Report on the Second Working Group Meeting of the "AG Marketing"

Friederike Paetz and Daniel Guhl

Abstract In this article, we report on the second working group meeting of the "AG Marketing" within the GfKl Data Science Society. The meeting was held online on August 17 and 18, 2020. The presented topics reflect ongoing trends of using innovative methods and models for preference measurement as well as new data sources and machine learning approaches in quantitative marketing.

Friederike Paetz
Technical University Clausthal, Julius-Albert-Str. 6, D-38678 Clausthal-Zellerfeld Germany
✉ friederike.paetz@tu-clausthal.de

Daniel Guhl
Humboldt University of Berlin, Spandauer Str. 1, D-10178 Berlin Germany
✉ daniel.guhl@hu-berlin.de

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST)
KIT SCIENTIFIC PUBLISHING
Vol. 7, No. 1, 2020

DOI: 10.5445/IR/1000123957
ISSN 2363-9881

Introduction

Because of the COVID-19 pandemic, we are experiencing unusual times and scientific exchange is currently being hampered in particular by the cancellation of numerous conferences. As organizers of the working group AG Marketing (PD Dr. Friederike Paetz, Clausthal University of Technology and Prof. Dr. Daniel Guhl, Humboldt University Berlin), we want to counteract this and continue the scientific dialogue. Therefore, we decided to hold the second working group meeting of the "AG Marketing" as an online conference!

We are delighted to report that despite the "spatial distance" between participants, the meeting was a success with exciting presentations and stimulating discussions. It was good to see familiar and new faces and to chat about quantitative marketing research.

Quantitative marketing research is of high importance for marketing academics and practitioners. Developing and applying sophisticated data science tools is relevant for marketing academics. Using these tools to understand consumer behavior, to improve customer segmentation, and to make profitable marketing decisions is crucial for business success. Eight presentations (four sessions with two talks each) were given that cover different fields of quantitative marketing.

The first session was on machine learning approaches to enhance consumer insights: Peter Kurz talked about how machine learning techniques and Artificial Neural Networks (ANN) can be used to improve experimental designs in choice-based conjoint analysis. Using several artificial data sets, he explored to which extent ANNs can generate ideal experimental designs. Furthermore, he challenged an experimental design based on the complete enumeration method with an ANN-generated design on a real data set (Kurz, P.). Nadine Schröder used machine learning approaches, e.g., Latent Dirichlet Allocation (LDA), as a text mining approach, to improve the understanding of customer reviews and to predict choices. She used several brand reviews from amazon and found that topics that are considered helpful for one brand do not necessarily affect the helpfulness for another brand. Based on this, she derived managerial implications (Schröder, N.).

The second session focused on discrete choice models. Narine Yegoryan presented how models accounting for attribute non-attendance (ANA) outperform popular models like the Mixed Multinomial Logit model or the Mixture-of-Normals Multinomial Logit model. Using several empirical data sets, she presented differences in the unobserved preference distributions of the models and outlined in which cases it is crucial to account for ANA (Yegoryan, N., Guhl, D., and Paetz, F.). Friederike Paetz challenged two types of product line optimization approaches, which employ preference data and Hierarchical Bayesian Multinomial Logit models. Using several artificial data sets, she outlined determinants for the performance of both different types of product line optimization approaches, e.g., degree of preference heterogeneity (Paetz, F., Steiner, W., and Hruschka, H.).

The research of the third session applied ideas from behavioral economics. Vlada Pleshcheva investigated how the framing of information on vehicles' environmental impact affects consumers' preferences for identical car quality improvements. Using data from a choice experiment, she recovered the distributions of consumer preferences by applying a Mixed Multinomial Logit model and showed that consumers fail to recognize how transport-related CO_2 emissions translate into higher financial costs and cause greater environmental costs (Pleshcheva, V.). Ossama Elshiewy proposed a brand choice model applied to real purchase data. The model allows consumers to make choices based on both internal and external reference prices as well as an interaction between these two reference price concepts by accommodating asymmetric reference price response, purchase incidence, and consumer response heterogeneity. He showed that both response types are identified in one model and that losses from external reference prices interact with both gains and losses from internal reference prices (Elshiewy, O. and Peschel, A. O.).

The fourth session dealt with improvements in the validity of conjoint experiments: Benedikt Brand reported on the validity of best-worst scaling (BWS) methods by employing multiple criteria. Using an empirical example, he verified BWS's high internal and external validity (Brand, B. M. and Kopplin, C. S.). Marcel Lichters gave a talk on the advantages of adaptive designs versus incentive alignment in choice-based conjoint analysis. Using multiple experimental studies, he highlighted the superior predictive validity of the incentive-aligned adaptive choice-based conjoint analysis. Even though this approach has the highest absolute costs in marketing practice, the relative costs for an improvement of 1 % in predicted hit-rates above chance level are lower compared to other approaches (Sablotny-Wackershauser, V., Lichters, M., Guhl, D., and Vogt, B.). Below we provide selected abstracts of the contributions that were presented at the second working group meeting of the AG Marketing:

1 Conjoint Meets AI
Peter Kurz

Background on Artificial Neural Network

In the past few decades, Artificial Neural Networks (ANNs) have been used to identify and model choice behavior in a wide variety of fields. To give some examples from the field of market research, ANNs have been applied to model price elasticities in fast moving consumer goods area and car ownership (e.g., Hensher and Ton, 2000). ANNs aim to efficiently recognize patterns in the data, without being explicitly programmed where to look. A key feature of ANNs lies in their capability to approximate any Data Generating Process (DGP), provided that sufficient processing units are available; this feature is known as the Universal Approximation Theorem (Hornik et al., 1989). However, despite the strong pragmatic appeal of ANNs, they have been criticized for being too much data driven and theory poor, in effect presenting the analyst with a black box-model of the DGP. This limitation has hampered their use by discrete choice modelers and market researchers. Whereas many researches in the last years worked on using ANNs to model the choice behavior, we don't know of actual papers using ANNs to generate Experimental Designs for Choice Models.

The challenge of creating optimal Experimental Designs

In day-to-day research work client studies get more and more demanding. The number of attributes and levels are constantly increasing, and sample sizes get even smaller. Therefore, in many cases it is not easy to find sufficient experimental designs. Studies with a large number of attributes (and therefore hundreds of parameters to estimate) combined with the necessary restrictions and prohibitions on the attribute level (that cannot be shown together) often brings the established algorithms to their limits. Furthermore, most of the experimental designs used in day-to-day research are developed to estimate only aggregate models of choice behavior.

The Power of Artificial Neural Networks Creating Experimental Designs

The aim of an ANN based design generation is to find a perfect design, considering the above-mentioned problems and minimize the statistical- and measurement error. The goal is to find a solution where all estimated values are equal to "0" when all answers are perfectly random. On the one side we know the answers (simply random figures) and on the other we know which attribute-level combinations we could show. Therefore, it is relatively easy to generate a large number of synthetic datasets to train ANN's. After a long enough training period the selected ANN can find nearly optimal solutions, even for very complex experimental designs. Using hundreds of synthetic datasets, we explore to what extent ANNs are able to generate ideal experimental designs when the underlying DGP is known to the analyst. We focus on standard criteria for good experimental designs like orthogonality, level balanced overlap and utility balance (see Huber, Zwerina 1996). Additionally, we will present first results from a real dataset using a split design: Choice tasks based on experimental design generated with the complete enumeration algorithm versus ANN generated choice tasks.

2 Did You Find This Content Helpful? Linking Brand Specific Review Contents to Helpfulness of a Product Review *Nadine Schröder*

Before making a purchase, many customers consult product reviews to get information on the product experience. As a way to structure the vast amount of reviews, platforms make use of the helpfulness function. Consumers who considered a certain review as helpful may vote accordingly. Consequently, a lot of studies have addressed what review characteristics influence the number of helpful votes. Interestingly, even though a survey among customers shows that information on product performance or consumer satisfaction is considered as helpful, studies related to review helpfulness have focused on effects of, e.g., star rating or reviewer characteristics as drivers of helpfulness. In fact, only a subgroup (e.g., Cao et al (2011)) has considered content related review aspects. In this regard, these studies mainly focused on readability or sentiments of product reviews. Some studies even use a text mining approach but do not investigate the resulting contents. In fact, to the best of our knowledge, no previous study has comprehensively analyzed which particular review topics are helpful for future customers when making their purchase decision. We extend prior research by using the Latent Dirichlet Allocation (LDA) as a text mining approach (Griffiths and Steyvers (2004)). The LDA allows us to identify review topics that are interpretable and do not depend on the identification of topic categories beforehand. In a second step, these topics serve as predictors in various types of count models (Zeileis et al (2014)) to assess the helpfulness of a review. We use reviews for four major laptop brands that were collected on amazon. We find that topics which are considered helpful for one brand not necessarily have an effect on helpfulness for another brand. Marketers may benefit from knowing helpful topics in different ways. In particular, they may adjust their product description or even future product development. Reviews with helpful topics might also be displayed more prominently.

3 Determinants for the Recovery of Product Lines' Revenues

Friederike Paetz, Winfried Steiner, Harald Hruschka

Optimal product line design is a challenging task for marketing managers, as managers have to take into account preference heterogeneity of potential consumers. Product line design approaches that are explicitly based on consumer preferences have proven their advantages compared to other design approaches. Nowadays, consumer preferences can be efficiently measured via conjoint-analytic approaches like conjoint choice analysis. The results of such conjoint studies, i.e., individual part-worth utility estimates, build the input for product line optimization tools.

In conjoint approaches, consumer preferences are determined for pre-specified attributes and attribute levels, and several factors affect the precise estimation of these preferences, e.g., the degree of preference heterogeneity. Predicted revenues for product offerings strongly depend on how good the true preference structure of consumer is recovered by the conjoint model, because estimated utility structures serve as an input for product line design tools which search for a promising or an optimal product line design solution. Companies should therefore be interested in the robustness of approaches used for optimal product line design. Here, both the underestimation and the overestimation of revenues is undesirable.

In a Monte Carlo study, we compare absolute differences between predicted revenues based on true part-worth utilities and predicted revenues based on (re-)estimated part-worth utilities. For the determination of revenues, we used the SMRT module of Sawtooth Software with a genetic algorithm as search method (see, e.g., Steiner and Hruschka (2002), P.V. and Jacob (1996)).

We compare different scenarios that vary in several experimental factors associated with the degree of the underlying preference heterogeneity. For all scenarios, optimal product lines are determined by using two different optimization approaches that differ in their consideration of the degree of preference heterogeneity. While the first approach combines the single best segment-specific product solutions to a product line, the second approach simultaneously determines an optimal product line for the entire market. The first approach is computationally faster than the second approach and primarily applied in practice.

We find that the recovery of true preferences measured by the correlation between true and re-estimated preference structures is significantly affected by the underlying degree of heterogeneity. The heterogeneity factors, e.g., separation between segments or inner-segment heterogeneity, show the same significant impact on the absolute difference between the product line revenues calculated from the true versus the re-estimated preferences in both optimization approaches. However, the simultaneous product line approach proved to be significantly more robust to biases in the input data, i.e., mis-specified part-worth utility estimates, and leads to more precise predictions of revenues. As a recommendation for companies, we suggest that marketing managers should rely on the more complex, i.e., computational more sophisticated, simultaneous optimization approach to obtain accurate predictions of product line revenues.

4 Metric and Scale Effects in Consumer Preferences for Environmental Benefits
Vlada Pleshcheva

The present study investigates how the framing of information on the environmental impact of vehicles affects consumers' preferences for identical improvements in car quality. In particular, the effects of two metrics (fuel consumption vs. CO_2 emissions) and three scales of one metric (CO_2 in kg/km vs. g/km vs. g/100 km) are examined.

For a rational agent, the presentation of fuel consumption (FC) and CO_2 to assess personal fuel costs and the environmental impact of a car is redundant because each metric presents a "translation" of the same underlying information (Ungemach et al. 2017). First, from a technical perspective, FC and CO_2 emissions are linearly connected by a constant factor and are thus isomorphic in describing the environmental friendliness of a car. Second, rescaling identical information should not change consumer decisions. However, as this study demonstrates, the type of information presented to consumers significantly affects the valuation of fuel savings and environmental benefits from a reduction in FC versus CO_2.

The research goal relates to the broad literature on how the framing of information affects consumers' decisions (Tversky & Kahneman 1981). A number of empirical studies have demonstrated that contextual features associated with a decision affect consumers' preferences and choices, sometimes resulting in preference reversal (Thaler et al. 2013). The current study's contribution lies in quantifying the differences in consumers' preferences for two measures of the same information that have not been previously directly compared. Although consumers' preferences for a reduction in FC and CO_2 emissions of cars are extremely important in the context of environmental policies, no prior work has directly compared consumers' preferences for them. Prior research on revealed preferences has not been able to separately identify these effects because the metrics are perfectly correlated, and research on stated preferences has either focused on one of these environmentally important attributes or considered both measures simultaneously and thus did not disentangle the separate effects of each metric.

The present study recovers the distributions of consumer preferences for FC and CO_2 independently based on consumer choices from optimally designed choice experiments and by applying a mixed (random coefficient) logit model. The estimation accounts for consumers’ unobserved heterogeneity in tastes for car attributes in addition to the observed heterogeneity in the respondents’ socio-demographic characteristics, car use experience, environmental attitudes, and knowledge.

The findings suggest that individuals fail to recognize how transport-related CO_2 emissions translate into ‘private’ costs and ultimately incur higher financial costs and cause greater environmental costs. The biases persist even when the environmentally friendly product is also cost-minimizing. The insights of this study serve to guide policymakers and car manufacturers on how to present information on car offers.

5 Examining Best-Worst Scaling’s Validity and Reliability: Worth a Try? *Benedikt Martin Brand, Cristopher Siegfried Kopplin*

As surveys employing (Likert) scale items suffer from several shortcomings, such as difficulties in interpreting rating score data, varying validity and reliability of items and constructs, and omitted reference domains for items, Finn and Louviere (1992) introduced the Best-Worst Scaling (BWS) attempting to overcome these shortcomings. This comparably novel methodology was developed by Louviere and Woodworth in 1990 as an extension of Thurstone’s paired comparison approach. As part of discrete choice modeling, respondents answering BWS surveys need to determine their best and worst item within a choice set over multiple rounds.

Even though this rather nascent method provides a couple of advantages, such as acquiring additional information about the worst choice, providing distinct demarcation between similar items, enabling inter-attribute comparisons, solving biases inherent to rating scales, and overcoming cultural response biases (Auger et al., 2007), it also contains some limitations. Due to the design algorithm generating multiple BWS constellations according to common choice design criteria (frequency balance, level balance, orthogonality, positional balance) in combination with selecting two items per choice set, difficulties arise in assessing BWS’ validity and reliability. Thus, many questions about BWS’ validity and reliability remain unanswered (Mi et al., 2019). Besides, applying

BWS in its initial composition only reveals the utilities of items relative to each other. Consequently, the items' absolute importance or effectiveness cannot be derived. Therefore, we contribute to current research by overcoming the before-mentioned limitations of BWS and by examining BWS' validity and reliability employing multiple criteria based on an empirical example. Hence, we analyze BWS' internal and external validity, focusing on hit rates, mean absolute error and root mean square error, its internal reliability in the form of test-retest reliability, and apply cross-validation using ranking tasks. Moreover, we evince possibilities for anchor scaling to reveal not only relative utilities but also absolute evaluation. Based on an empirical example dealing with effective measures to reduce product returns and thereby reduce the related negative environmental impact, consumers (n=288) were asked to evaluate 13 items. Results yielded high hit rates, very low mean absolute errors and root mean square errors, verifying BWS' high internal validity. Moreover, criteria scrutinizing internal reliability demonstrate a high consistency, especially for the chosen worst items. Regarding predictive validity, the BWS choices were forecasted moderately precise based on random subsample draws and with a varying amount of respondents used for test vs. training data categorization. Here, the selected best items were predicted more often correct compared with the worst items.

References

Auger P, Devinney T, Louviere J (2007) Using best–worst scaling methodology to investigate consumer ethical beliefs across countries. Journal of Business Ethics 70(3):299–326. DOI: 10.1007/s10551-006-9112-7.

Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. Decision Support Systems 50:511–521. DOI: 10.1016/j.dss.2010.11.009.

Finn A, Louviere J (1992) Determining the appropriate response to evidence of public concern: The case of food safety. Journal of Public Policy & Marketing 11(2):12–25. DOI: 10.1177/074391569201100202.

Gilbride T, Allenby G, Brazell J (2006) Models for heterogenous variable selection. Journal of Marketing Research 43(3):420–430. DOI: 10.1509/jmkr.43.3.420.

Griffiths T, Steyvers M (2004) Finding scientific topics. Proceedings of the National Academy of Sciences 101:5228–5235. DOI: 10.1073/pnas.0307752101.

Hensher D, Ton T (2000) A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transportation Research Part E 36(3):155–172. DOI: 10.1016/S1366-5545(99)00030-7.

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2(5):359–366. DOI: 10.1016/0893-6080(89)90020-8.

Huber J, Zwerina K (1996) The importance of utility balance and efficient choice designs. Journal of Marketing Research 33(3):307–317. DOI: 10.1177/002224379603300305.

Mi X, Tang M, Liao H, Shen W, Lev B (2019) The state-of-the-art survey on integrations and applications of the best worst method in decision making: Why, what, what for and what’s next? Omega 87:205–225. DOI: 10.1016/j.omega.2019.01.009.

P.V. B, Jacob V (1996) Genetic algorithms for product design. Management Science 42:1105–1117. DOI: 10.1287/mnsc.42.8.1105.

Steiner W, Hruschka H (2002) Produktliniengestaltung mit Genetischen Algorithmen. Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (ZfbF) 54(7):575–601. DOI: 10.1007/BF03372688.

Thaler R, Sunstein C, Balz J (2013) Choice Architecture. In: The Behavioral Foundations of Public Policy, Shafir E (ed), Princton University Press, Princeton, pp. 428–439. ISBN: 978-0-691137-56-8.

Tversky A, Kahneman D (1981) The Framing of Decisions and the Psychology of Choice. Science 211(4481):453–458. DOI: 10.1126/science.7455683.

Ungemach C, Camilleri A, Johnson E, Larrick R, Weber E (2017) Translated Attributes as Choice Architecture: Aligning Objectives and Choices Through Decision Signposts. Management Science 64(5):2445–2459. DOI: 10.1287/mnsc.2016.2703.

Yegoryan N, Guhl D, Klapper D (2020) Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. Journal of Business Research 111:290–304. DOI: 10.1016/j.jbusres.2019.01.061.

Zeileis A, Kleiber C, Jackman S (2014) Regression models for count data in r. Journal of Statistical Software 27:1–25. URL: `https://cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf`.