B 20280

MARKETING

ZFP – Journal of Research and Management

Friederike Paetz, Maren Hein, Peter Kurz, and Winfried J. Steiner
Latent Class Conjoint Choice Models: A Guide for Model Selection, Estimation, Validation, and Interpretation of Results

Dirk Temme and Sarah Jensen
Missing Data – Better "Not to Have Them", but What If You Do? (Part 1)

Thomas Otter
A Note on Confidence Intervals and Model Specification

Volume 41
4/2019
4. Quarter 2019

C.H.BECK · Vahlen · Munich
www.marketing-zfp.de

2450201904

Latent Class Conjoint Choice Models: A Guide for Model Selection, Estimation, Validation, and Interpretation of Results

By Friederike Paetz, Maren Hein, Peter Kurz, Winfried J. Steiner*

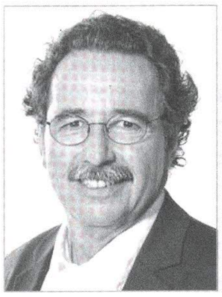
The consideration of preference heterogeneity in consumer choice behavior has become state of the art. In addition, the identification of consumer segments remains essential for marketing managers. For disaggregate consumer choice data representing the basis of segmentation, the latent class multinomial logit (MNL) model is currently the most popular approach for estimating segment-specific preferences.

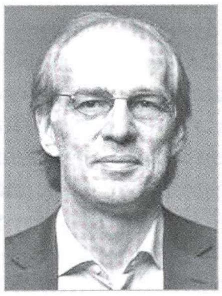
After addressing the theoretical background of the latent class MNL model, we use an empirical choice-based conjoint data set to illustrate model estimation and validation, as well as how the estimation results should be interpreted. A particular focus lies on the model selection process, i.e., the determination of an appropriate number of segments. We further work out interpretation pitfalls when the existing preference heterogeneity of consumers is ignored. This will ultimately provide a guide for applying the latent class MNL model regarding model selection, estimation, validation, and interpretation of results both from a statistical and managerial perspective.

1. Motivation

Accommodating preference heterogeneity in consumer choice behavior has become state of the art in marketing theory and practice. Companies account for preference heterogeneity for example via price discrimination or product differentiation. Using these kinds of marketing activities, firms aim to increase revenues and profits through a better absorption of consumers' willingness to pay and/or by offering more attractive product variations for certain consumers or consumer segments.

Developing successful marketing strategies makes it necessary to know consumers' preferences. Several approaches have been proposed here, e.g., multidimensional scaling (MDS) or conjoint analysis, that have proven their ability to determine customer preferences. For example, MDS draws inferences for the optimal positioning of products in a multidimensional space. Similarity/dissimilarity judgements of consumers for pairs of products in particular are transformed into distances. Based on these distances, customer perceptions and preferences for these products are represented within a perceptual space which can be interpreted in terms of meaningful psychological attribute dimensions (e.g., Kruskal 1964). Conjoint analysis approaches on the other hand circum-





Friederike Paetz is Assistant Professor of Marketing (Akademischer Rat) at Clausthal University of Technology, Institute of Management and Economics, Julius-Albert-Str. 6, 38678 Clausthal-Zellerfeld, Germany, Phone: +49/5323 72 7682, Fax: +49/5323 72 7659, E-Mail: friederike.paetz@tu-clausthal.de

Maren Hein works at the State Office for Statistics Lower Saxony (Dezernentin), Göttinger Chaussee 76, 30453 Hannover, Germany, Phone: +49/511 98981021, E-Mail: maren.hein@statistik.niedersachsen.de

Peter Kurz is Managing Partner, Innovation & Methods, at bms marketing research & strategy, Landsberger Str. 487, 81241 München, Germany, Phone: +49/89 88 96 94 42, Fax: +49/89 88 96 94 44, E-Mail: p.kurz@bms-net.de

Winfried Steiner is Professor of Marketing at Clausthal University of Technology, Institute of Management and Economics, Julius-Albert-Str. 2, 38678 Clausthal-Zellerfeld, Germany, Phone: +49/5323 72 7650, Fax: +49/5323 72 7659, E-Mail: winfried.steiner@tu-clausthal.de
* Corresponding Author.

vent the necessary subsequent transformation of psychological attribute dimensions to address the physical product attributes and levels needed for product design decisions. In contrast to the MDS approach, they directly elaborate on manageable attributes and attribute levels, determining related part-worth utilities of customers for product attributes or their levels. This means that conjoint analysis approaches are more directly applicable for redesigning existing, or designing new products or product lines, and they are predominantly used in practical applications (e.g., Green et al. 1981; Chen and Hausman 2000; Pullman et al. 2002; Fruchter et al. 2006; Fruchter and Fligler 2007; Feit et al. 2010; Tsafarakis 2016; Luchs et al. 2016). The conjoint analysis approach is currently (and not unexpectedly) the leading tool for measuring consumer preferences. The resulting part-worth utility estimates are used for (new) product and product line design, pricing and market segmentation (e.g., DeSarbo et al. 1995; Johnson and Olberts 1996; Steiner and Hruschka 2000, 2002a, 2002b; Krieger et al. 2004; Baier and Brusch 2009; Schön 2010; Steiner 2010; Schlereth et al. 2011; Gensler et al. 2012; Tuma and Decker 2013).

Diverse variants of conjoint analysis have been proposed, including traditional conjoint analysis (TCA), adaptive conjoint analysis (ACA), limit conjoint analysis (LCA), or choice-based conjoint analysis (CBC) (TCA: Luce und Tukey 1964; Green und Rao 1971; Johnson 1974; ACA: Johnson 1987; LCA: Voeth and Hahn 1998; CBC: Louviere and Woodworth 1983). Today, the CBC variant is by far the most widely used conjoint approach both in academia and market research practice (Hartmann and Sattler 2002; Buyer et al. 2010; Orme 2019). The main reason for the dominance of the CBC approach is that it most closely mimics the real choice behavior of consumers by asking respondents to repeatedly choose their preferred alternative from a set of several alternatives (choice sets). Another reason for its widespread use is a result of today's powerful computation power which allows the efficient estimation of part-worth utilities from choice data, even on the individual respondent level via the application of hierarchical Bayesian estimation techniques (Orme 2000) [1]. Since unordered first choice data as collected in most CBC studies provide much less information compared to rating or ranking data collected in TCA, CBC models in former times were usually estimated on a more aggregate respondent level (at least by pooling respondents at the segment level, if not on a completely aggregated level (Vriens et al. 1998, p. 238)). In other words, segment-specific CBC models were already estimable before hierarchical Bayesian estimation techniques became available due to respondents being partly aggregated, thus strongly softening the degrees-of-freedom problem (i.e., estimating a large number of individual parameters based on relatively little individual information). Nevertheless, thanks to the introduction of Bayesian estimation methods and the availability of high-performance personal computers, preference heterogeneity can now be accommodated from choice data at any desired respondent level, even on the individual respondent level. According to Peter Kurz, head of research and development at Kantar TNS in Munich from 2006 to 2018, more than 95 % of all conjoint studies in market research practice use the CBC approach. Sawtooth Software, the worldwide leading company for software solutions in the field of conjoint analysis, estimates that 80 % of Sawtooth customers currently use the CBC approach to determine consumer preferences (Orme 2019).

At first glance, it appears most promising to accommodate preference heterogeneity on the finest, i.e. most disaggregated individual customer level. This allows the bias to be reduced when capturing preference heterogeneity compared to a more aggregated CBC model. Provided the statistical efficiency does not suffer from too sparse data on the individual respondent level, this lower bias should generate a better forecasting performance e.g. with regard to the prediction of choice shares based on part-worth utility estimates. In addition, estimating individual consumer preferences enables the customization of products to individual consumers, as well as (from a theoretical perspective) a perfect absorption of a consumer's surplus, leading to maximum revenues (e.g., Schoder et al. 2006; Franke et al. 2009; Paetz 2018). On the other hand, from a practitioner's point of view, it is not necessarily advisable to derive managerial implications from individual-level part-worth utility estimates. Consumer goods companies commonly use second- or third-degree price discrimination, and set different prices for latent or known consumer segments, or offer only a small number of product variants rather than individualized product solutions. This is why segment-specific utility estimates provide helpful information for marketing managers to determine segment-specific marketing strategies.

There is an ongoing debate within this context among marketing academics whether to consider preference heterogeneity via a discrete or a continuous modeling approach (Wedel and Kamakura 2000; Natter and Feurstein 2002; Moore 2004; Karniouchina et al. 2009; Keane and Wasi 2013). On the one hand, utilizing the discrete approach conforms to the assumption of perfectly homogeneous segments and hence to the use of a segmentation framework employed for example by a latent class approach. However, the assumption of strictly separated homogeneous segments may be overly restrictive (Wedel et al. 1999, p. 222). On the other hand, addressing heterogeneity via the continuous modeling approach requires the a priori specification of a statistical distributional form, e.g. a Gaussian or gamma distribution. In this case, individual-level part-worth utilities are estimated by hierarchical Bayesian (HB) models. Here, the assumption of using a pre-specified distributional form to represent consumer heterogeneity can be criticized because estimation results are sensitive to this distributional form (Wedel and Kamakura 2000, p. 327).

Several studies have compared the performance of HB and latent class conjoint choice models. From a statistical point of view, the findings relating to the performance of HB versus latent class (conjoint) choice models are not unambiguous. In empirical CBC studies, the HB approach tends to be superior to the latent class approach regarding model fit and forecasting accuracy (see e.g., Allenby and Ginter 1995; Allenby et al. 1998; Moore et al. 1998; Moore 2004; Karniouchina et al. 2008; Keane and Wasi 2013). However, several counterexamples are known where neither approach outperformed its counterpart in terms of forecasting accuracy (e.g., Huber et al. 1998; Teichert 2001a; Teichert 2001b). In the context of simulation studies, i.e. studies in which true preference structures are assumed to be known, the findings yield a comparable performance of both approaches in terms of parameter recovery and forecasting accuracy (e.g., Andrews et al. 2002a; Gensler 2003). So no unequivocal recommendation of one single approach can be made, and the decision about which approach to use depends on the researcher's preference for and the purpose of the respective study. Louviere (2006, p. 178) strongly argues for the use of a discrete representation of heterogeneity in applied economics because segmentation models often fit the data at least as well as random parameter models such as HB models, and are much easier to estimate and interpret.

The following will focus on the segmentation approach, and specifically on the latent class MNL model to determine segment-specific preferences. Two classes of segmentation approaches can be distinguished in the context of CBC: one-step approaches (to which the latent class segmentation approach belongs) and two-step approaches (a priori segmentation and post-hoc segmentation approaches). The a priori segmentation approach widely applied in the past obtains segment-specific preferences for pre-specified segments. Here, based on individual background variables of respondents like socio-demographic or psychographic variables, consumers in a first step are clustered into distinctive segments. Subsequently, homogeneous preferences within each pre-specified segment are determined via conjoint analyses. Preferences between segments obviously only differ here if the individual background variables of the respondents actually relate to the differences in preferences. Although the existence of this kind of relationship already was criticized in early literature (e.g., Green and Krieger 1991; DeSarbo et al. 1992), it is now undisputed that at least socio-demographic variables only seldom constitute appropriate predictors for customer behavior. Post-hoc CBC segmentation approaches use HB estimation techniques to estimate individual preferences in a first step. In the second step, cluster algorithms such as k-means or two-step clustering procedures are used to group consumers with similar preferences into distinctive segments. Two-step segmentation approaches have always been applied in marketing practice (Wedel and Kamakura 2000, p. 5; Fennell et al. 2003; Lee 2006; Lopes 2012; Crabbe et al. 2013), although it is well-known that they perform worse compared to one-step segmentation approaches when it comes to forecasting accuracy (Steiner and Baumgartner 2004; Paetz 2016; Paetz 2018). In addition, with the post-hoc segmentation approach, it is at least questionable to search for segments in the second step after having estimated individual part-worth utilities in the first step via the standard HB approach. This standard HB approach assumes a multivariate normal distribution as first stage prior for individuals' preferences, disavowing the possibility of clearly separated segments from its initiation. In contrast to the two-step segmentation approach, one-step segmentation approaches (like the latent class modeling approach) simultaneously group respondents into segments according to their stated preferences and estimate segment-specific utilities. This circumvents the critique of two-step segmentation approaches which optimize different and unconnected statistical criteria in the two-step segmentation process (Gustafson et al. 2003).

In the following, we focus on the latent class multinomial logit (MNL) model as the most popular one-step segmentation model for CBC data. The prominence of the latent class MNL model is reflected in both a vast amount of proposed academic studies (e.g., Teichert 2000; Greene and Hensher 2003; Temme 2007; Goossens et al. 2014; Elshiewy et al. 2017) and the extensive application of the model in market research practice, the latter of which is supported by the widespread use of the latent class module by Sawtooth Solutions Inc. (Sawtooth Software 2004). After addressing the theoretical background of the latent class MNL model, we use an empirical CBC data set to illustrate how the model is estimated and validated, as well as how the estimation results are to be interpreted. A particular focus lies on the model selection process, i.e., the identification of an appropriate number of segments. We further work out pitfalls in interpretation when existing preference heterogeneity of consumers is ignored. We ultimately provide a guide for applying the latent class MNL model concerning model selection, estimation, validation, and interpretation of results from both a statistical and managerial perspective. Our contribution further elaborates on the *Marketing ZFP – Journal of Research and Management* article "Multinomial Logit Models in Marketing – From Fundamentals to State-of-the-art" in which Elshiewy et al. (2017) summarized the most prominent MNL model variants for analyzing choice behavior in marketing research, among them the latent class MNL model. As with their article, our paper is intended to help academics (e.g., advanced students or Ph.D. students of business economics or business-related degree programs) and practitioners (e.g., working at market research institutes or market research departments) theoretically understand and empirically apply the latent class MNL approach.

The remainder of the paper is structured as follows: In Section 2, we first review the simple MNL model and discuss important measures to assess the statistical model performance and interpret estimation results. Next, we extend this basic framework to the latent class MNL model that accommodates unobserved heterogeneity in customer preferences on the segment level. One of the crucial steps here is to determine the right number of segments, which is referred to as *model selection* and discussed in Section 3. In Section 4, we provide an empirical application of the latent class approach that illustrates the model estimation and model selection process, and we offer a guideline for model validation and the interpretation of results. In Section 5, we conclude and briefly discuss advanced segmentation approaches.

2. Random Utility Models

The use of random utility models is common practice in conjoint choice analysis (Train 2003, p. 18). The MNL model and its extensions are very popular here. For didactic reasons, we start with the simple (aggregate) MNL model, consider its derivation based on random utility theory, and provide the basics for model calibration using maximum likelihood estimation. We further discuss statistical performance measures to assess model fit and model validation as well as economic measures to interpret parameter estimates (e.g., attribute importance). Subsequently, we extend the simple MNL to the latent class MNL to account for unobserved preference heterogeneity. Again, we provide details on model derivation and the estimation procedure, and further point to model-specific characteristics such as how respondents are assigned to segments.

2.1. Multinomial Logit Model: Theory and Estimation

We assume that a respondent *j* behaves utility maximizing when making a choice decision. In a certain choice occasion *t*, a respondent is assumed to choose that alternative *i* out of several alternatives $m=1,..,I$ which provides the highest utility U_{jti} to her/him. However, the utilities U_{jti} are unknown to the researcher in discrete choice experiments or choice-based conjoint studies (Elshiewy et al. 2017, p. 33). In particular, the researcher can only observe the respondent's most preferred alternative from her/his stated preferences, but not the exact utility value for the chosen alternative nor the exact utility values for non-chosen alternatives. Therefore, she/he has to rely on explanatory variables that describe the different alternatives and are captured within a design matrix X_i. The design matrix X_i contains the coding of all attribute levels ($k=1,..,K_l$) of *L* relevant attributes ($l=1,...,L$) that make up the alternatives. Furthermore, the researcher is most likely unable to specify all relevant determinants underlying a respondent's choice decision or utility. The researcher therefore has to incorporate this uncertainty in addition to the deterministic utility components in the model, leading to so-called random utility models (McFadden 1974). In particular, random utility models assume that the stochastic utility U_{jti} is composed of a deterministic and a random part:

$$U_{jti} = \sum_{l=1}^{L} \sum_{k=1}^{K_l} x_{jtilk} \cdot \beta_{jlk} + \varepsilon_{jti}, \qquad (1)$$

where x_{jtilk} is a binary variable (as a component of the design matrix X_i) that equals one if alternative *i* in choice occasion *t* possesses level *k* of attribute *l*, and zero otherwise. β_{jlk} represents the part-worth utility of respondent *j* for level *k* of attribute *l*, and ε_{jti} is a random error term [2]. The first part of the utility is the deterministic part which is formulated by the researcher, with the second part constituting the stochastic part of the utility. The stochastic part captures all effects impacting the utility of the respondent, which are not observable or controllable by the researcher and therefore not incorporated into the deterministic part (i.e., as exploratory variables). The random error term is assumed to follow a specific probability distribution. Commonly, ε_{jti} is assumed to be Gaussian or i.i.d. Gumbel distributed, which respectively results in the multinomial probit (MNP) model or the MNL model. The MNL model has the desirable property of a closed-form solution for choice probabilities. Accordingly, the probability that respondent *j* chooses alternative i^* in choice occasion *t* is:

$$P_{jt}(i^*) = \frac{\exp(\mu \cdot \sum_{l=1}^{L} \sum_{k=1}^{K_l} x_{jti^*lk} \cdot \beta_{jlk})}{\sum_{m=1}^{I} \exp(\mu \cdot \sum_{l=1}^{L} \sum_{k=1}^{K_l} x_{jtmlk} \cdot \beta_{jlk})}, \qquad (2)$$

where μ ($\mu > 0$) is the scale parameter of the MNL model. For model estimation, μ is implicitly set to one (Train 2003, p. 41).

Whereas the existence of a closed-form solution for choice probabilities represents an advantage, the MNL assumes independent error terms both across choice decisions of one respondent and across choice decisions of different respondents. In addition (and this might be more critical), the MNL model involves proportional substitution patterns across alternatives known as the *independence of irrelevant alternatives* (IIA) property (Luce 1959). Stated otherwise, the ratio of choice probabilities between two alternatives is independent of the existence of other alternatives. It is however important to note that the IIA property holds only on the aggregation level on which the part-worth utility parameters have been estimated. If we account for preference heterogeneity on the segment level by estimating for example a latent class model (i.e., part-worth utilities vary for different segments), the IIA property will hold only on the segment level, but is softened on more aggregated levels, e.g., for aggregate choice share predictions. Therefore, while the simple MNL model ($\beta_j = \beta \ \forall j$) is fully prone to the IIA property, the latent class MNL is not.

Maximum likelihood estimation can be performed to determine the homogeneous part-worth utility vector β of the simple MNL model. Assuming that respondents' choice decisions are independent, the log-likelihood function can be stated as (e.g., Balderjahn 1993):

$$LL(\beta) = \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{I} \delta_{jti} \cdot \ln P_{jti}, \quad (3)$$

where the binary variable δ_{jti} equals one if respondent j has chosen alternative i in choice occasion t (i.e., from choice set t), and zero otherwise. Because the first partial derivative has no closed-form solution, maximization of the log-likelihood requires iterative procedures such as Newton-Raphson algorithms or gradient-based approaches (e.g., Gensler 2006, p. 257). However, the log-likelihood function is globally concave, and iterative procedures are able to find the unique maximum (McFadden 1974).

2.2. Multinomial Logit Model: Validation and Economic Evaluation

Both statistical performance measures and economic criteria should be used to evaluate the estimation results of an MNL model. Three classes of measures exist for assessing the statistical model performance: fit, predictive validity, and parameter recovery statistics. Model fit can be measured in terms of the root likelihood, pseudo-R^2 and in-sample hit rates.

The *root likelihood* is computed as the geometric mean of in-sample hit probabilities and can therefore be interpreted as the estimated average choice probability of alternatives chosen by the respondents across choice occasions.

$$Root\ Likelihood = \sqrt[J \cdot T]{\prod_{j=1}^{J} \prod_{t=1}^{T} \prod_{i=1}^{I} P_{jti}^{\delta_{jti}}} \quad (4)$$

Consequently, the higher the average hit probability, the better the model fits the observed choice data. The root likelihood ranges between the reciprocal of the number of alternatives in choice occasion t (lower bound) and 1 (upper bound). For instance, if a choice task contains four alternatives, ¼ = .25 is the lowest possible root likelihood value corresponding to the chance probability of a null model (Sawtooth Software 2017).

The *pseudo-R^2*, also known as likelihood-ratio index, McFadden's R^2, or percent certainty, results from the comparison of the log-likelihood of the estimated MNL model with predictors X ($LL(X)$) and the log-likelihood of the null model without predictors ($LL(0)$), and is calculated as (McFadden 1974; Hauser 1978; Ben-Akiva and Lerman 1985):

$$\rho^2 = 1 - \frac{LL(X)}{LL(0)} \quad (5)$$

The pseudo-R^2 ranges between 0 and 1. A value of 0 corresponds to $LL(\beta) = LL(0)$, implying that the predictor variables do not explain anything about a respondent's utility, meaning part-worth utilities are zero. A pseudo-R^2 of 1 indicates that the exploratory variables completely explain respondents' utilities respectively their choices. In practical applications, a value between 0.2 and 0.4 yet suggests a good model fit (McFadden 1979).

Similarly, the log-likelihood values of two nested models can be compared to decide if an MNL model that contains one or more additional predictors (also called the full model) fits the choice data significantly better than an MNL model that does not contain these additional predictors (also called the restricted model). As a special case, the restricted model can represent the null model as well. The corresponding likelihood ratio (LR) test for nested models is calculated via (Train 2003, p. 74):

$$LR = 2 \cdot (LL_{full} - LL_{restricted}) \quad (6)$$

The LR test statistic is approximately $\chi^2(n)$-distributed, where n is equal to the difference in the number of parameters between the full and restricted model (e.g., Franses and Paap 2001, p. 98).

The *in-sample hit rate* reports the share of correctly predicted choices of respondents across choice occasions. The hit rate lies in the range between 0 % and 100 %, where higher hit rates correspond to a better model fit (e.g., Vriens et al. 1996; Andrews et al. 2002b).

The predictive performance of an MNL model can be evaluated in terms of holdout choice sets which represent the choices of respondents that were not considered (i.e., held out) for model estimation. Equivalent to the in-sample hit rate, the *out-of-sample hit rate* represents the share of correctly predicted choices of respondents in holdout choice occasions. For predictive purposes in particular, it furthermore makes sense to compute the *relative hit rate* which relates the hit rate to the number of alternatives that were offered in a certain choice occasion. For example, if three alternatives were offered in a choice set, the chance rate of the null model for a correctly predicted choice is already one-third, hence one-third yields the lower boundary against which the hit rate should be contrasted (otherwise the estimated model would predict worse than the null model). The relative hit rate is calculated as the ratio of the hit rate of the estimated model to the hit rate of the null model. If for instance the relative hit rate equals 2, the estimated MNL predicts two times better than chance. The computation of relative hit rates is especially helpful when comparing the out-of-sample performance between studies where the number of alternatives offered in choice occasions differs. In addition, the *root likelihood* (see model fit) can also be computed based on holdout choices.

The *mean absolute error* (*MAE*) as an aggregate-level measure across respondents also refers to holdout choices and is computed as the mean absolute deviation between observed shares-of-choice W_i to predicted shares-of-choice $\widehat{W}_i$ across alternatives $(i=1,...,I)$ of each holdout task.

$$MAE = \frac{1}{I} \cdot \sum_{i=1}^{I} | W_i - \widehat{W}_i |$$

The MAE equals 0 if predicted shares-of-choice coincide with observed shares-of-choice, and would represent a perfect (ex-post) forecast in this case. The forecasting accuracy decreases with an increasing MAE value. With regard to the interpretation of MAE values, it is important to note that the MAE also depends on the number of holdout alternatives, i.e. it tends to be smaller for an increasing number of holdout alternatives (ceteris paribus). We recommend calculating the predicted shares-of-choice based on the first choice rule that assigns a probability of 1 to the alternative with the highest predicted utility for a respondent in a given holdout choice scenario. This is because choice shares here do not depend on the scaling of the estimated part-worths (e.g., Hein et al. 2019a; 2019b).

Similarly, the *root mean square error (RMSE)* relates to the squared deviation between observed shares-of-choice W_i and predicted shares-of-choice $\widehat{W}_i$ across alternatives $(i=1,...,I)$ of each holdout task:

$$RMSE = \sqrt{\frac{1}{I} \cdot \sum_{i=1}^{I} (W_i - \widehat{W}_i)^2}$$

Using this measure, larger prediction errors are penalized more strongly compared to smaller prediction errors. By taking the square root, the RMSE is directly interpretable in terms of measurement units (Leeflang et al. 2000). Like the MAE measure, values of zero indicate no prediction error, i.e. a perfect prediction of observed shares-of-choice. Note that the MAE and RMSE measures are only meaningful for fixed holdout tasks (i.e., if all respondents were offered the same alternatives in the holdout task) or if random holdouts were identical for subgroups of respondents (i.e., for those respondents who saw the same version of the choice task).

In the case of simulated data, i.e. if true part-worth utilities for simulated respondents are available, parameter recovery can be evaluated in addition to model fit and predictive validity. Parameter recovery can be addressed in relative terms via the *Pearson correlation* between true and re-estimated part-worth utilities (e.g., Hein et al. 2019a; 2019b) and/or in absolute terms via the *root mean squared error* (RMSE(β)) between true and re-estimated part-worth utilities (e.g., Andrews et al. 2002b; Hein et al. 2019a; 2019b). The latter is computed as:

$$RMSE(\beta) = \sqrt{\frac{1}{P} \cdot \sum_{p=1}^{P} (\beta_p^{estimated} - \beta_p^{true})^2}, \qquad (7)$$

where P denotes the number of estimated parameters corresponding to the length of the estimated part-worth utility vector. The ability to correctly recover attribute level preferences with CBC studies is especially relevant for product design decisions because firms want to find good (if not optimal) attribute levels for their products.

From an economic perspective, the researcher should at first check if parameter estimates are plausible. This means that the *signs* or the *order* of the parameter estimates correspond to her/his a priori expectations. For instance, if a price attribute is included in a study, and price was coded linearly using one single parameter, we would expect a negative sign, except when prices were perceived as a proxy for quality. Accordingly, if the price was coded using the part-worth utility model (i.e., one price parameter for each specified price level), we would expect that higher price levels to be less preferred to lower price levels.

It is important to note that it is not possible to directly interpret the magnitude of estimated (part-worth) utilities. The MNL model returns interval-scaled parameters that are invariant to the addition of and/or multiplication by a constant value. It is therefore possible to add a constant and/or multiply all part-worth estimates using a constant without changing the results so that the utility for the most preferred alternative remains the highest utility in inter-alternative comparisons (Train 2003, p. 27). Due to the interval scale property of utilities, only differences in utilities matter which, for example, allow the comparison of differences between two levels of one attribute versus two levels of another attribute. Using utility differences for interpreting part-worth utility estimates further forms the basis for calculating *(relative) attribute importance*. The importance of one attribute l corresponds to its part-worth utility range, and the relative attribute importance (ri_l) to its part-worth utility range divided by the sum of part-worth utility ranges of all attributes:

$$ri_l = \frac{\max_{k=1,...,K_l} \beta_{lk} - \min_{k=1,...,K_l} \beta_{lk}}{\sum_{m=1}^{I} (\max_{k=1,...,K_m} \beta_{mk} - \min_{k=1,...,K_m} \beta_{mk})} \qquad (8)$$

The relative importance of an attribute ranges between 0 and 1, where a smaller value indicates a smaller importance of the focal attribute for a consumer's choice decision. The interpretation of the attribute importances should relate to the mean attribute importance that depends on the number of attributes. If, for example, five attributes are considered the mean relative attribute importance is ⅕ = .2. Consequently, while a relative attribute importance of .2 should be interpreted as small in the case of three relevant attributes, it should be interpreted as large if 10 or more attributes were considered.

2.3. Latent Class Multinomial Logit Model

We so far have focused on the aggregate MNL model, which treats all consumers as having equal (homogeneous) preferences and yields one part-worth utility vector for all respondents. Obviously, not considering preference heterogeneity may lead to the estimation of preferences of an average consumer who might actually not exist. For example, consider the following example of raw onions in salads. Raw onions are typically detested or liked by salad eaters. Hence, the preference distribution may consist of two clearly separated segments, i.e., "raw onion likers" and "raw onion haters". Disregarding preference heterogeneity by the estimation of an aggregate MNL model and, therefore, the specification of one and the same part-worth utility parameters for all respondents results in an "average" part-worth utility that can be interpreted as an average consumer who is neutral about raw onions in her/his salad. This would not reflect the true preference distribution.

Accommodating preference heterogeneity in a conjoint choice model requires the assumption of a specific distributional form of heterogeneity. If we assume a *continuous* distribution, we will account for heterogeneity on an individual level, i.e., β_j differs across all respondents, resulting in (HB) random effects modeling. If we assume a *discrete* distribution of preference heterogeneity, we will account for heterogeneity on the segment level. Here, the part-worth utility vectors β_j of all respondents j, which belong to the same segment (class) s, are equal, e.g., $\beta_j = \beta_s \ \forall j$ in segment s.

In addition to estimating segment-specific part-worth utility vectors β_s, $s=1,...,S$, the researcher has to determine relative segment weights π_s. For a pre-specified number of segments, estimation of segment-specific parameters can be performed by maximizing the following likelihood function

$$LL(\theta) = \prod_{j=1}^{J} \sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} \prod_{i=1}^{I} P_{jt(i)s}^{\delta_{jti}}, \tag{9}$$

where θ is the vector of all parameters to be estimated, i.e., β_s and π_s, $s=1,...,S$. The binary variable δ_{jti} equals one if respondent j has chosen alternative i on choice occasion t, and zero otherwise. Maximization of the (log-) likelihood function can be performed by iterative procedures such as the Expectation-Maximization (EM) algorithm. The EM algorithm is widely used to maximize the log-likelihood function in latent class models (Greene 2008, p. 596; Tuma and Decker 2013, p. 3). A brief description of the EM algorithm is provided by Greene (2001), and a detailed description can be found in McLachlan and Krishnan (2007). However, several other iterative procedures such as the modified Newton gradient search of Kamakura and Russel (1989) or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (applied e.g., in a latent class MNL context by Boxall and Adamowicz 2002) enable the estimation of the parameter vector θ. Free software packages for the estimation of latent class MNL models such as the R-package gmnl allow a selection between varying estimation procedures, i.e., BFGS algorithm, Berndt-Hall-Hall-Hausman (BHHH) algorithm and a Newton Raphson algorithm (Sarrias and Dazanio 2017).

In contrast to the log-likelihood of the aggregate MNL model, the log-likelihood of the latent class MNL model is not globally concave. Hence, the iterative approaches mentioned above may stick in one of several existing local optima. This is why the researcher is advised to start the iterative procedure from different starting values to determine the global optimum or find at least a near-optimal solution. As a consequence, there is ongoing research for improved random-starting methods for the EM algorithm (e.g., Schepers 2015). If different "optimal" solutions are obtained for different starting values, the solution with the best found value of the log-likelihood should be retained (Grün and Leisch 2008).

One characteristic of the latent class MNL model is its fuzzy assignment of respondents to segments (cf. Wedel and DeSarbo 1994). Hence, respondents are not strictly assigned to exactly one segment, but receive a certain segment membership probability instead. For example, if three segments exist, a certain respondent may be assigned a probability of 95 % to belong to segment 1 and membership probabilities of 3 % and 2 % to belong to segments 2 and 3. At first glance, this fuzzy assignment appears to complicate a strict segment profiling and targeting. However, latent class models by definition assume a disjunct classification and therefore that each respondent can be assigned to exactly one segment (of course being the one with the highest segment membership posteriori probability (DeSarbo et al. 1995; Paetz and Guhl 2017)). Following Gensler (2003, p. 135), the posteriori probability τ_{js} that respondent j belongs to segment s is calculated via

$$\tau_{js} = \frac{\pi_s \cdot \prod_{t=1}^{T} \prod_{i=1}^{I} P_{jt(i)s}^{\delta_{jti}}}{\sum_{r=1}^{S} \pi_r \cdot \prod_{t=1}^{T} \prod_{i=1}^{I} P_{jt(i)r}^{\delta_{jti}}} \tag{10}$$

The strict assignment of respondents ("crisp segmentation") enables the profiling of segments based on respondents' individual background characteristics, e.g., (socio-) demographic variables like gender or age, or psychological variables such as personal values or personality variables (e.g., Jain and Kaur 2006; Huang and Sarigöllü 2008; Paetz 2016). A concise profiling helps marketing managers comprehensively describe focal consumer segments and improve the targeting of segments using the marketing mix.

The same performance measures as discussed for the aggregate MNL model can be used to evaluate the results of a latent class MNL model for a fixed number of segments. Here we will abstain from a formal presentation to prevent redundancy, and simply refer the reader back to Section 2.2.

3. Model Selection

Both statistical and economic selection criteria have to be considered to determine the optimal number of segments (Ramaswamy and Cohen 2013). For statistical criteria, the use of information criteria, posterior membership probabilities, an entropy-based measure, as well as the inspection of segment sizes and standard errors of parameter estimates are suggested. Economic criteria on the other hand relate to the managerial interpretability of segment solutions, which constitutes the basis for deriving positioning and targeting strategies (Tuma and Decker 2013). It is specifically important to analyze whether segment-specific differences regarding both the estimated utility structures and attribute importances exist (e.g., Kamakura et al. 1994; DeSarbo et al. 1995; Vriens et al. 1998). In addition, cross-tabulations between adjacent segment solutions (2-segment solution versus 3-segment solution, 3-segment solution versus 4-segment solution, etc.) can be inspected to see how segment solutions compare, i.e., to obtain a feeling about how stable an existing segment solution is if an additional segment is allowed (Sawtooth Software 2004).

3.1. Statistical Criteria for Model Selection

Percent Certainty, Chi Square, and Root Likelihood

In a first step, the evolution of the percent certainty, root likelihood and the chi square statistic for different numbers of segments can be inspected. As with the percent certainty (see Section 2.2), the chi square statistic is also based on the comparison between the log-likelihood of the considered segment solution and the log-likelihood of the null model:

$$\chi^2 = 2 \cdot (LL(\theta) - LL(0)) \tag{11}$$

All three statistics evaluate the goodness-of-fit of a segment solution without penalizing the complexity of the estimated model. Therefore, the three measures get better as the number of segments increases because a larger number of segments always leads to a better model fit (Sawtooth Software 2004). As a consequence, the evolution of the incremental improvement in model fit from adding more and more segments is the point here upon which to decide the right number of segments.

Information Criteria

Information criteria are widely used to determine how many segments to retain (Soromenho 1994; Andrews et al. 2002b; Andrews and Currim 2003; Nylund et al. 2007; Tuma and Decker 2013). In general, information criteria are based on the log-likelihood function augmented by some penalty term in order to penalize models for complexity and indicate the optimal number of segments at their minimum value (Andrews and Currim 2003; Tuma and Decker 2013). We next describe the Akaike's information criterion (AIC, Akaike 1974), the Bayesian information criterion (BIC, Schwarz 1978) and the consistent Akaike's information criterion (CAIC, Bozdogan 1987) in more detail. These are widespread in empirical applications (Tuma and Decker 2013).

The *AIC* is calculated as

$$\text{AIC} = -2 \cdot \text{LL}(\theta) + 2 \cdot n_\theta, \tag{12}$$

where $-2 \cdot \text{LL}(\theta)$ is -2 times the maximized log-likelihood function of the latent class MNL model, and n_θ denotes the number of parameters estimated. The AIC is known to tend to overestimate the optimal number of segments (Ramaswamy et al. 1993, p. 109).

The *BIC* is obtained by adding the penalty term $\ln(\text{N}) \cdot n_\theta$ instead of $2 \cdot n_\theta$:

$$\text{BIC} = -2 \cdot \text{LL}(\theta) + \ln(\text{N}) \cdot n_\theta, \tag{13}$$

where N is the number of observations, i.e., the number of choice tasks per respondents times the number of respondents in the data set (Jain et al. 1994, p. 320; Frühwirth-Schnatter 2011, p. 14; Sawtooth Software 2012). The BIC involves a stronger penalization than the AIC.

The *CAIC* has been proposed for empirical applications with a large number of observations. It penalizes overparameterization more strongly than the BIC (Ramaswamy et al. 1993; Andrews and Currim 2003):

$$\text{CAIC} = -2 \cdot \text{LL}(\theta) + [(\ln \text{N})+1] \cdot n_\theta \tag{14}$$

Note that all information criteria are heuristics to determine the appropriate number of segments. They can point to different "optimal" segment solutions (Ramaswamy et al. 1993). Further information criteria such as AIC3, ABIC, WAIC1, or WAIC2 have been proposed in the relevant literature (e.g., Sclove 1987; Andrews and Currim 2003; Watanabe 2010).

Posterior Membership Probabilities

As mentioned, the latent class model returns segment membership probabilities for respondents, i.e., respondents are not formally assigned to exactly one segment per se. On the other hand, less fuzzy segment membership probabilities are desirable for profiling and targeting segments. An indicator of how well the segment solution fits the data is therefore the average maximum membership probability. It is important to note that an average membership probability of 100 % can only be obtained with artificial data, whereas in empirical applications the average membership probability will be much lower (Sawtooth Software 2004). According to Greene and Hensher (2013), very small average maximum membership probabilities indicate model overfitting.

Entropy

Based on segment membership posterior probabilities of respondents τ_{js}, $j=1,...,J$, $s=1,...,S$, the related entropy measure provides information about the quality of the separation of the segments, and therefore about a valid assignment of respondents to segments (cf. Wedel and Kamakura 2000, p. 92; Ramaswamy et al. 1993; Paetz and Guhl 2017). The entropy is calculated as

$$E_S = 1 + \frac{\sum_{j=1}^{J} \sum_{s=1}^{S} \tau_{js} \cdot \ln(\tau_{js})}{J \cdot ln(S)} \tag{15}$$

The entropy measure is also bounded between 0 and 1. The entropy approximates 1 if the membership posteriori probabilities of respondents approximate 1 for one segment, and 0 for the other segments. In contrast, if the assignment of respondents to certain segments is extremely fuzzy, the entropy measure approximates 0. An entropy value close to 0 therefore indicates that the centroids of the conditional parametric distributions are not well-separated for the number of segments considered.

Segment Size

It is further important to take segment sizes into account when determining the adequate number of segments. If a considered segment solution contains segments with only tiny segment sizes, this would argue against that segment solution (Sawtooth Software 2012) because management decisions should also be based on segment sizes (Tuma and Decker 2013). So if a segment contains only a small number of respondents, it is not worth addressing that segment because it is not theoretically meaningful (Garver et al. 2008). As a rule of thumb, Garver et al. (2008) suggested a minimum of 30 respondents per segment, while Peter Kurz, head of research and development at Kantar TNS in Munich from 2006 to 2018, suggests a minimum of 70 respondents for practical applications [3]. The latter suggestion is based on the fact that sample sizes used for commercial applications of choice-based conjoint studies in market research practice are usually much larger (between 500 and 1,500 respondents) than in academic studies.

Standard Errors

It is also useful to consider the standard errors of parameter estimates for selecting the right number of segments. Large standard errors indicate model overfitting and are associated with imprecise parameter estimates (Greene and Hensher 2013; Hensher et al. 2015). Large standard errors often coincide with the existence of tiny segment sizes.

In addition to the criteria discussed, *in-* and *out-of-sample hit rates* can support the model selection process as well.

3.2. Economic Criteria

Segment-Specific Preferences and Attribute Importances

Aggregate analyses of choice data using the simple MNL model assumes that consumers show the same preference structure. Only one part-worth utility vector is estimated for the entire sample. Contrarily, conducting a segment-level analysis using the latent class MNL model approach provides segment-specific part-worth utility vectors and corresponding segment-specific attribute importances (DeSarbo et al. 1995). The description and interpretation of segment solutions forms the basis for the derivation of positioning and targeting strategies (Tuma and Decker 2013). Therefore, beyond all statistical criteria, a strong indicator of an adequate segment solution is that noticeable differences in estimated part-worth structures and/or attribute importances across segments must exist (DeSarbo et al. 1995).

Cross-Tabulation of Segment Solutions

After assigning each respondent to the segment for which she/he reveals the highest segment membership posterior probability, segment solutions can be further compared by cross-tabulating different solutions versus one another. For example, the 2-segment solution can be tabulated against the 3-segment solution, the 3-segment solution can be tabulated against the 4-segment solution, and so on (Sawtooth Software 2004; 2012). Doing this helps assess how stable segment solutions are if an additional segment is allowed. For example, if only one segment of a 2-segment solution is split into two classes in the 3-segment solution while the other segment is retained in the 3-segment solution, the 3-segment solution would be reasonable. Note that depending on the sample size, an increasing number of segments increases the probability of tiny segment sizes.

4. Empirical Application

In the following, we use an empirical choice-based conjoint data set to illustrate model estimation and validation, and how the estimation results are to be interpreted. A particular focus lies on the model selection process, i.e., the identification of an appropriate number of segments. First, we describe our empirical data set, and subsequently estimate the aggregate MNL model and interpret the results. After that, we account for unobserved preference heterogeneity by estimating latent class MNL models with different numbers of segments. The focus here lies on model selection, i.e., the determination of the appropriate number of segments for our data set. Finally, we contrast the aggregated solution to the best segment solution and work out pitfalls in interpretation, if existing preference heterogeneity of consumers is ignored.

4.1. Data

For an illustration, we use an empirical choice-based conjoint data set for the product category *car tires* comprising 872 UK respondents. This data was collected in 2012 and provided by Kantar TNS, one of the largest market research institutes worldwide. The focal product of car tires can be described along nine relevant attributes (with three to six attribute levels respectively), where the star levels of the first five attributes relate to an established rating system for tires:

- road holding on wet road: 3 stars, 4 stars, 5 stars
- braking distance on dry road: 3 stars, 4 stars, 5 stars
- shock resistance: 3 stars, 4 stars, 5 stars
- absorption of road noise and vibration: 3 stars, 4 stars, 5 stars
- reduction of the car's CO2 emissions: 3 stars, 4 stars, 5 stars
- reduction of fuel consumption: none, $2/full tank, $4/full tank
- longevity: 18,000 miles, 24,000 miles, 30,000 miles, 36,000 miles, 42,000 miles
- tire brand: brand A, brand B, brand C, brand D, brand E, brand F
- price: $38, $42, $47, $52, $56

Respondents were asked to choose their preferred tire across 15 choice sets, resulting in a total of 13,080 observations. Each choice set contained three tire alternatives as well as a "no purchase" option. Since the no purchase option received a choice share of only 6.83 %, the study design (specification of attributes and attribute levels) could be assumed as well-considered.

We used 13 choice sets (= 11,336 observations) for model estimation, and two choice sets (= 1,744 observations) for holdout validation. As with all other attributes, we here also selected for the price attribute the part-worth utility model instead of a linear specification, the latter of which is often used in conjoint-analytic approaches (Steiner and Meißner 2018). Estimation of a part-worth model for the price attribute (i.e., one part-worth parameter for each price level) allows nonlinearities to be captured in the price-utility relationship, and is therefore more flexible (Orme 2007). It is well-known that price response functions often reveal nonlinearities due to existing price thresholds (e.g., Baumgartner and Steiner 2007; Steiner and Meißner 2018).

4.2. Aggregate MNL Model

For the aggregate MNL model, part-worth utilities from respondents' choices were determined via traditional maximum likelihood estimation. In contrast to the latent class MNL model, the aggregate MNL model totally ignores respondent heterogeneity because one part-worth utility vector for all respondents is estimated. This is equivalent to setting the number of segments to S=1. We obtained a log-likelihood of –12,499, a chi square of 6,432 (as the *LL*(0) is –15,715), and a corresponding percent certainty of .21, where the latter indicates only a moderate model fit. The root LL and the in-sample hit rate are .33 and .53, suggesting that the model fits the data better than would be expected by chance (with 4 alternatives in each choice set, the chance rate is .25). Similarly, the out-of-sample hit rate of .53 indicates that the predictive validity of the aggregate MNL model is better than that of a null model. For AIC, BIC, and CAIC we obtained values of 25,050, 25,240 and 25,266, respectively. The three information criteria were later used for model selection when the aggregate MNL (which corresponds to a 1-segment solution) was compared to the latent class solutions for larger numbers of segments. *Tab. 1* presents the parameter estimates for the aggregate MNL model. As seen, most of the parameter estimates turned out statistically significant ($p < .05$) and had high face validity. The order of magnitudes of the estimated effects for the attribute levels all fell within a priori expectations (except for the brand attribute where no a priori expectation existed). For the first five attributes, respondents preferred the highest number of stars. As expected, the lowest number of stars was least preferred. Also, higher reductions in fuel consumption (as measured in dollar value per full tank) were preferred to lower reductions. Higher tire longevity was preferred to lower longevity. Higher price levels were furthermore less preferred to lower price levels. And with regard to the type of brand, respondents preferred brand E most, whereas brand C was the least favored. Altogether, the aggregate results suggest that the optimal car tire should have a good performance on all features, and that respondents prefer a tire of the brand E at the lowest price of $38. An inspection of the attribute importance shows that the longevity of a tire is most important for the respondents (25.13 %), followed by road holding on wet roads (18.32 %), price (14.89 %), and braking distance on dry roads (12.32 %). We considered nine attributes, so the mean relative importance of attributes is $\frac{1}{9} \approx 11$ %. Hence, the importances of the remaining attributes in a range between 4 % and 8 % are quite small. However, as mentioned, disregarding preference heterogeneity may lead to the estimation of preferences for an average consumer who does not actually exist. In order to account for preference heterogeneity, i.e., to prove whether the data contains a certain level of heterogeneity that the MNL model is not able to capture, we conduct a segment-level analysis in the following by varying the number of segments from 2 up to 8 segments, subsequently comparing the best segment solution to the aggregate results.

4.3. Latent Class MNL Models and Statistical Model Selection

We estimated latent class MNL models for up to eight segments, using five replications each with different starting values to avoid convergence to local optima, or at least to minimize the risk of highly suboptimal solutions (cf. Grün and Leisch 2008, p. 235; Melnikov and Maitra 2010, p. 88). We retained only the best replication for each latent class MNL model, i.e., the solution with the highest chi square statistic (cf. Sawtooth 2004). The results of the model fit and predictive performance statistics of each segment solution are displayed in *Tab. 2*. The results of the aggregate MNL model (1-segment solution) discussed above are also reported here for comparison.

As expected, the higher the number of segments, the better the model fit as indicated by LL values that decrease (in absolute terms) with an increasing model complexity. Because the percent certainty, the chi square statistic, and the root LL are closely related to the LL value, we can observe that these measures also improve with a higher number of segments. Information criteria play a central role for model selection when penalizing for an increasing model complexity, i.e., weighing between a better model fit from an increasing number of segments and the additional number of parameters to be estimated for more segments. The results show that the AIC, BIC and CAIC suggest different segment solutions. The AIC has its minimum for eight segments. The AIC is however known for its rather soft penalization, bringing with it the risk of an overestimation of the number of segments. The BIC statistic offers a larger penalization that depends on the sample size and here decreases monotonically from 2 segments up to 6 segments. For a still higher number of segments (7- and 8-segment solutions) the BIC again increases and thus turns out worse. Note, however, that the

Attribute	Attribute level	Estimate	Std. error
Road holding on wet roads	3 stars	**-.476**	.017
	4 stars	**.062**	.015
	5 stars	**.414**	.014
Braking distance on dry roads	3 stars	**-.338**	.016
	4 stars	**.077**	.015
	5 stars	**.261**	.015
Shock resistance	3 stars	**-.141**	.016
	4 stars	-.003	.015
	5 stars	**.143**	.015
Absorption of road noises and vibrations	3 stars	**-.147**	.016
	4 stars	.028	.015
	5 stars	**.119**	.015
Reduction of the car's CO2 emissions	3 stars	**-.107**	.015
	4 stars	.011	.015
	5 stars	**.096**	.015
Reduction of fuel consumption	None	**-.173**	.016
	$2/full tank	**-.037**	.015
	$4/full tank	**.209**	.015
Longevity	18,000 miles	**-.705**	.027
	24,000 miles	**-.270**	.024
	30,000 miles	**.111**	.023
	36,000 miles	**.349**	.022
	42,000 miles	**.516**	.022
Tire brand	Brand A	-.026	.026
	Brand B	**-.057**	.027
	Brand C	**-.135**	.027
	Brand D	-.044	.026
	Brand E	**.154**	.026
	Brand F	**.109**	.026
Price	$38	**.262**	.022
	$42	**.237**	.022
	$47	**.092**	.023
	$52	**-.130**	.024
	$56	**-.461**	.026

Tab. 1: Parameter estimates for the aggregate MNL model

Notes: Statistically significant results ($p < .05$) are indicated in bold

improvement in BIC for the 6-segment solution compared to the 5-segment solution is only marginal (BIC difference of only 5) in comparison to the improvement from four to five segments (BIC difference of 172). According to Raftery (1995), only a difference in BIC values larger than 10 between two models suggests that the more complex model is really superior. The BIC therefore points towards more of a 5-segment solution. Using the CAIC, the model complexity is penalized even more heavily than compared to the BIC. As seen in *Tab. 2*, the CAIC is lowest for the 5-segment solution, again suggesting that a 5-segment solution is appropriate. Altogether, the information criteria suggest a five-segment solution. Taking a look at the standard errors for the 5-segment solution versus the 6-segment solution confirms this preliminary assessment. Here we observe large stan-

# segments	1	2	3	4	5	6	7	8
# parameters	26	53	80	107	134	161	188	215
LL	-12,499	-11,551	-11,069	-10,772	-10,560	-10,431	-10,318	-10,230
Percent Certainty [in %]	21	27	30	32	33	34	34	35
Root LL (in-sample)	.332	.361	.377	.387	.394	.398	.402	.406
Chi Square	6,432	8,328	9,291	9,887	10,311	10,568	10,794	10,970
AIC	25,050	23,208	22,299	21,757	21,387	21,184	21,013	20,890
BIC	25,240	23,597	22,886	22,542	22,370	22,365	22,392	22,467
CAIC	25,266	23,650	22,966	22,649	22,504	22,526	22,580	22,682
Entropy	1	.986	.951	.948	.951	.958	.963	.964
Avg. max. membership probability	1	.990	.941	.917	.908	.901	.889	.873
In-sample hit rate [in %]	53	54	60	63	64	64	65	67
Out-of-sample hit rate [in %]	53	55	61	60	60	60	61	61

Tab. 2: Number of parameters and performance measures

Segments	1	2	3	4	5	6	7	8
2-segment solution	17.5 %	82.5 %						
3-segment solution	15.7 %	52.3 %	32.0 %					
4-segment solution	21.0 %	24.8 %	39.4 %	14.9 %				
5-segment solution	18.6 %	21.1 %	23.6 %	14.5 %	17.2 %			
6-segment solution	18.5 %	11.1 %	25.9 %	22.6 %	16.9 %	5.0 %		
7-segment solution	22.5 %	16.3 %	6.6 %	14.7 %	5.0 %	10.9 %	23.9 %	
8-segment solution	6.9 %	7.5 %	21.3 %	4.6 %	7.7 %	18.1 %	11.3 %	22.6 %

Tab. 3: Relative segment sizes

dard errors for part-worths for one group of the 6-segment solution, suggesting model overfitting.

In addition to the information criteria and the inspection of standard errors, the entropy measure is around 95 % for the 5-segment solution, indicating that the five segments are sufficiently distinct and well-separated. The average maximum membership probability of about 91 % also confirms that respondents as a rule can be treated as belonging to one of the five segments. The in-sample hit rate as a measure of model fit considerably increases from the 1-segment solution (53 %) to the 5-segment solution (64 %), as does the out-of-sample hit rate as a measure of predictive validity from 53 % for the 1-segment solution to 60 % for the 5-segment solution. The out-of-sample performance only improves marginally beyond five segments. Note that the choice sets contained three "real" alternatives plus the none option, which implies a chance rate of 25 %. All segment solutions therefore are clearly superior to the null model. The inspection of the segment sizes (see *Tab. 3*) reveals that all segments of the 5-segment solution are sufficiently large, consisting of at least 127 respondents (14.5 % out of 872 respondents for the smallest segment). In contrast, the 6-segment solution comprises one segment with only 44 members, which is far lower than 70 (as suggested for larger sample sizes, see Section 3.1) and near the lower bound of 30 as suggested by Garver et al. (2008). Hence, according to the statistical criteria discussed so far, a 5-segment solution appears appropriate for the empirical data at hand. In the following, the appropriateness of the 5-segment solution is checked from an economic perspective.

4.4. Economic Interpretation of the 5-Segment Solution

Beyond the pure statistical measures for model selection discussed above, cross-tabulations of segment solutions provide evidence of the stability of the latent segment so-

Segment	1	2	3	4	5	Total
1	17	16	0	1	149	**183**
2	15	5	201	0	0	**221**
3	126	210	3	0	0	**339**
4	2	0	0	127	0	**129**
Total	**160**	**231**	**204**	**128**	**149**	**872**

Tab. 4: Cross-tabulation of the 4-segment versus 5-segment solution

lutions obtained from the latent class approach. *Tab. 4* shows the cross-tabulation between the 4-segment solution and the 5-segment solution, the latter of which we so far favored based on the statistical criteria.

As seen, groups 1, 2, and 4 of the 4-segment solution are largely preserved in the 5-segment solution, and segment 3 which was the largest one in the 4-segment solution (339 respondents) is split up into two smaller segments that are well interpretable regarding differences in their preferences. The finer 5-group segmentation therefore appears well-justified.

The segment-specific attribute importances as well as the segment-specific parameter estimates for the 5-segment solution are displayed in *Tab. 5* and *Tab. 6*, respectively. The first segment consists of respondents for whom price (18 %) and longevity of a tire (15.70 %) are most important. A closer inspection of the parameter estimates for the price attribute shows that the respondents prefer a medium price. Obviously, respondents of segment 1 consider the price to be a partial indicator of quality because they reject low prices (which can be plausible for durable goods like tires which are infrequently purchased). Further, segment 1 as a rule prefers a higher longevity to a shorter lifespan of the tire. Members of segment 2 primarily care about the longevity of the tire (45.65 %) and clearly prefer the best level of 42,000 miles. All other attributes are apparently less important. Segment 3 assigns the highest importances to the two attributes of road-holding under wet road conditions (37.66 %) and braking distance of the tire under dry road conditions (22.08 %). Here, as expected, respondents prefer tires with a better rating (more stars) to tires with worse ratings (less stars), respectively. Respondents of segment 4 attach greater importance to the brand (21.41 %) and also care about the longevity of the tire (18.33 %) and the attribute of road holding on wet roads (16.03 %). In particular, they prefer brand E, mostly followed by brand F, and obviously reject the other four brands. They further favor a better road holding on wet roads and a higher longevity of the tire. Finally, segment 5 has a clear focus on the price of the tire (55.25 %) and attaches some importance to the longevity of the tire (17.35 %). Respondents in this segment are extremely price-sensitive (in contrast to segment 1) and favor cheap price levels. All other attributes are unimportant.

4.5. Comparison between the Aggregate (1-Segment Solution) and the 5-Segment Solution

The comparison of results for the aggregate MNL model (1-segment solution) and the 5-segment solution obtained from the latent class model reveals that the 5-segment solution clearly outperforms the 1-segment solution. As reported in *Tab. 2*, the log-likelihood improves from –12,499 to –10,560. The likelihood-ratio test yields a test statistic of 3,878, which is $\chi^2(109)$ distributed and therefore indicates a statistically significantly better

Relative segment size	Segment 1 (18.6%)	Segment 2 (21.1%)	Segment 3 (23.6%)	Segment 4 (14.5%)	Segment 5 (17.2%)
Road holding on wet roads	12.478	8.987	37.659	16.025	3.829
Braking distance on dry roads	12.035	6.190	22.077	10.145	4.506
Shock resistance	8.783	3.621	7.730	4.662	3.190
Absorption of road noise and vibration	9.906	4.253	5.412	6.753	3.759
Reduction of the car's CO2 emissions	4.885	3.054	4.219	7.390	2.231
Reduction of fuel consumption	8.151	12.042	2.837	4.411	6.925
Longevity	15.704	45.646	10.664	18.330	17.352
Tire brand	10.054	5.116	5.047	21.405	2.957
Price	18.003	11.093	4.354	10.879	55.252

Tab. 5: Segment-specific attribute importances

Attribute	Attribute level	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Road holding on wet roads	3 stars	-.167(.033)	-.385(.036)	-1.939(.074)	-.510(.057)	-.182(.043)
	4 stars	-.018(.032)	.062(.033)	.432(.045)	.171(.049)	.061(.042)
	5 stars	.184(.031)	.323(.033)	1.507(.049)	.339(.047)	.121(.041)
Braking distance on dry roads	3 stars	-.186(.033)	-.270(.035)	-1.157(.050)	-.290(.053)	-.195(.043)
	4 stars	.033(.032)	.051(.034)	.294(.038)	.042(.049)	.033(.042)
	5 stars	.153(.031)	.218(.033)	.863(.042)	.248(.047)	.162(.041)
Shock resistance	3 stars	-.121(.033)	-.107(.034)	-.360(.041)	-.150(.051)	-.140(.043)
	4 stars	-.006(.032)	-.072(.034)	.013(.039)	.052(.049)	.028(.042)
	5 stars	.126(.031)	.179(.033)	.347(.039)	.098(.048)	.112(.042)
Absorption of road noise and vibration	3 stars	-.164(.033)	-.142(.034)	-.257(.040)	-.197(.051)	-.187(.043)
	4 stars	.114(.031)	-.051(.034)	.019(.039)	.035(.049)	.110(.042)
	5 stars	.050(.032)	.193(.033)	.238(.038)	.161(.048)	.076(.041)
Reduction of the car's CO2 emissions	3 stars	-.089(.033)	-.111(.034)	-.203(.040)	-.194(.051)	-.096(.043)
	4 stars	.049(.032)	-.018(.034)	.020(.038)	-.005(.049)	.015(.042)
	5 stars	.040(.032)	.130(.033)	.183(.039)	.198(.047)	.081(.041)
Reduction of fuel consumption	None	-.072(.032)	-.463(.036)	-.103(.039)	-.107(.050)	-.278(.043)
	$2/full tank	-.079(.033)	-.024(.034)	-.054(.039)	-.019(.049)	.009(.042)
	$4/full tank	.150(.031)	.487(.033)	.157(.039)	.126(.048)	.269(.041)
Longevity	18,000 miles	-.278(.052)	-2.002(.094)	-.643(.064)	-.614(.086)	-.845(.072)
	24,000 miles	-.032(.050)	-.690(.061)	-.154(.061)	-.140(.075)	-.309(.067)
	30,000 miles	.105(.049)	.290(.050)	.131(.060)	.082(.072)	.165(.063)
	36,000 miles	.164(.048)	.803(.049)	.333(.059)	.357(.067)	.527(.064)
	42,000 miles	.042(.049)	1.599(.053)	.332(.060)	.314(.069)	.462(.065)
Tire brand	Brand A	.038(.056)	-.052(.069)	.010(.069)	-.229(.087)	.151(.073)
	Brand B	-.104(.058)	-.190(.067)	-.006(.067)	-.224(.089)	.050(.073)
	Brand C	-.110(.057)	-.002(.070)	-.314(.070)	-.477(.094)	-.083(.073)
	Brand D	-.057(.057)	-.032(.067)	.077(.067)	-.158(.085)	-.077(.074)
	Brand E	.173(.055)	.061(.069)	.148(.069)	.658(.073)	-.027(.076)
	Brand F	.061(.056)	.214(.069)	.085(.069)	.431(.076)	-.015(.075)
Price	$38	-.317(.053)	.221(.059)	.045(.059)	.255(.069)	1.736(.071)
	$42	.065(.049)	.227(.061)	.144(.061)	.171(.070)	1.161(.067)
	$47	.189(.048)	.173(.059)	.004(.059)	.096(.071)	.379(.068)
	$52	.075(.048)	.027(.060)	.060(.060)	-.200(.077)	-.644(.082)
	$56	-.011(.050)	-.648(.061)	-.254(.061)	-.322(.079)	-2.632(.163)

Notes: Standard errors are in parantheses.

Tab. 6: Segment-specific estimates of parameters

model fit of the 5-segment solution ($p \leq .001$). The percent certainty markedly increases from 21 % to 33 %, with similar changes in the root LL from .33 to .39, while chi square improves from 6,432 for the aggregate solution to 10,311 for the 5-segment solution. The information criteria in particular strongly improves compared to the aggregate solution, with huge differences of 3,663 (AIC), 2,870 (BIC), and 2,762 (CAIC). In addition, the in-sample hit rate increases from 53 % for the 1-segment solution up to 64 % for the 5-segment solution. The out-of-sample hit rate is also much better for the 5-segment solution (60 %) compared to the aggregate MNL (53 %). Hence, the empirical data contain heterogeneous preferences that the aggregate MNL model is not able to capture. Once respondents' heterogeneity is accommodated by conducting a segment-level analysis, model fit and predictive validity considerably improve compared to the aggregate solution. Further, the inspection of the attribute importances and parameter estimates show that the five segments identified differ in their preferences. In contrast, the aggregated MNL model leads to the estimation of preferences of an average consumer.

The most important attribute is the longevity of the tire for this average consumer (25.13 %) and to a smaller extent road holding on wet roads (18.32 %) and price (14.89 %), see Section 4.2. The segment-level analysis instead reveals that there is one segment that strongly cares about the longevity of the tire, but that other segments have clearly different preferences. For example, a second segment is very price-sensitive and does not care about tire features. A third segment is brand-loyal and favors only a specific brand. These preferences are not detected and hence cannot be addressed via marketing strategies if the aggregate MNL model were to build the basis for utility estimation and decision making. In other words: relying on the aggregate solution may in fact result in wrong marketing implications.

5. Conclusions and Limitations

Accommodating preference heterogeneity in consumer choice modeling has become state of the art in marketing theory and practice. The identification of heterogeneous consumer segments is particularly relevant for marketing managers. In this field, the latent class multinomial logit (MNL) model is currently the most popular approach for estimating segment-specific preferences. Latent class MNL models simultaneously assign respondents to segments and yield segment-specific preferences.

This paper provided the theoretical background for the estimation of latent class MNL models. We then applied the methodology to an empirical choice-based conjoint data set for tires to illustrate model estimation, validation, and interpretation of the results. An important issue in this context is the selection of the appropriate number of segments which is guided by both statistical and economic criteria. In our study, statistical criteria pointed to a 5-segment solution which could very well be further interpreted from an economic point of view in the sense that segment-specific preferences and/or attribute importances clearly differed between segments. We finally compared the solution obtained from the aggregate MNL model to the selected 5-segment solution to illustrate that relying on the wrong model (in this case the aggregate MNL that completely ignores heterogeneity) may lead to wrong marketing implications. The 5-segment solution strongly outperformed the aggregated solution in terms of the statistical criteria used and therefore indicated that heterogeneous consumer preferences exist in the tire market. Put differently, the aggregate solution would claim preferences of an average consumer that do not actually exist there. Relying on the right segment model is crucial for targeting customers and ultimately for a company's success.

Using the latent class MNL approach, it is generally possible that a certain number of segments seems favorable based on statistical criteria, but that the selected segment solution is not well interpretable from an economic point of view. For example, two segments may not substantially differ with regard to attribute importances and preference structures. Another phenomenon often observed in practical applications is a so-called 'catch-all' segment capturing all responses which cannot be assigned to another segment. Accordingly, the utility functions of respondents in this segment cannot be adequately reproduced by a mixture of utility functions of other segments as well (e.g., Teichert 2001b). As a consequence, all or many attributes in this segment might turn out somewhat important which further limits the meaningfulness of this segment. In these cases, a trade-off between pure statistical model selection and economic interpretability must be solved, where economic interpretability should be the decisive criterion.

As such, statistical criteria should be used to guide the model selection process in the right direction, even though the final decision on the "optimal" number of segments should be based on economic criteria. For large sample sizes, it can further happen that statistical criteria improve up to an unrealistic large number of segments, and that no reasonable segment structure can be identified. In this context, successive cross-tabulations of segment solutions might indicate that adding another segment destroys the previously identified segment structure. Then, the market might not consist of well-separable segments at all, and a continuous approach in terms of an HB model might be more appropriate for addressing heterogeneity when compared to a discrete approach like the latent class model.

Profiled segments based on socio-demographic and/or psychographic background variables are needed to target consumers of segments via marketing-mix activities. Usually, the profiling step is performed once the segments have been determined. However, so-called concomitant variable mixture models were proposed that extend the standard latent class approach by reparameterizing the membership probabilities to depend on consumer's background characteristics (e.g., Kamakura et al. 1994). Therefore, assigning respondents to segments, estimating segment-specific utility structures, and profiling the segments are performed simultaneously in these approaches. Further advanced statistical approaches for market segmentation based on consumer choice data like mixture-of-normals models or Dirichlet process prior models have been proposed and applied in the academic marketing literature (e.g., Allenby et al. 1998; Kim et al. 2004; Baumgartner and Steiner 2007; Gilbride and Lenk 2010; Voleti et al. 2017). While the latent class approach assumes strictly homogeneous segments, mixture-of-normals and Dirichlet process prior models additionally allow for inner-segment heterogeneity. No commercial software is yet available for the estimation of these advanced models. Here the user is advised to utilize the free programming language R or the SAS package.

Finally, a latent class multinomial probit model that accounts for both heterogeneity and utility dependencies of alternatives within and across choice sets was recently proposed by Paetz and Steiner (2017). As with probit models in general, model estimation is computationally much more expensive for increasing numbers of alternatives and segments.

Notes

[1] For a discussion on further advantages and disadvantages of the CBC approach, see Steiner and Meißner (2018).

[2] Along with the part-worth utility model used here, other utility models exist, including vector or ideal point models. An overview of the different utility models is provided in Green and Srinivasan (1978, 1990), and Steiner and Meißner (2018).

[3] Interview with Peter Kurz on October 9th, 2019.

References

Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the Heterogeneity of Demand. *Journal of Marketing Research*, 35, 384–389.

Allenby, G. M., & Ginter, J. L. (1995). Using Extremes to Design Products and Segment Markets. *Journal of Marketing Research*, 32, 392–403.

Andrews, R. L., Ainslie, A., & Currim, I. S. (2002a). An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *Journal of Marketing Research*, 39, 479–487.

Andrews, R. L., Ansari, A., & Currim, I. S. (2002b). Hierarchical Bayes Versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery. *Journal of Marketing Research*, 39, 87–98.

Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, 40, 235–243.

Baier, D., & Brusch, M. (2009). *Erfassung von Kundenpräferenzen für Produkte und Dienstleistungen*. In: D. Baier, & M. Brusch (Eds.) Conjointanalyse, Berlin: Springer, 3–18.

Balderjahn, I. (1993). *Marktreaktionen von Konsumenten. Ein theoretisch-methodisches Konzept zur Analyse der Wirkung marketingpolitischer Instrumente*, Berlin: Duncker & Humblot.

Baumgartner, B., & Steiner, W. J. (2007). Are Consumers Heterogeneous in Their Preferences for Odd and Even Prices? Findings from a Choice-Based Conjoint Study. *International Journal of Research in Marketing*, 24, 312–323.

Ben-Akiva, M. E. & Lerman, S. R. (1985). *Discrete Choice Analysis. Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Boxall, P. C., & Adamowicz, W. L. (2002). Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach. *Environmental and Resource Economics*, 23, 421–446.

Bozdogan, H. (1987). Model Selection and Akaike's Information criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrica*, 52, 345–370.

Buyer, D., Cleff, T., & Frank D. (2010). Von Präferenzen zu Marktanteilen – Bestandsaufnahme zum Einsatz von Conjoint-Analysen zur Absatzprognose in der Marktforschungspraxis. *Planung & Analyse*, 4, 45–49.

Chen, K. D., & Hausman, W. H. (2000). Mathematical Properties of the Optimal Product Line Selection Problem Using Choice-Based Conjoint Analysis. *Management Science*, 46, 327–332.

Crabbe, M., Jones, B., & Vandebroek, M. (2013). Comparing Two-Stage Segmentation Methods for Choice Data with a One-Stage Latent Class Choice Analysis. *Communications in Statistics – Simulation and Computation*, 42, 1188–1212.

DeSarbo, W. S., Ramaswamy, V., & Cohen, S. H. (1995). Market Segmentation with Choice-Based Conjoint Analysis. *Marketing Letters*, 6, 137–147.

DeSarbo, W. S., Wedel. M., Vriens, M., & Ramaswamy, V. (1992). Latent Class Metric Conjoint Analysis. *Marketing Letters*, 3, 273–288.

Elshiewy, O., Guhl, D., & Boztuğ, Y. (2017). Multinomial Logit Models in Marketing – From Fundamentals to State of the Art. *Marketing ZFP – Journal of Research and Management*, 39, 32–49.

Fennell, G., Allenby, G. M., Yang, S. & Edwards, Y. (2003). The Effectiveness of Demographic and Psychographic Variables for Explaining Brand and Product Category Use. *Quantitative Marketing and Economics*, 1, 223–244.

Feit, E., Beltramo, M. A., & Feinberg, F. M. (2010). Reality Check: Combining Choice Experiments with Market Data to Estimate the Importance of Product Attributes. *Management Science*, 56, 785–800.

Franses, P. H., & Paap, R. (2001). *Quantitative Models in Marketing Research*. Cambridge: Cambridge University Press.

Franke, N., Keinz, P. & Steger, C. (2009). Testing the Value of Customization: When Do Customers Really Prefer Products Tailored to Their Preferences?, *Journal of Marketing*, 73. 103–121.

Fruchter, G. E., & Fligler, A. (2007). The Various Brands in the Optimal Product Line vs. the Optimal Single Product. *Innovative Marketing*, 3, 15–22.

Fruchter, G. E., Fligler, A., & Winer, R. S. (2006). Optimal Product Line Design: Genetic Algorithm Approach to Mitigate Cannibalization. *Journal of Optimization Theory and Applications*, 131, 227–244.

Frühwirth-Schnatter, S. (2011). Model-based Clustering of Time Series – A Review from a Bayesian Perspective. Working Paper, Chicago.

Garver, M. S., Williams, Z., & Taylor, G. S. (2008). Employing Latent Class Regression Analysis to Examine Logistics Theory: An Application of Truck Driver Retention. *Journal of Business Logistics*, 29(2), 233–257.

Gensler, S. (2003). *Heterogenität in der Präferenzanalyse*, Wiesbaden: DUV.

Gensler, S. (2006). Ermittlung von Präferenzen für Produkteigenschaften mit Hilfe der Choice-Based Conjoint Analyse. *Wirtschaftswissenschaftliches Studium*, 35, 254–258.

Gensler, S., Hinz, O., Skiera, B., & Theysohn, S. (2012). Willingness-to-pay Estimation with Choice-Based Conjoint Analysis: Addressing Extreme Response Behavior with Individually Adapted Designs. *European Journal of Operational Research*, 219, 368–378.

Gilbride, T. J., & Lenk, P. J. (2010). Posterior Predictive Model Checking: An Application to Multivariate Normal Heterogeneity, *Journal of Marketing Research*, 47, 5(5), 896–909.

Goossens, L. M. A, Utens, C. M. A., Smeenj, F. W. J., Donkers, B., van Schayck, O. C. P., & Rutten-van Mölken, M. P. M. (2014). Should I Stay or Should I Go Home? A Latent Class Analysis of a Discrete Choice Experiment on Hospital-At-Home. *Value in Health*, 17, 588–596.

Green, P. E., Carroll, J. D., & Goldberg, S. M. (1981). A General Approach to Product Design Optimization Via Conjoint Analysis. *Journal of Marketing*, 45, 17–37.

Green, P. E., & Krieger, A. M. (1991). Segmenting Markets with Conjoint Analysis. *Journal of Marketing*, 55, 20–31.

Green, P. E., Krieger, A. M. & Wind, Y. (2001). Thirty Years of Conjoint Analysis: Reflections and Prospects. *Interfaces*, 31, 56–73.

Green, P., & Rao, V. (1971). Conjoint Measurement for Quantifying Judgemental Data. *Journal of Marketing Research*, 8, 355–363.

Green, P. E., & Srinivasan, V. S. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, 5, 103–123.

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

Greene, W. (2001). Fixed and Random Effects in Nonlinear Models. Working Paper EC-01-01, Stern School of Business, Department of Economics, https://archive.nyu.edu/bitstream/2451/26204/2/1-1.pdf, (last accessed: October 5, 2019)

Greene, W. H. (2008). *Econometric Analysis*. 6th Ed, New Jersey: Pearson Prentice Hall.

Greene, W., & Hensher, D. (2003). A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B: Methodological*, 37, 681–698.

Greene, W., & Hensher, D. (2013). Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model. *Applied Economics*, 45(14), 1897–1902.

Grün, B., & Leisch, F. (2008). Identifiability of Finite Mixtures of Multinomial Logit Models with Varying and Fixed Effects. *Journal of Classification*, 25, 225–247.

Gustafson, A., Herrmann, A., & Huber, F. (2003). Conjoint Analysis as an Instrument of Market Research Practice. In A. Gustafson, A. Herrmann, & F. Huber (Eds.) *Conjoint Measurement: Methods and Applications*, 3rd Ed., Berlin: Springer, 5–46.

Hartmann, A., & Sattler, H. (2002). Commercial Use of Conjoint Analysis in Germany, Austria, and Switzerland. *Research Paper on Marketing and Retailing*, 0–14.

Hauser, J. R. (1978). Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach. *Operations Research*, 26 (3), 406–421.

Hensher, D., Rose, J. M., & Greene, W. H. (2015). *Applied Choice Analysis*. New York: Cambridge University Press.

Hein, M., Kurz, P., & Steiner, W. (2019a). Analyzing the Capabilities of the HB Logit Model for Choice-based Conjoint Analysis: A Simulation Study. *Journal of Business Economics*, First Online 06 March 2019.

Hein, M., Kurz, P., & Steiner, W. (2019b). On the Effect of HB Covariance Matrix Prior Settings: A Simulation Study. *Journal of Choice Modelling*, 31, 51–72.

Huang, R., & Sarigöllü, E. (2008). Benefit Segmentation of Tourists to the Caribbean. *Journal of International Consumer Marketing*, 20, 67–83.

Huber, J., Arora, N., & Johnson, R. M. (1998). Capturing Heterogeneity in Consumer Choices. ART Forum. American Marketing Association, Chicago.

Jain, S. K., & Kaur, G. (2006). Role of Socio-Demographics in Segmenting and Profiling Green Consumers. *Journal of International Consumer Marketing*, 18, 107–146.

Jain, D. C., Vilcassim, N. J., & Chintagunta, P. K. (1994). A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data. *Journal of Business & Economic Statistics*, 12(3), 317–328.

Johnson, R. M. (1974). Trade-Off-Analysis of Consumer Values. *Journal of Marketing Research*, 11, 121–127.

Johnson, R. (1987). Accuracy of Utility Estimation in ACA. Sawtooth Software, https://www.sawtoothsoftware.com/download/techpap/acaaccur.pdf, (last accessed: October 4, 2019)

Johnson, R. M., & Olberts, K. A. (1996). Using Conjoint Analysis in Pricing Studies: Is One Price Variable Enough? Sawtooth Software, https://www.sawtoothsoftware.com/download/techpap/onevar.pdf, (last accessed: October 4, 2019)

Kamakura, W. A., & Russell, G. J. (1989). A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research*, 26(4), 379–390.

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant Variable Latent Class Models for Conjoint Analysis. *International Journal of Research in Marketing*, 11, 451–464.

Karniouchina, E. V., Moore, W. L., van der Rhee, B., & Verma, R. (2009). Issues in the Use of Ratings-Based Versus Choice-Based Conjoint Analysis in Operations Management Research. *European Journal on Operational Research*, 197, 340–348.

Keane, M., & Wasi, N. (2013). Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior. *Journal of Applied Economics*, 28, 1018–1045.

Kim, J. G., Menzefricke, U., & Feinberg, F. M. (2004). Assessing Heterogeneity in Discrete Choice Models Using a Dirichlet Process Prior. *Review of Marketing Science*, 2, 1–39.

Krieger, A., Green, P. E. & Wind, Y. (2004). *Adventures in Conjoint Analysis: A Practioner's Guide to Trade-Off Modeling and Applications*. Philadelphia: Wharton School, University of Pennsylvania.

Kruskal, J. B. (1964). Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, 29, 1–27.

Lee, B. C. Y. (2006). Consumer Perceived Importance of Channel Authorization: A Post Hoc segmentation Approach to Dealing with Gray Markets. *Australasian Marketing Journal*, 14, 10–22.

Leeflang, P. S. H., Wittink, D. R.,Wedel, M., & Naert, P. A. (2000). *Building Models for Marketing Decisions*. Boston: Kluwer.

Lopes, S. D. F. (2012). Post Hoc Segmentation Combining Conjoint Analysis and a Two-Stage Cluster Analysis. *African Journal of Business Management*, 6, 169–176.

Louviere, J. L. (2006). What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and the Analysis of Discrete Choice Experiments. *Environmental and Resource Economics*, 34, 173–188.

Louviere J., & Woodworth, G. (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments. *Journal of Marketing Research*, 20, 350–367.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Luce, R., & Tukey, J. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1, 1–27.

Luchs, M. G., Swan, K. S., & Creusen, M. E. H. (2016). Perspective: A Review of Marketing Research on Product Design with Directions for Future Research. *Journal of Product Innovation Management*, 33, 320–341.

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, In P. Zarembka (Ed.), *Frontiers in Econometrics*, New York: Academic Press, 105–142.

McFadden, D. L. (1979). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. In D. Hensher and P. Stopher (Eds.), *Behavioural Travel Modelling*, London: Croom Helm, 279–318.

McLachlan, G. J., & Krishnan, T. (2007). *The EM Algorithm and Extensions*, 2nd Ed., Hoboken: Wiley.

Melnikov, V., & Maitra, R. (2010). Finite Mixture Models and Model-Based Clustering. *Statistical Surveys*, 4, 80–116.

Moore, W. L. (2004). A Cross-validity Comparison of Rating-Based and Choice-Based Conjoint Analysis Models. *International Journal of Research in Marketing*, 21, 299–312.

Moore, W. L, Gray-Lee, J., & Louviere, J. J. (1998). A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation. *Marketing Letters*, 9, 195–208.

Natter, M., & Feurstein, M. (2002). Real World Performance of Choice-Based Conjoint Models. *European Journal of Operational Research*, 137, 448–458.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 14, 535–569.

Orme, B. (2000). Hierarchical Bayes: Why all that attention? https://www.sawtoothsoftware.com/download/techpap/hbwhy.pdf, (last accessed: October 18, 2019)

Orme, B. (2007). Three Ways to Treat Overall Price in Conjoint Analysis, https://www.sawtoothsoftware.com/download/techpap/price3ways.pdf, (last accessed: October 4, 2019)

Orme, B. (2019). 2016 Customer Feedback Results. https://www.sawtoothsoftware.com/about-us/news-and-events/news/1693-results-of-2016-sawtooth-software-user-survey, (last accessed: October 4, 2019)

Paetz, F. (2016). Persönlichkeitsmerkmale als Segmentierungsvariablen: Eine empirische Studie. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 68, 279–306.

Paetz, F. (2018). Improving the Forecasting Accuracy of 2-Step Segmentation Models. In A. Fink, A. Fügenschuh, M. J. Geiger (Eds.), *Operations Research Proceedings 2016*, Berlin: Springer, 57–62.

Paetz, F., & Guhl, D. (2017). Understanding Differences in Segment-specific Willingness-to-pay for the Fair Trade Label. *Marketing ZFP – Journal of Research and Management*, 39, 38–47.

Paetz, F., & Steiner, W. J. (2017). The Benefits of Incorporating Utility Dependencies in Finite Mixture Probit Models. *OR Spectrum*, 39(3), 793–819.

Pullman, M. E., Moore, W. L., & Wardell, D. G. (2002). A Comparison of Quality Function Deployment and Conjoint Analysis in New Product Design. *Product Innovation Management*, 19, 354–364.

Raftery, A. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111–163.

Ramaswamy, V., & Cohen, S. H. (2013). Latent Class Models for Conjoint Analysis. In A. Gustafsson, A. Hermann, & F. Huber (Eds.), *Conjoint Measurement*, 3rd Ed., Berlin: Springer-Verlag.

Ramaswamy, V., DeSarbo, W. S, Reibstein, D. J., & Robinson, W. T. (1993). An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science*, 12, 103–124.

Sarrias, M., & Daziano, R. A. (2017). Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The gmnl Package. *Journal of Statistical Software*, 79, 1–46.

Sawtooth Software (2004). The CBC Latent Class Technical Paper, Version 3. https://www.sawtoothsoftware.com/download/techpap/lctech.pdf (last accessed: October 4, 2019)

Sawtooth Software (2012). Latent Class Manual, Version 4.5. https://www.sawtoothsoftware. com/download/techpap/lclass_manual.pdf, (last accessed: October 4, 2019)

Sawtooth Software (2017). The CBC System for Choice-Based Conjoint Analysis, Version 9. https://www.sawtoothsoftware.com/download/techpap/cbctech.pdf, (last accessed: October 4, 2019)

Schepers, J. (2015). Improved Random-Starting Method for the EM Algorithm for Finite Mixtures of Regressions. *Behavior Research Methods*, 47, 134–146.

Schlereth, C., Eckert, C., & Skiera, B. (2011). Estimation of Willingness to Pay Intervals by Discrete Choice Experiments. *Proceedings of the 40th Conference of the European Marketing Academy (EMAC)*, Ljubljana, Slovenia.

Schoder, D., Sick, S., Putzke, J., & Kaplan, A. M. (2006). Mass Customization in the Newspaper Industry: Consumers' Attitudes Toward Individualized Media Innovations. *International Journal on Media Management*, 8, 9–18.

Schön, C. (2010). On the Optimal Product Line Selection Problem with Price Discrimination. *Management Science*, 56, 896–902.

Schwarz, G. (1978). Estimating the Dimensions of a Model. *The Annals of Statistics*, 6, 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333–343.

Soromenho, G. (1994). Comparing Approaches for Testing the Number of Components in a Finite Mixture Model. *Computational Statistics*, 9, 65–78.

Steiner, W. J. (2010). A Stackelberg-Nash Model for New Product Design. *OR Spectrum*, 32(1), 21–48.

Steiner, W. J., & Baumgartner, B. (2004). Conjoint-Analyse und Marktsegmentierung. *Zeitschrift für Betriebswirtschaft (ZfB)*, 74(6), 1–25.

Steiner, W. J., & Hruschka, H. (2000). Conjointanalyse-basierte Produkt(linien)gestaltung unter Berücksichtigung von Konkurrenzreaktionen. *OR Spektrum*, 22(1), 71–95.

Steiner, W. J., & Hruschka, H. (2002a). A Probabilistic One-Step Approach to the Optimal Product Line Design Problem Using Conjoint and Cost Data. *Review of Marketing Science*, Working Papers 1(4), Art. 4.

Steiner, W. J., & Hruschka, H. (2002b). Produktliniengestaltung mit Genetischen Algorithmen. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (ZfbF)*, 54(7), 575–601.

Steiner, M., & Meißner, M. (2018). A User's Guide to the Galaxy of Conjoint Analysis and Compositional Preference Measurement. *Marketing ZFP – Journal of Research and Management*, 40, 3–25.

Teichert, T. (2000). Das Latent-Class Verfahren zur Segmentierung von wahlbasierten Conjoint-Daten: Befunde einer empirischen Anwendung. *Marketing ZFP – Journal of Research and Management*, 3, 227–239.

Teichert, T. (2001a). *Nutzenschätzung in Conjoint-Analysen*. Wiesbaden: Deutscher Universitätsverlag.

Teichert, T. (2001b). Nutzenermittlung in wahlbasierter Conjoint-Analyse: Ein Vergleich von Latent-Class- und Hierarchischem Bayes-Verfahren. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 53, 798–822.

Temme, J. (2007). Discrete-Choice-Modelle. In Albers, S., D. Klapper, U. Konradt, A. Walter, J. Wolf (Eds.), *Methodik der empirischen Sozialforschung*. 2nd Ed., Wiesbaden: Gabler, 327–341.

Train, K. E. (2003). *Discrete Choice Models with Simulation*, Cambridge: Cambridge University Press.

Tsafarakis, S. (2016). Redesigning Product Lines in a Period of Economic Crisis: A Hybrid Simulated Annealing Algorithm with Crossover. *Annals of Operations Research*, 247, 617–633.

Tuma, M. N., & Decker, R. (2013). Finite Mixture Models in Market Segmentation: A Review and Suggestions for Best Practices. *Electronic Journal of Business Research Methods*, 11, 2–15.

Voeth, M., & Hahn, C. (1998). Limit Conjoint-Analyse. *Marketing ZFP – Journal of Research and Management*, 20, 119–132.

Voleti, S., Srinivasan, V., & Ghosh, P. (2017). An Approach to Improve the Predictive Power of Choice-Based Conjoint Analysis. *International Journal of Research in Marketing 34*, (2), 325–335.

Vriens, M., Wedel, M., & Wilms, T. (1996). Metric Conjoint Segmentation Methods: A Monte Carlo Comparison. *Journal of Marketing Research*, 33(1), 73–85.

Vriens, M., Opperwal, H., & Wedel, M. (1998). Ratings-Based Versus Choice-Based Latent Class Conjoint Models – An Empirical Comparison. *Journal of the Market Research Society*, 40, 237–248.

Watanabe, S. (2010). Equations of states in singular statistical estimation. *Neural Networks*, 23, 20–34.

Wedel, M., & DeSarbo, W. S. (1994). A Review of Recent Developments in Latent Class Regression Models. In R. Bagozzi (ed.), *Advanced Methods of Marketing Research*, Cambridge: Blackwell Pub., 352–388.

Wedel, M., Kamakura, W., Arora, N., Bemmaor N. J. C., Erod, T., Johnson, R., Lenk, P., Neslin, S., & Poulson, C. (1999). Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling. *Marketing Letters*, 10, 219–232.

Wedel, M., & Kamakura, W. (2000). *Market Segmentation: Conceptual and Methodological Foundations*, 2nd Ed., Boston Dordrecht London: Kluwer.

Keywords

Latent Class Model, Discrete Choice Data, Multinomial Logit Model, Heterogeneity