.

2 Intuitive Presumptions and Research Questions

Due to the evolutional development of conjoint analytical data collection methods, many different methods and approaches arose. There are traditional methods to avoid information overload by shrinking potential attribute combinations through orthognal designs (Addelman (1962)) or modern methods to collect as many data as possible by each respondent by asking as few questions as possible (Johnson (1987), Green and Krieger (1996), Netzer et al. (2008)). Besides them, there are other approaches, which produce much better validity values due to the usage of modern computer technology (e.g., Allenby et al. (1995), Johnson and Orme (2007)). Additionally there is an intuitive presumption of learning effects of research company employees regarding the question of "How to do a good CA".

Against this background and even with the potential negative impacts due to the distraction potentials of the internet, a validity gain over time in the conjoint analytical research area is expected. On the one hand, this expectation is based on the developments in technology and on the other hand on the previously given methodological developments in the past. Both expectations are probably compensating given negative impacts mentioned in section 1. Furthermore, support for this presumption could be found in other scientific areas (see Day and Montgomery 1983, Landeta 2006). Insofar, the research question could be given as "Is there a validity gain in CA over time?" and potential intuitive answers to this research question can be derived as hypotheses in such a manner:

H_1: The mean validity of CAs is increasing over time.
H_2: The mean validity variance CAs is decreasing over time.

By following these hypotheses, a brief introduction into validity measures should be given here. Just to clearify the understanding of them within the scope of this paper. Typcially, validity in CA context is measured by internal and external validity values. Internal validity values are represented through the Root Likelihood (RLH) value, which is measuring the correlation between estimated respondent answers and the given ones. Other correlation coefficients are also possible (and typical; e.g., R^2), but not within the scope of this paper. Typically the RLH vlaues are measuring the validity of hierarchical bayes (HB) estimation models of CAs, which were applied to all data sets here. The Mean Absolute Error (MAE) measures the errors between the estimated data model and the prior calculated simulation results. Therefore, the MAE values are also a measure for the internal model fit. For measuring the external validity, here, First-Choice Hit-Rates (FCHR) are given. Upcoming analyses here are based on linear regression models and F-Statistics (to test hypothesis H_1) and on Breusch-Pagan- (BPT) and Goldfeld-Quandt-Tests (GQT) to check for heteroscedasticity (and therefore to test hypothesis H_2).

3 Database of Recent Commercial CAs

To investigate the research question and to support or reject the given hypotheses, a database containing 2093 data sets of commercial CAs over a time $(1996-2011)$ was analyzed. The data sets are provided from a german market research institute. Therefore, this database just contains data about german CAs. To come up with the meta-data examination, it should be summarized here, that the given data contains information about…

- …the end date, the topic, the user amount, the drop-off rate,
- the representativeness of the study, the questionnaire duration,
- the purpose, the usage of incentives, the multimedia usage,
- the questionnaire type (CASI, CAPI), the used approach (e.g. ACA, CBC),
- the features and levels and (of course)
- some validity values (RLH, MAE, FCHR) for each CA.

In a first analysis step, a brief overview of the meta-data information is given in an aggregated manner in table 1. The given summary is following the given characteristics in Wittink et al. (1994). The given results showing w.r.t. the application context a similar result, as given by Green et al. (1981), Louviere/Woodworth (1981), and Green and Srinivasan (1990). W.r.t. the purpose, the results showing a similar result as given in Cattin and Wittink (1982), Wittink and Cattin (1989), Wittink et al. (1994), and Sattler and Hartmann (2008). Therefore, with these given distributions and results, it could be stated, that the upcoming results here probably can be transfered to other international CAs as well. Even if the database here is just taken from the german market. Furthermore, it should be outlined, that the data examination shows an meaningful result regarding the usage of incenvices and the usage of CASI over time. Both usages increased highly significant over time ($p < 0.001$).

4 Validity and Variance Analysis of Recent Commercial CAs

For the upcoming data analysis it has to be repeated here, that all data sets are collected over a time period of 16 years. Due to technological and method-

Table 1. Characteristics of given database
Legend: ACA, Adaptive CA; CBC, Choice Based CA; HCA, Hierarchical CA; CVA, Conjoint Value Analysis; CAWI/CAPI, Computer-assisted (Web/Personal) Interview; Avg., Average

Category	**% of Usage**	**Purpose**	**% of Usage**
Consumer Goods (to expend)	29.35%	Product Optimization	52.41%
Consumer Goods (to utilize)	20.79%	Pricing	31.49%
Telco Services	17.40%	Drug Admission	6.31%
Financial Services	10.95%	New Product Development	3.15%
Medical Services	9.70%	Distribution	1.82%
Other Services	7.93%	Other	4.83%
Other	3.87%		
CA Method	**% of Usage**	**Avg. Duration**	**% of Usage**
CBC	89.10%	Up to 20 minutes	7.03%
ACA	5.73%	Up to 30 minutes	33.56%
HCA	3.20%	Up to 45 minutes	31.60%
CVA	1.96%	Up to 60 minutes	27.82%
		Up to 90 minutes	1.82%
Computer-aided collection type	**% of Usage**	**Usage of Incentives**	**% of Usage**
CAWI	66.65%	Incentives used	69.47%
CAPI	25.99%	No Incentives used	30.53%
other	7.36%		

ological development over time, the given results are not directly comparable (e.g., different estimation software versions, iteration amounts, etc.). Therefore, all recent commercial CAs were recalculated on the same computer, with the same software base and with the same estimation parameters ($10k$ burn-in and estimation iterations, prior variance of 2, 5 degrees of freedom and an acceptance rate of 35%) to create a homogeneous database.

As mentioned in section 2, F-Statistics and linear regression models where used to analyze the internal validity values. Given p-values indicating the significance levels and the regression coefficent b will be used to indicate the validity development of the specific dataset over time (positive b-values indicating a gain, negative values the opposite). The regression model here is given through the usual regression formula:

$$\hat{y_i} = b \cdot x_i + \alpha$$

The first investigation is about the CBC and ACA approaches in general. Table 2 shows the results in an appropiate manner.

Given results showing no significant validity gain over time rather the opposite is shown over the whole time period. A more deep view on further

Table 2. Summary of Validity Development Over Time for Different Time Frames. *Legend: E, Exponent*

Time Frame	ACA	CBC
1996-2011	$N = 119; b = -0.035; p < 0,01$	$N = 1772; b = -0,014; p < 0,001$
2002-2011	$N = 28$; data set to small	$N = 1683; b = -0,007; p < 0,01$
2006-2011	$N = 8$; data set to small	$N = 1382; b = 0,003644; p > 0,1$

given validity values is inline with the given results above (see table 3) – No validity gain over time.

Table 3. Internal and External Validity Development in CA *Legend: E, Exponent*

CA Approach	Validity Criterion	Result
Time frame: $1996 - 2002$		
ACA	R^2 $(N = 89)$	$b = -2.2E^{-6}$; $p > 0.1$
	MAE $(N = 92)$	$b = 1.8E^{-4}$; $p > 0.1$
	$FCHR$ $(N = 92)$	$b = -6.7E^{-4}$; $p > 0.1$
CBC	R^2 $(N = 88)$	$b = -3.7E^{-6}$; $p < 0.1$
	MAE $(N = 140)$	$b = 4.7E^{-4}$; $p > 0.1$
	$FCHR$ $(N = 140)$	$b = 1.8E^{-5}$; $p > 0.1$
Time frame: $2003 - 2011$		
ACA	Dataset too small, for research implications $(N = 28)$	
CBC	RLH $(N = 1770)$	$b = 4.5E^{-4}$; $p < 0.1$
	MAE $(N = 1719)$	$b = 1.1E^{-3}$; $p < 0.001$
	$FCHR$ $(N = 1719)$	$b = -4.6E^{-3}$; $p < 0.001$

Therefore, hypothesis H_1 has to be rejected. No significant increase of CA' validity values over time could be detected. In fact, the opposite development was found for some validity values. To complete these findings, a variance analysis was applied to check for result dispersons within the data set. On this, all RLH values got regressed and tested for homo- and heteroscadisticity by using GQT and (studentized) BPT. The statistical test and p-values are summarized in table 4. Compared to the results above, they are showing a consistent result. The validity variance of CAs over time was not decreased. Significant p-values there indicating a heteroscedasticity, what has to be interpreted as an variance increase over time. Non significant values indicating homoscedasticity, what has to be interpreted as constant variance over time. The CBC validity values $(1996 - 2011)$ from table 4 were also plotted in figure 1 and using the example of the FCHR. A visual analysis of figure's shape shows the increased variance. Hypothesis H_2 could not be accepted either. A validity gain based on validity variance could not be detected and therefore, support for both hypotheses could not be found.

Table 4. Validity variance values and Variance Development in CA
Legend: E, Exponent; BPT, Breusch-Pagan-Test; GQT, Goldfeldt-Quandt-Test; BP, Breusch-Pagan's value of test statistic; GQ, Goldfeld-Quandt's value of test statistic; VC, Validity Criterion *$p < 0,1$; **$p < 0,05$; ***$p < 0,01$; ****$p < 0,001$

CA Approach	VC	studentized BPT	BPT	GQT
Time frame: 1996 – 2011				
ACA ($N = 119$)	RLH	$BP = 0.085$	$BP = 0.252$	$GQ = 1.334$
	MAE	$BP = 3.59$*	$BP = 1.81$	$GQ = 0.97$
	$FCHR$	$BP = 0.202$	$BP = 0.087$	$GQ = 1.102$
CBC ($N = 1772$)	RLH	$BP = 8.003$***	$BP = 3.854$**	$GQ = 0.93$
($N = 1859$)	MAE	$BP = 4.353$**	$BP = 2.178$	$GQ = 0.889$
($N = 1859$)	$FCHR$	$BP = 7.33$***	$BP = 3.757$*	$GQ = 1.16$**
Time frame: 2003 – 2011				
ACA	Dataset too low, for research implications ($N = 41$)			
CBC ($N = 1683$)	RLH	$BP = 0.238$***	$BP = 0.108$**	$GQ = 0.967$
($N = 1719$)	MAE	$BP = 3.253$*	$BP = 1.6$	$GQ = 0.932$
($N = 1719$)	$FCHR$	$BP = 12.921$****	$BP = 6.521$**	$GQ = 1.134$**
Time frame: 2006 – 2011				
CBC ($N = 1382$)	RLH	$BP = 0.193$***	$BP = 0.085$**	$GQ = 0.957$
	MAE	$BP = 6.694$***	$BP = 3.287$*	$GQ = 0.905$
	$FCHR$	$BP = 5.415$**	$BP = 2.802$*	$GQ = 1.057$

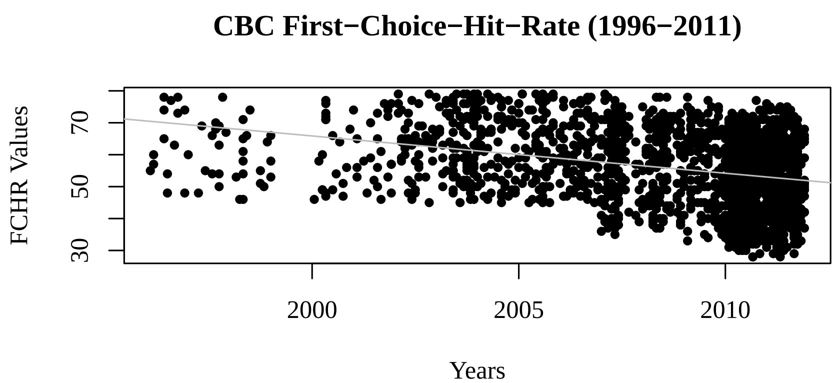

Fig. 1. Graphical Representation of CBC's First-Choice-Hit-Rates over Time

5 Conclusion and Outlook

The given analysis here has not proven the intuitiv presumption of a validity gain over time. Neither the internal and external validity values nor the validity variance of more than 2000 analyzed commercial CAs from the last 16 years have shown support for the presumptions and the derived hypotheses. Even modern technological and methodological developments in the past seeming to have no positive effects on CA' validity values. Maybe the mentioned negative impacts in section 1 are cutting through, which means, that the negative impacts of modern internet on respondent's distraction are compensating the potential positive impact of such new modern approaches and technological enhancements. This should be tested and investigated in future.

References

ADDELMAN, S. (1962): Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments. *Technometrics, 4, 21–46.*

ALLENBY, G. M. and ARORA, N. and GINTER, J. L. (1995): Incorporating Prior Knowledge into the Analysis of Conjoint Analysis. *Journal of Marketing Research, 32, 152–162.*

CATTIN, P. and WITTINK, D. R. (1982): Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing, 46, 44–53.*

CHEN, Y.-C. and SHANG, R.-A. and KAO, C.-Y. (2009): The Effects of Information Overload on Consumers' Subjective State Towards Buying Decision in the Internet Shopping Environment *Electronic Commerce Research and Applications, Elsevier Science Publishers B. V., 8, 48–58.*

DAY, G. S. and MONTGOMERY, D. B. (1983): Diagnosing the Experience Curve. *Journal of Marketing, 47, 44-58.*

DESARBO, W. and FONT, D. and LIECHTY, J. and COUPLAND, J. (2005): Evolutionary Preference/Utility Functions: A Dynamic Perspective. *Psychometrika, 70, 179–202.*

GREEN, P. E. and CARROLL, J. D. and GOLDBERG, S. M. (1981): A General Approach to Product Design Optimization via Conjoint Analysis. *Journal of Marketing, 45, 17–37.*

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

GREEN, P. E. and KRIEGER, A. M. and WIND, Y. (2001): Thirty Years of Conjoint Analysis: Reflections and Prospects. *Interfaces, INFORMS, 31, 56–73.*

GREEN, P. E. and RAO, V. R. (1971): Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research, 8, 355–363.*

GREEN, P. E. and SRINIVASAN, V (1990): Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *The Journal of Marketing, 54, 3–19.*

JACOBY, J. (1984): Perspectives on Information Overload. *Journal of Consumer Research, 4, 432–435.*

JOHNSON, R. M. (1987): Adaptive Conjoint Analysis. *Conference Proceedings of Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, 253–265.*

JOHNSON, R. M. and ORME, B. K. (2007): A New Approach to Adaptive CBC. *Sawtooth Software Inc., 1–28.*

LANDETA, J. (2006): Current Validity of the Delphi Method in Social Sciences. *Technological Forecasting and Social Change, 73, 467–482.*

LOUVIERE, J. J. and WOODWORTH, G. (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research, 20, 350–367.*

MCCULLOUGH, J. and BEST, R. (1979): Conjoint Measurement: Temporal Stability and Structural Reliability. *Journal of Marketing Research, 16, 26–31.*

MEISSNER, M. and DECKER, R. Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research, 52, 591–610.*

NETZER, O. and TOUBIA, O. and BRADLOW, E. and DAHAN, E. and EVGENIOU, T. and FEINBERG, F. and FEIT, E. and HUI, S. and JOHNSON, J. and LIECHTY, J. and ORLIN, J. and RAO, V. (2008): Beyond Conjoint Analysis: Advances in Preference Measurement. *Marketing Letters, 19, 337–354.*

PELZ, J.R. (2012): *Aussagefähigkeit und -willigkeit aus Sicht der Informationsverarbeitungstheorie Aussagefähigkeit und Aussagewilligkeit von Probanden bei der Conjoint-Analyse.* Gabler, Wiesbaden.

SÄNN, A. and BAIER, D. (2012): Lead User Identification in Conjoint Analysis Based Product Design. In: Gaul, W. A., Geyer-Schulz, A., Schmidt-Thieme, L., and Kunze, J. (Eds.): *Studies in Classification, Data Analysis, and Knowledge Organization, Challenges at the Interface of Data Analysis, Computer Science, and Optimization.*, Springer, Berlin Heidelberg, 43, 521–528.

SATTLER, H. and HARTMANN, A. (2008): Commercial Use of Conjoint Analysis. In: M. HÖCK, and K.I. Voigt, (Eds.): *Operations Management in Theorie und Praxis.* Gabler, Wiesbaden, 103–119.

SELKA, S. and BAIER, D. and BRUSCH, M. (2012): Improving the Validity of Conjoint Analysis by Additional Data Collection and Analysis Steps. In: Gaul, W. A., Geyer-Schulz, A., Schmidt-Thieme, L., and Kunze, J. (Eds.): *Studies in Classification, Data Analysis, and Knowledge Organization, Challenges at the Interface of Data Analysis, Computer Science, and Optimization.*, Springer, Berlin Heidelberg, 43, 529–536.

TOUBIA, O. and DE JONG, M.G. and STIEGER, D. and FÜLLER, J. (2012): Measuring Consumer Preferences Using Conjoint Poker. *Marketing Science, 31, 138–156.*

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

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