

Eine Evaluation experimenteller Designs und deren Einfluss auf die Schätzergebnisse von Choice Based Conjoint Analysen mittels Monte Carlo Simulation

Nico Lehmann

Institut für Industriebetriebslehre und Industrielle Produktion (IIP)
Karlsruher Institut für Technologie (KIT)

Abstract

Das Ziel der statistischen Versuchsplanung bei Choice Experimenten ist es, möglichst effiziente Designs zu erstellen. Im Rahmen dieser Arbeit werden Methoden zur Erstellung eines (3,4,3,4)-Designs mittels Monte Carlo Simulation miteinander verglichen. Es zeigt sich, dass bei allen Methoden die Haupteffekte geschätzt werden können und sich die geschätzten Parameter den wahren Parametern mit steigendem Stichprobenumfang nähern. Die relativen Unterschiede in Form des Root Mean Square Error (RMSE) sind vergleichsweise hoch, können jedoch in ihrer absoluten Höhe nachrangig behandelt werden. Sollen Haupteffekte und Zwei-Wege-Interaktionseffekte geschätzt werden, so sind diese Effekte nicht bei allen Methoden zur Designerstellung schätzbar. Sowohl die absoluten Werte des RMSE sind höher gegenüber einer ausschließlichen Haupteffektschätzung als auch die relativen Unterschiede zwischen den Designs. Eine steigende Anzahl an Blöcken hat allgemein einen positiven Einfluss auf die statistische Effizienz und den RMSE. Der marginale Einfluss sinkt jedoch mit steigender Anzahl, sodass er v.a. bei sehr wenigen Blöcken von Bedeutung ist. Zudem ist der Einfluss geringer als der eines steigenden Stichprobenumfangs. Die durch die Methode der Designerstellung oder durch eine geringe Anzahl an Blöcken auftretenden Ineffizienzen können durch einen erhöhten Stichprobenumfang kompensiert werden und sind daher insbesondere bei beschränkten Stichprobengrößen von Relevanz.

1 Einleitung

Choice Experimente stellen eine indirekte Befragungsmethode dar, um Präferenzen hinsichtlich kommerzieller und nicht-kommerzieller Güter zu erheben. Dabei werden den Probanden jeweils mehrere Alternativen (Stimuli) mit differierenden Attributen in sog. Choice Sets vorgelegt, aus denen sie eine Alternative als die bevorzugte auswählen können [1]. Die Anzahl an Attributen und deren Ausprägungen (Level), die Anzahl an Alternativen pro Choice Set, die Anzahl an Choice Sets pro Proband und die Erstellung der Choice Sets sind Teil der statistischen Versuchsplanung von Choice Experimenten und sollten überlegt gewählt werden [2]. Häufig werden Experimente zuerst durchgeführt und danach ausgewertet, ohne im Vorfeld das Design statistisch zu überprüfen [3]. Das Ziel der Versuchsplanung bei Choice Experimenten ist es, statistisch effiziente Designs zu erstellen, um auf diese Weise möglichst viele Informationen aus den Antworten der Probanden zu gewinnen [1, 3, 4].

Im Rahmen dieser Arbeit werden Verfahren zur Erstellung eines Designs, welches zur Erhebung von Präferenzen hinsichtlich Stromtarifattributen genutzt werden soll, miteinander verglichen. Das Design umfasst die vier Attribute *Anbieter* (drei Level), *Strommix* (vier Level), *Anteil regionaler Erzeugung* (drei Level) und *Preisauflschlag* (vier Level), woraus ein (3,4,3,4)-Design resultiert. Für die numerischen Attribute *Anteil regionaler Erzeugung* und *Preisauflschlag* wird jeweils eine Ordinalskalierung angenommen, da Nichtlinearitäten in den Präferenzen der Probanden nicht ausgeschlossen werden können. Die Attribute *Anbieter* und *Strommix* sind nominalskaliert. Das vollfaktorielle Design umfasst somit $3 \times 4 \times 3 \times 4 = 144$ mögliche Stimuli, die bei drei Stimuli pro Choice Set in bis zu $144 \times 144 \times 144 = 2.985.985$ Kombinationen angeordnet werden können. Je mehr Stimuli pro Choice Set vorgelegt werden, desto statistisch effizienter ist ein Design [5]. Allerdings kann eine hohe Anzahl an Stimuli pro Choice Set zu Überforderungen der Probanden oder zu einem Antwortverhalten, welches Heuristiken nutzt, führen [6, 7]. Meist ist es nicht möglich, einem Probanden alle Stimuli vorzulegen, weshalb in der Praxis häufig auf teilfaktorielle Designs zurückgegriffen wird [8] und/oder die möglichen Choice Sets in Blöcke aufgeteilt werden, um auf diese Weise die Anzahl an Choice Sets pro Proband zu reduzieren [2, 9]. Eine Abkehr vom vollfaktoriellen Design geht jedoch mit Qualitätseinbußen bei den Parameterschätzern einher, weshalb in der Praxis eine Abwägung zwischen Durchführbarkeit und Qualität getroffen werden muss [2].

2 Hintergrund und Methodik

Die Auswahl eines Designs sollte spezifisch nach den zu beantwortenden Forschungsfragen getroffen werden [2]. Meist werden jedoch die Ziele der Orthogonalität und der Ausgeglichenheit angestrebt [9–11]. Die Orthogonalität stellt sicher, dass Parameterschätzungen im Modell unabhängig voneinander erfolgen. Bei ausgeglichenen Designs treten alle Attributsausprägungen gleich häufig auf, wodurch die Attribute die gleiche Trennschärfe (englisch: *statistical power*) besitzen [2] und zugleich unkorreliert mit dem Interzept sind [11, 12]. Letzteres setzt eine Effektkodierung bei Faktoren voraus [13–15].

Übliche Maße zur Bewertung der Effizienz von Designs basieren auf der Informationsmatrix ($X^T X$) [12, 16]. Jeder Stimulus innerhalb eines Choice Sets t wird durch eine Zeile der Matrix dargestellt, wobei sich X aus den Informationsmatrizen der einzelnen Auswahl-situationen X_t ergibt und jedem Probanden T Choice Sets vorgelegt werden. Die Anzahl aller Choice Sets ergibt sich durch Multiplikation von T mit der Anzahl an Probanden N . Die Varianz der P geschätzten Nutzenparameter $diag(\Sigma_p)$ ist proportional zur Informationsmatrix, sodass ein Design mit einer „kleineren“ Informationsmatrix genauere Parameterschätzungen erzeugt. [2, 16] Die Eigenwerte einer Matrix sind hierfür ein geeignetes Maß. Die D-Effizienz ist das in der Praxis meistverwendete Gütekriterium bei der Erstellung von Designs [9] und berechnet sich durch Minimierung des geometrischen Mittelwerts der Varianzen der Parameterschätzer [2]. In der Praxis wird für statistisch effiziente Designs häufig das Kriterium der Orthogonalität bis zu einem gewissen Grad aufgegeben, um die D-Effizienz zu maximieren [4].

$$D = |X^T X|^{1/P} \quad (1)$$

Die Definition in Formel (1) nimmt keine Normierung hinsichtlich der Anzahl an Probanden vor, was einen Vergleich der D-Effizienzen bei unterschiedlichen Stichprobengrößen N erlaubt [17]. Aus dem Verhältnis der D-Effizienzen zweier Designs lässt sich die relative D-Effizienz bestimmen [18]. Diese Definition wird im Folgenden verwendet. Die Inverse der Informationsmatrix $(X^T X)^{-1}$ ist proportional zur Kovarianzmatrix Σ_p des zu schätzenden Vektors der Nutzenparameter $\hat{\beta}_p$ [12, 16], weshalb die Kovarianzmatrix aus dem multinomialen logistischen (MNL) Regressionsmodell der N simulierten Probanden zur Berechnung der D-Effizienz verwendet werden kann. Mit diesem Ansatz kann die D-Effizienz für verschiedene, vorgegebene Nutzenfunktionen und -werte berechnet werden. Der Nutzen $u_{n,i}$ von Proband n für Stimulus i ergibt sich aus der Verknüpfungsfunktion Ψ , die sich aus einem deterministischen Nutzen v für das Level l von Attribut k und einer stochastischen Komponente ε zusammensetzt (vgl. Formel (2)) [8].

$$u_{n,i} = \Psi[v_{n,i}(x_{i,k,l}), \varepsilon_{n,i}] \quad (2)$$

Allen Probanden wird im Folgenden die gleiche Nutzenfunktion unterstellt, wobei keine Präferenzen hinsichtlich der Attributsausprägungen angenommen werden und die stochastische Komponente ε einer Gumbel-Verteilung (Extremwert Typ I) folgt (vgl. Formel (3)).

$$u_{n,i} = u_i = \varepsilon_i \quad (3)$$

Die auf dieser Basis bestimmte D-Effizienz wird mit D_0 bezeichnet. Neben einer Bestimmung der D-Effizienz können die wahren Nutzenparameter auf Basis der simulierten Probanden geschätzt werden. Je näher die geschätzten Nutzenparameter $\hat{\beta}_p$ an den wahren Nutzenparametern β_p von Null liegen, desto präziser erfolgt eine Nutzenschätzung [14]. Ein weiteres, mit der D-Effizienz stark zusammenhängendes Gütekriterium sind die Standardfehler der Schätzung. Je geringer diese sind, desto effizienter ist die Schätzung [19].

Insgesamt werden 19 Designs getestet, die mittels acht verschiedener Methoden erstellt werden. Diese umfassen die *Shifting*- [20, 21], die *Mix & Match 1*- [14, 21–23] und die *Mix & Match 2*-Methode [14, 22] sowie eine Methode nach *Cook & Nachtsheim* [24, 25]. Die vier verbleibenden Methoden namens *Random*, *Complete Enumeration*, *Shortcut* und *Balanced Overlap* stammen vom kommerziellen Anbieter *Sawtooth Software Inc.* und sind in [23, 26] beschrieben. *Shifting*, *Mix & Match 1* und *Mix & Match 2* verwenden als Basis zur Erstellung der Choice Sets normalerweise die Stimuli eines orthogonalen Haupteffektplans [16]. Da ein solcher für das vorliegende (3,4,3,4)-Design nicht gefunden werden konnte [27], werden für diese drei Methoden andere Versuchspläne als Ausgangsbasis verwendet. Für die ersten drei Designs mit je 144 Blöcken wird der vollfaktorielle Versuchsplan (VfV) mit 144 Stimuli als Ausgangsbasis verwendet. Außerdem werden mittels des Fedorov-Algorithmus [9] 48 der 144 Stimuli des vollfaktoriellen Versuchsplans ermittelt, die einen möglichst großen Beitrag zur D-Effizienz eines Versuchsplans leisten. Da die D-Effizienz abhängig vom unterstellten Modell ist [24], wird bei der Ermittlung der Stimuli mittels des Fedorov-Algorithmus sowohl ein Modell als Input vorgegeben, bei dem nur Haupteffekte (HE) angenommen werden, als auch eines mit allen Zwei-Wege-Interaktionseffekten (ZWIE). Diese je 48 Stimuli werden als Basis zur Erstellung der Choice Sets mittels *Shifting*, *Mix & Match 1* und *Mix & Match 2* verwendet, woraus sechs weitere Designs mit je vier Blöcken resultieren. Die fünf verbleibenden Methoden verwenden als Ausgangsbasis den vollfaktoriellen Versuchsplan, unterscheiden sich jedoch mit vier bzw. 144 Blöcken und ergeben somit zehn weitere Designs. Jeder Block eines Designs besteht aus zwölf Choice Sets mit je drei Stimuli und jeder simulierte Proband wird sukzessive einem Block zugewiesen. Um den Einfluss der 19 Designs auf die Schätzergebnisse zu bestimmen, werden Monte-Carlo-Simulationen [28] mit je 500 Wiederholungen für verschiedene Stichprobenumfänge N (25, 50, 300, 750) und geschätzte Anzahlen an Parametern $\hat{\beta}_p$ (HE, ZWIE) durchgeführt. Als Metrik für die Genauigkeit der Schätzergebnisse wird der Root Mean Squared Error (RMSE) verwendet [29] und anschließend der Median der 500 Wiederholungen ermittelt. Der RMSE berechnet sich aus der Wurzel des mittleren quadratischen Fehlers und ist damit abhängig vom Skalenniveau. Er kann jedoch im Gegensatz zu relativen Metriken auch bei wahren Parametern von Null verwendet werden [30].

3 Ergebnisse

In Abbildung 1 sind die relativen D_0 -Effizienzen für verschiedene (3,4,3,4)-Designs, basierend auf je 750 simulierten Probanden, dargestellt. Bei einer ausschließlichen Schätzung von HE hat eine Reduzierung der Blöcke von 144 auf 4 bei den Methoden selbst kaum eine negative Auswirkung auf die Design-Effizienz. Jedoch unterscheidet sich die Effizienz zwischen den Methoden um bis zu 32,4%. Werden zusätzlich zu den HE alle ZWIE geschätzt, so resultiert bei der *Shifting*-Methode ein singuläres Design. Bei den verbleibenden Methoden hat eine Reduzierung der Blöcke von 144 auf 4 mit bis zu 53,9% Unterscheid im Vergleich zum effizientesten Design erheblich negative Auswirkungen. Der Effekt kann jedoch abgemildert werden, wenn bei der Auswahl der Stimuli für *Mix & Match 1* und *Mix & Match 2* durch den Fedorov-Algorithmus (siehe Kapitel 2) ZWIE

explizit berücksichtigt werden. Der geringste absolute Effizienzeinbruch durch eine Reduzierung der Blöcke bei den ZWIE ist beim Verfahren nach *Cook & Nachtsheim* zu beobachten.

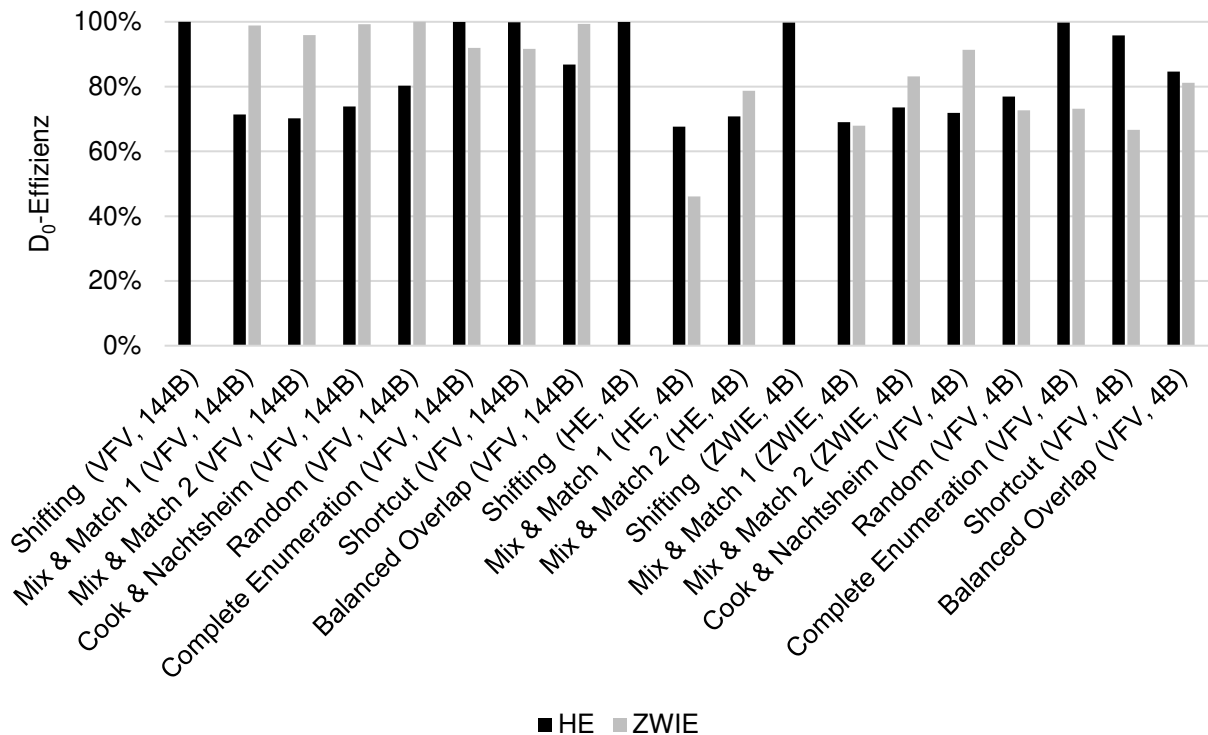


Abbildung 1: D_0 -Effizienz für verschiedene (3,4,3,4)-Designs unter Berücksichtigung von Haupteffekten (HE) und Zwei-Wege-Interaktionseffekten (ZWIE), basierend auf 750 simulierten Probanden.

Tabelle 1 gibt den RMSE für unterschiedliche Anzahlen simulierter Probanden N (25, 50, 300, 750) und Anzahlen geschätzter Parameter P (HE, ZWIE) wider. Die Aussage hinsichtlich der Anzahl an Blöcken bei der D_0 -Effizienz bleibt auch beim RMSE unverändert: Bei ausschließlicher Schätzung der HE hat eine Erhöhung der Anzahl an Blöcken bei Betrachtung jeweils der gleichen Designerstellungsmethode nahezu keine Auswirkungen. Wird die Betrachtung auf einen Vergleich zwischen allen 19 Designs erweitert, so liegen die Unterschiede beim RMSE bei bis zu 26,4%. In ihrer absoluten Höhe sind die Unterschiede jedoch vernachlässigbar gering. Bei einer Schätzung der HE und aller ZWIE ergibt sich ein anderes Bild: Die relativen Unterschiede zwischen den Designs beim RMSE sind mit bis zu 475,3% deutlich höher, wobei das Design *Mix & Match 1 (HE, 4B)* hier einen Extremwert darstellt. Ohne dieses liegen die maximalen Unterschiede bei 66,4%. Eine sinkende Anzahl an Blöcken erhöht den RMSE. Die negativen Auswirkungen bei *Mix & Match 1* und *Mix & Match 2* können auch beim RMSE reduziert werden, wenn beim Fedorov-Algorithmus ZWIE explizit berücksichtigt werden (siehe Kapitel 2). Obwohl eine steigende Anzahl an Blöcken den RMSE reduziert, ist der positive Effekt geringer als der eines steigenden Stichprobenumfangs. Ein erhöhter Stichprobenumfang lässt die relativen Unterschiede zwischen den Designs unverändert, wirkt jedoch stark auf die absolute Höhe des RMSE. Folglich kommen die Unterschiede beim RMSE zwischen den

Designs v.a. bei geringen Stichprobengrößen zum Tragen. Trotz der Unterschiede zwischen den Designs lässt sich anhand der Tabelle 1 erkennen, dass sich der RMSE bei steigendem Stichprobenumfang gegen Null nähert.

Tabelle 1: Median RMSE für verschiedene (3,4,3,4)-Designs über 500 Monte-Carlo-Wiederholungen.

Design		<u>Geschätzte Parameter und Anzahl simulierter Probanden</u>							
		HE		ZWIE		HE		ZWIE	
		25	50	300	750	25	50	300	750
144 Blöcke	Shifting (VFV, 144B)	0,044	-	0,030	-	0,012	-	0,008	-
	Mix & Match 1 (VFV, 144B)	0,050	0,183	0,036	0,123	0,014	0,048	0,009	0,030
	Mix & Match 2 (VFV, 144B)	0,052	0,188	0,037	0,127	0,015	0,049	0,009	0,031
	Cook & Nachtsheim (144B)	0,050	0,189	0,036	0,125	0,014	0,049	0,009	0,031
	Random (144B)	0,050	0,187	0,034	0,124	0,014	0,048	0,009	0,030
	Complete Enumeration (144B)	0,043	0,198	0,031	0,132	0,013	0,052	0,008	0,033
	Shortcut (144B)	0,043	0,201	0,030	0,134	0,012	0,053	0,008	0,033
	Balanced Overlap (144B)	0,047	0,187	0,033	0,126	0,013	0,050	0,008	0,031
4 Blöcke	Shifting (HE, 4B)	0,043	-	0,031	-	0,012	-	0,008	-
	Mix & Match 1 (HE, 4B)	0,053	1,051	0,037	0,665	0,015	0,247	0,010	0,160
	Mix & Match 2 (HE, 4B)	0,053	0,248	0,036	0,164	0,015	0,064	0,009	0,040
	Shifting (ZWIE, 4B)	0,042	-	0,030	-	0,012	-	0,008	-
	Mix & Match 1 (ZWIE, 4B)	0,053	0,304	0,036	0,197	0,015	0,078	0,009	0,049
	Mix & Match 2 (ZWIE, 4B)	0,052	0,229	0,036	0,151	0,014	0,059	0,009	0,038
	Cook & Nachtsheim (4B)	0,051	0,208	0,036	0,138	0,014	0,055	0,009	0,034
	Random (4B)	0,052	0,275	0,036	0,180	0,014	0,070	0,009	0,044
	Complete Enumeration (4B)	0,044	0,257	0,030	0,174	0,012	0,068	0,008	0,043
	Shortcut (4B)	0,045	0,293	0,031	0,186	0,013	0,074	0,008	0,046
Balanced Overlap (4B)	0,047	0,249	0,033	0,163	0,014	0,064	0,009	0,040	

4 Schlussfolgerungen

Für die Bewertung von Designs ist aus statistischer Sicht wichtig, dass das zu schätzende Modell identifizierbar und möglichst präzise (i.S.d. Erwartungstreue und Suffizienz) ist. Mit allen 19 betrachteten Designs können die HE geschätzt werden, jedoch erlauben zwei Designs keine zusätzliche Schätzung der ZWIE. Sind die gewünschten Parameter schätzbar, so reduziert ein erhöhter Stichprobenumfang den RMSE und die geschätzten Nutzenparameter nähern sich den wahren Nutzenparametern an. Auch die absolute D_0 -Effizienz steigt mit steigendem Stichprobenumfang, wobei die D_0 -Effizienz und der RMSE zwar stark, jedoch nicht perfekt negativ korrelieren. Vor der Durchführung eines Experiments im Feld sollte daher ein Design auf Identifizierbarkeit und Präzision getestet werden. Eine steigende Anzahl an Blöcken hat allgemein einen positiven Einfluss auf die statistische Effizienz und senkt den RMSE, der marginale Einfluss sinkt jedoch mit steigender Anzahl an Blöcken. Daher empfiehlt sich eine Anzahl an Blöcken kleiner oder gleich dem Stichprobenumfang. Zuletzt sei angemerkt, dass eine Erhöhung der statistischen Effizienz negative Auswirkungen auf die Antworteffizienz der Probanden haben kann. Diese sollte vor der Durchführung ebenfalls Berücksichtigung finden.

Literatur

- [1] S. Gensler, O. Hinz, B. Skiera und S. Theysohn, „Willingness-to-pay estimation with choice-based conjoint analysis: Addressing extreme response behavior with individually adapted designs“, *European Journal of Operational Research*, Jg. 219, Nr. 2, S. 368–378, 2012.
- [2] J. L. Lusk und F. B. Norwood, „Effect of Experimental Design on Choice-Based Conjoint Valuation Estimates“, *American Journal of Agricultural Economics*, Jg. 87, Nr. 3, S. 771–785, 2005, <http://www.jstor.org/stable/3697912>.
- [3] F. Triefenbach, „Design of Experiments: The D-Optimal Approach and Its Implementation As a Computer Algorithm“. Bachelor’s Thesis, Umea, 2008. [Online] Verfügbar unter: <http://www8.cs.umu.se/education/examina/Rapporter/FabianTriefenbach.pdf>. Zugriff am: 26. August 2019.
- [4] M. Ryan, J. R. Kolstad, P. C. Rockers und C. Dolea, *How to Conduct a DCE for Health Workforce Recruitment and Retention in Remote and Rural Areas.: a user guide with case studies*. Geneva: WHO Press, 2012.
- [5] J. Pinnell, „Comment on Huber: Practical Suggestions for CBC Studies“, Sawtooth Software Inc., Sequim, WA, RESEARCH PAPER SERIES, 2004. [Online] Verfügbar unter: <https://www.sawtoothsoftware.com/support/technical-papers/cbc-related-papers/the-benefits-of-accounting-for-respondent-heterogeneity-in-choice-modeling-1999>. Zugriff am: 30. September 2019.
- [6] Sawtooth Software Inc., „CBC Technical Paper“. Version 9, TECHNICAL PAPER SERIES, 2017. [Online] Verfügbar unter: <https://www.sawtoothsoftware.com/support/technical-papers/cbc-related-papers/cbc-technical-paper-2013>. Zugriff am: 30. September 2019.
- [7] S. Hess, J. M. Rose und J. Polak, „Non-trading, lexicographic and inconsistent behaviour in stated choice data“, *Transportation Research Part D: Transport and Environment*, Jg. 15, Nr. 7, S. 405–417, 2010.
- [8] S. Gensler, *Heterogenität in der Präferenzanalyse: Ein Vergleich von hierarchischen Bayes-Modellen und Finite-Mixture-Modellen*. Wiesbaden: Deutscher Universitäts-Verlag, 2003.
- [9] F. Reed Johnson *et al.*, „Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force“ (eng), *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, Jg. 16, Nr. 1, S. 3–13, 2013.
- [10] B. Orme, „Three Ways to Treat Overall Price in Conjoint Analysis“, Sequim, WA, RESEARCH PAPER SERIES, 2007. [Online] Verfügbar unter: <https://www.sawtoothsoftware.com/support/technical-papers/general-conjoint-analysis/three-ways-to-treat-overall-price-in-conjoint-analysis-2007>. Zugriff am: 13. Juni 2019.
- [11] W. F. Kuhfeld, *Marketing Research Methods in SAS*. [Online] Verfügbar unter: <http://support.sas.com/techsup/technote/mr2010.pdf>. Zugriff am: 29. April 2019.

- [12] W. F. Kuhfeld, „Efficient Experimental Designs Using Computerized Searches“, SAS Institute, Inc., Sequim, WA, RESEARCH PAPER SERIES, 1997. [Online] Verfügbar unter: <https://www.sawtoothsoftware.com/support/technical-papers/design-of-conjoint-experiments/efficient-experimental-designs-using-computerized-searches-1997>. Zugriff am: 27. Oktober 2019.
- [13] M. Bech und D. Gyrd-Hansen, „Effects coding in discrete choice experiments“ (eng), *Health economics*, Jg. 14, Nr. 10, S. 1079–1083, 2005.
- [14] J. J. Louviere, D. A. Hensher und J. D. Swait, *Stated choice methods*. New York: Cambridge University Press, 2000.
- [15] S. Gensler, „Ermittlung von Präferenzen für Produkteigenschaften mit Hilfe der Choice-Based Conjoint Analyse, Teil I“ (de), *WiSt - Wirtschaftswissenschaftliches Studium*, Jg. 35, Nr. 5, S. 254–258, 2006.
- [16] D. J. Street, L. Burgess und J. J. Louviere, „Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments“, *International Journal of Research in Marketing*, Jg. 22, Nr. 4, S. 459–470, 2005.
- [17] Sawtooth Software Inc., *Testing the CBC Design*. [Online] Verfügbar unter: https://www.sawtoothsoftware.com/help/lighthouse-studio/manual/hid_web_cbc_designs_6.html. Zugriff am: 13. Januar 2020.
- [18] W. F. Kuhfeld, R. D. Tobias und M. Garratt, „Efficient Experimental Design with Marketing Research Applications“, *Journal of Marketing Research*, Jg. 31, Nr. 4, S. 545, 1994.
- [19] M. Messer und G. Schneider, *Statistik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019.
- [20] D. Bunch, J. Louviere und D. Anderson, *A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes*, 1996.
- [21] H. Aizaki und K. Nishimura, „Design and Analysis of Choice Experiments Using R: A Brief Introduction“, *Agricultural Information Research*, Jg. 17, Nr. 2, S. 86–94, 2008, <https://stat.ethz.ch/pipermail/r-help/attachments/20100618/a8d2f366/attachment-0001.pdf>. Zugriff am: 19. März 2019.
- [22] Sawtooth Software Inc., *Experimental Designs for CBC*. Webinar. [Online] Verfügbar unter: http://www.sawtoothsoftware.com/webinar/sawtooth_software_expdesign_webinar.pdf. Zugriff am: 13. Januar 2020.
- [23] K. Chrzan und B. Orme, „An Overview and Comparison of Design Strategies for Choice-Based Conjoint Analysis“ in *Proceedings of the Sawtooth Software Conference*, Sawtooth Software Inc., Hg., Sequim, WA, 2000, S. 161–177.
- [24] R. D. Cook und C. J. Nachtsheim, „Computer-Aided Blocking of Factorial and Response-Surface Designs“, *Technometrics*, Jg. 31, Nr. 3, S. 339, 1989.
- [25] J. Horne, *Package ‘choiceDes’: Design Functions for Choice Studies*. The R project for statistical computing <https://www.r-project.org/>, 2018.
- [26] Sawtooth Software Inc., „The CBC Latent Class Technical Paper“, Sequim, WA, TECHNICAL PAPER SERIES, 2004. [Online] Verfügbar unter: <https://www.sawtoothsoftware.com/support/technical-papers/sawtooth-software-products/cbc-latent-class-technical-paper-2004>. Zugriff am: 17. Juli 2019.

- [27] B. Wheeler, *Package ,AlgDesign': Algorithmic Experimental Design*. The R project for statistical computing <https://www.r-project.org/>, 2019.
- [28] C. H. Leschinski, *Package 'MonteCarlo': Automatic Parallelized Monte Carlo Simulations*. The R project for statistical computing <https://www.r-project.org/>, 2019.
- [29] B. Hammer, M. Frasco und E. LeDell, *Package 'Metrics': Evaluation Metrics for Machine Learning*. The R project for statistical computing <https://www.r-project.org/>, 2018.
- [30] R. J. Hyndman und A. B. Koehler, „Another look at measures of forecast accuracy“, *International Journal of Forecasting*, Jg. 22, Nr. 4, S. 679–688, 2006.