Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis: Business Analitycs & Quantitative Marketing

Exploring properties of Menu-Based Conjoint Logit models with real holdout tasks.

Jedrzej Wynimko (570635)

Supervisor: prof. dr. Richard Paap
Second assessor: prof. dr. Eoghan O'Neill
Company supervisor: Remco Don

April 30, 2022

Abstract

In the era of mass customization of products and services the menu choiceboard offerings are becoming more prevalent across different industries. Consequently, methods for analyzing such menu-based choices which allow to learn about the customer preferences and produce market forecasts are gaining more focus among the researchers and practitioners.

In this study, we explore and compare the properties of different Logit Menu-Based Conjoint modelling configurations employing holdout tasks. The data are gathered utilizing menu tasks of high complexity to better mimic the real-life menu situations. We consider three underlying models: decompositional Serial-Cross Effects (SCE), single-model combinatorial Exhaustive Alternatives (EA) and Modular (MOD), a combinatorial approach incorporating cross-price effects between menu subsections. The price sensitivity is captured assuming product dependency by Alternative-Specific effects and in a novel application of the Generic effects formulation, which assumes that price sensitivity is invariant of the menu item. We also vary the main price effect form to be represented by a linear or a non-linear function. Parameters of applied model arrangements are estimated by the Aggregate Logit, the Hierarchical Bayes and the Hierarchical Bayes with trimming. We compare the models by means of predictive performance measures utilizing holdout tasks: Mean Absolute Error (MAE) and hit rate as well as goodness of data fit and computational time.

The findings are translated into recommendations aiding future researchers and practitioners in the appropriate modelling choices in analyzing MBC datasets. In general, less complicated methods seem to work best for analyzing the highly complex menu situations.

Contents

1 Introduction

In recent years, more and more companies began to offer the opportunity to customize products and services according to unique preferences and needs of their customers. The shift from mass production economy to one driven by mass customization is largely attributed to the development of the Internet and other advances in information technology which fundamentally changed the market conditions (Liechty et al., 2001). Companies' operations became more flexible, and consumers could now access information on the competition more easily and compare offers across the whole market without additional effort. Importantly, customization transformed the role of customers in the market, who ceased being a passive member, but began to play an active role in co-creating value (Liechty et al., 2001).

The change in consumers' decision-making process provided the companies with an unprecedented market opportunity, as marketers strongly believe that tailored products are more appealing to the customers, who are willing to pay extra for this possibility. Mass customization is also beneficial for the producers' operations. It allows to eliminate the problem of unloading unwanted goods at non-optimal prices or introducing a product failure into the market, which in both cases translates to a low- or no-profit strategy (Cohen and Liechty, 2007). The attractiveness of build-your-own products comes not only from the possibility to adjust the product prior to the purchase decision, but also from providing a superior value in comparison to ready off-the-shelf products by meeting the individual expectations of each consumer (Liechty et al., 2001). For these reasons the suppliers' focus shifted from creating optimal preassembled products to identifying the demanded product features and their levels. Moreover, it became crucial to understand how consumers want to customize the offer, which customers are interested in such possibility and to recognize the premium they are willing to pay for it.

In order to make such personalization feasible, companies are designing their products and services in a modular structure which reflects the customers' multi-choice decisions made for the available functionalities or add-on items. The offered portfolio is presented in a menu form showcasing items, features, components, prices or even delivery options, as well as bundles of products offered at discount prices when purchased together. Examples of such build-your-own choices can be found in almost all industries, starting with the food industry where fast-food joints are offering menus involving choices *a la carte* of all products, as well as bundles like extra-value meals, from which the customers can construct their preferred meals. Mass customization is also present in telecom, banking and insurance industries offering a range of services that can be adjusted to exact needs of a business or an individual customer (Orme (2020), Neuerburg et al. (2021), Ben-Akiva and Gershenfeld (1998)). Many high-technology suppliers, such as automotive and computer manufacturers quickly adapted this modular sales approach as their products comprise of hundreds of components, which determine their performance and price. Dell was one of the first companies that implemented such a supply model. Magretta (1998) describe how giving the customers a direct access to an on-line purchasing choiceboard, detailed information about the different configuration and possibility of customization became one of the main drivers that led Dell to reinvent the personal computer commerce and position themselves as a market leader. They also emphasize the connection towards the brand that customers feel when being involved in the design process of their product, which positively influences the market share.

The unique aspect of menu-based sales is that the choices made by the consumer are interrelated and made at the same occasion. Wind and Mahajan (1997) are the first to recognize the importance of analyzing products and services as a bundle of features rather then a set of independent characteristics. They emphasize the need for research methods that would not only aid the design of customized products, but also provide guidelines for improving the preassembled products offered to consumers not willing to pay the premium for a tailored product.

Conjoint analysis is a multi-attribute utility-measurement approach popular in marketing research first applied by Green and Rao (1971). It works on the assumption that humans make decisions by evaluating the overall desirability of a complex product or service, regarded as a function of the values of the product characteristics. Through experimentally designed questionnaires the respondents are asked to evaluate different product profiles consisting of multiple conjoined attributes. This decompositional preference technique allows researchers to investigate the trade-offs between different levels of product features and estimate people's preferences for these characteristics. According to Orme (2020), the most popular conjoint method to date is Choice-Based Conjoint (CBC), highly regarded by practitioners as it closely mimics buyer decision process in competitive contexts. Respondents choose out of a set of products that vary in features and their levels simulating different market conditions. The main advantage of such set up is revealing preferences and price sensitivities while taking into account the intangible forces in the market. However, it comes short with insights and reliable forecasts in the context of customizable products and services. Menu-Based Conjoint (MBC) is an extension of CBC, with a purpose of simulating the build-your-own purchase decision through multiple selection experiments. The choice alternatives are not mutually exclusive and lead to multivariate outcomes. Moreover, MBC incorporates the idea of bundling products at a discount price when purchased together which is a popular strategy that enables the supplier to increase the probability of more products being chosen, thus boosting their profit. Bakken and Bond (2004) recognize the benefits from the consumers perspective as well, for whom introducing bundles reduces the cognitive burden of making complicated choices.

As mass customization progresses, and the possibilities of personalizing purchased products and services become more complex, there is an increasing interest in designing and analyzing the MBC questionnaires. The modeling approaches proposed by Ben-Akiva and Gershenfeld (1998) and Liechty et al. (2001) became the cornerstone of the methodology. Followed by Cohen and Liechty (2007), who presents a comparative study of three different approaches to menu-choice modeling. Neuerburg et al. (2021) emphasize the research gap in the properties of various modeling approaches, such as the model behaviour and predictive performance under different study and data settings, and suggest that increasing number of practitioner-based studies, for instance Orme (2011), Orme (2020), Cordella et al. (2012), Dippold-Tausendpfund and Neuerburg (2018), stresses the need for academic guidance in the model selection process. One of the reasons hindering academic research is the high cost of data acquisition through questionnaires designed specifically for MBC studies.

In this study we aim to quantitatively explore the properties of Logit-based methods used for MBC analysis that are most applied in the market research practice. By varying aspects of the applied MBC modelling techniques we seek to identify how different configurations influence not only the predictive performance of considered approaches, but also goodness of fit to the data or parameter estimation time. The comparative study is performed on a high complexity menu, which accurately represents real-life menu situations present in the market. We consider three underlying models and test different parameter formulations, which are investigated for the linear and non-linear functional forms of the price effect. Moreover, the parameter estimation is performed by three techniques. Applying different performance measures allows to produce guidelines for the model selection process in a wider context, not limited only to forecasting accuracy. We formulate our insights as advantages and disadvantages of the modeling approaches to aid future researchers and practitioners in composing the most appropriate approach according to the aim of their analysis and chosen performance indicators.

The paper proceeds as follows. In Section 2 we discuss the relevant literature. Section 3 introduces and characterizes a menu-based dataset gathered for the purpose of this study. Section 4 outlines the applied methodology and performance measures. Section 5 includes a discussion of the obtained results. Finally, the recommendations and suggestions for future research are presented in Section 6.

2 Literature

The emergence of conjoint theory in marketing literature can be attributed to Green and Rao (1971), who utilized the conjoint measurement, a way of capturing the joint effects of two or more variables proposed by mathematical psychologists Luce and Tukey (1964), in quantifying complex purchase decisions and estimating preferences based on rank-ordered data. The method had been quickly rising in popularity among researchers, especially after the conditional logit for analyzing discrete choice behavior (McFadden, 1973) was introduced. Conceptualizing the choice analysis led to fundamental changes in conjoint studies - instead of using full-factorial designs where respondents ranked concepts according to their preferences, the industry transitioned to questionnaires presenting a set of products out of which respondents pick the one they would purchase. This approach was established as CBC, which introduced competitive context and allowed to incorporate a 'None' alternative, thus more closely mimicking the purchase process. The results were analyzed at the population level without considering individual preferences until the 1990s, when Hierarchical Bayes (HB) estimation permitted analysis on both the group- and individual-level.

MBC was introduced as an extension of the CBC methodology to empower reliable analysis of customized products. From researchers' perspective the menu-choice presents serious challenges. Firstly, the number of outcomes drastically increases with the number of options in each task (Kamakura and Kwak, 2020). Secondly, the model should incorporate bundles of products offered at a discount price. Finally, the interrelated nature of demand for individual items due to substitutability or complementarity and the income effect has to be addressed. Ben-Akiva and Gershenfeld (1998) are the first to conceptualize an analytical approach for multi-choice experiments in a marketing application, emphasizing the need for a more realistic framing of choice tasks in order to minimize different types of response biases. They propose a two-level nested logit where they treat each choice alternative as a combination of selected items, resulting in 2^J possible outcomes for a menu with J items - approach later named as the Exhaustive Alternatives (EA) (Orme, 2011). The underlying assumption is that the respondent makes a trade-off between combinations considering all possible outcomes of the menu situation. In the upper level they obtain probability of a buy or no-buy decision, whereas the lower level represents the conditional probability of a particular feature combination, given a decision to buy. Ben-Akiva and Gershenfeld (1998) show that the specification is appropriate to menus including both *a la carte* products and bundles with a discount price.

Liechty et al. (2001) propose an alternative method employing a Bayesian approach with multivariate Probit (MVP) model, where the probability of choice is a function of intrinsic attractiveness, feature-specific effects and cross-effects of other features' characteristics. An advantage of the MVP specification is that it is more theoretically complete, as it allows to model correlations capturing cross-dependencies across the menu items. The model uses random effects formulation to accommodate customer heterogeneity in feature attractiveness and allows for real-world constraints in menu choices. Moreover, they include the first comparative study of underlying MBC methods. Building on the utility specification of Ben-Akiva and Gershenfeld (1998) they formulate a random-effects multinomial probit (MNP) model and test the predictive performance of MNP and two versions of MVP – with and without correlations between feature-specific utilities. The obtained empirical results suggest that MVP is the superior method quadrupling the hit rate of MNP and including the correlation structure slightly improves the predictive performance. However, a serious drawback of this study is low complexity of the menu task, which includes only six features rendering 64 potential outcomes, a number further reduced to only 20 portfolios after imposing menu constraints.

Cohen and Liechty (2007) discuss the results of the same study in a less technical manner, additionally describing a method that translates the menu-choice into a series of binary models that address a choice of each item separately. However, they do not analyze this technique arguing that the model would yield biased estimates, as it does not incorporate the interdependent nature of multiple choices. Nevertheless, Orme (2011) proposes a Serial Cross-Effects (SCE) model, an extension of the idea of Cohen and Liechty (2007), by interconnecting binary models via the cross-effects terms. The simplicity of this model translates into its high flexibility. It has an ability to tackle situations with more than just two mutually exclusive outcomes but first and foremost it can be used to break down very complex models and simplify their estimation. On the other hand, Orme (2011) emphasize that when applying this approach building many separate models can be inconvenient and prone to errors. Pruning the model of any non-significant effects or imposing utility constraints are the recommended solutions. Another disadvantage is that the technique does not formally recognize combinatorial outcomes, which can be detrimental to individual-level predictions.

In the same publication Orme (2011) carry out a comparative study on the predictive ability of three models: EA, SCE and a two-stage model (Bakken and Bond, 2004). The last method assumes that the choice of *a la carte* items is only made if all bundles are first rejected and is only appropriate for menu situations with such a prohibition. For each approach, estimation was performed by Aggregate Logit, Latent Class and HB (with and without covariates). The model fit and predictive power were measured by R-squared and Mean Absolute Error (MAE) using holdout tasks. Although, all methods performed very well estimation with aggregated logit surprisingly showed the best results. Orme (2011) found that SCE did slightly worse than EA and two-stage models. They recognized that this result might be different if the menu choice had no constraints, because SCE did not incorporate the logical exclusions in menu outcomes. In line with other research and expectations, incorporating covariates in HB estimation did not significantly improve the predictive accuracy.

The majority of research around MBC focuses on Logit-based methods, due to their popularity among practitioners, owing to a very convenient property of having a closed-form expression for its likelihood. As mentioned above, Liechty et al. (2001) are the first to propose a probit model as an alternative. Neuerburg (2015) compare the performance of Independent MNL (Cohen and Liechty, 2007), SCE and a multivariate multinomial Probit (MVMNP). The Probit-based method allows to obtain estimates in a single complex model and thus identify substitutive and complementary effects through correlations of error terms. To make the comparison comprehensive, they test the models on multiple synthetic datasets with varying characteristics, like menu complexity, and measure the predictive performance employing both combinatorial and item hit rate, where the former represents predictions of choice patterns. The results clearly show the Probit model to be inferior to Logit-based methods in terms of accuracy and computation time, giving support to the popular Logit-based techniques. One of the shortcomings of the study is rather low complexity of tested menu situations, where the most complex case translates to only 60 million possible combinations. Neuerburg et al. (2021) further extends this research by incorporating an EA model into the analysis and employing model fit and reservation price recoverability to better characterize the properties of the models. They discover that the attempt to model menu choices in a single overall model of high complexity does not improve the quality of predictions. The study also shown that the differences in hit rates between small (n=100) and large sample sizes (n=500) are rather small, which has important implications for applicability.

Some researchers' consider variations of the popular methods. Cordella et al. (2012) aim to address the main pitfalls of the EA and SCE models by employing a mixed approach named Choice Set Sampling (CSS) model. Firstly, in order to reduce the choice set of all possible combinatorial outcomes they implement Importance Sampling of Alternatives. Secondly, they complement the combinatorial model with binary logit tasks, where the choice of each feature for every respondent is modeled as a function of its own price effect only. The aim is to identify the potential individual price barriers for each feature regardless of other effects. Comparison of the performance of CSS and SCE provided further evidence, that conceptually simpler methods are superior, as the latter outperformed the former in all holdout tasks. Cordella et al. (2012) also pointed out that this conclusion is further supported both by the computational power and the time needed to construct the CSS.

Pfaff (2021) analyze three different models, with variations, aiming to identify the best predictive model using two datasets gathered using different menu designs. The simple design allows choice between individual items as well as bundles including some of them. The menu vs *a la carte* design incorporates a base product which can be customized by adding either individual items or predefined bundles. For the analysis, they implement a Probit-based SCE and a Logit-based EA, each as an aggregate model and random effects model. For the EA, they tackle the issue of large number of possible outcomes by limiting the number of considered combinations to a subset, which is constructed in two ways: including only 40 most frequently chosen combinations, and by utilizing the Stratified Importance Sampling (SIS). For the latter, the outcome space is split into disjoint strata based on the number of items in a combination, from which a fixed number of alternatives is drawn with equal probabilities. Additionally, building on the work of Liechty et al. (2001) on MVP, they propose an aggregate Multivariate Choice (MVC) model in which the latent utilities of all items are considered separately but are allowed to be correlated, thus capturing the interdependencies between menu items. In total they compare seven different specifications on two datasets and find that best predictive performance is obtained for EA with SIS, closely followed by MVC and aggregate SCE. However, the differences in performance between above models are marginal and there is not enough evidence to support one of them as superior. Pfaff (2021) discuss how each of those three methods might be considered the best depending on the aim of the study, preferred performance measures or underlying behavioural model. They point out that MVC is more informative then SCE, but only for menus with small number of items and the former is much more computationally demanding. On the other hand SCE does not account for choice restrictions, unlike the other two models. Such observations on model characteristics are what we aim to explore in depth and on a wider scope in this study.

Considering the growing importance and popularity of conjoint methods, especially MBC, it is important to mention the fundamental publications among the practitioner community. Orme (2020) and Orme and Chrzan (2017)) provide the guidelines which are the foundations for the commercial application of conjoint analysis. In a comprehensible read they cover the underlying theory, the full scope of different conjoint techniques, questionnaire designs, estimation methods with their pitfalls and best practices on how to avoid them. Orme (2019) extend this work to cover in detail MBC analysis. Although, the aim is to guide a researcher from designing the experiment and preparing the data files all the way to creating a market simulator for predictions using Sawtooth Software, the manual extensively discusses previous research on the topic as well as the most important findings. An important contribution is systemizing the distinction between Generic and Alternative-Specific effects formulations, where the former simplifies the sensitivity to price changes to be constant among menu items.

From the presented literature two main techniques for MBC emerge: EA which allows to model menu choices in a combinatorial manner and SCE which deconstructs the extensive choiceboard tasks into a series of binary models. We found a lot of evidence that more complex modeling approaches are inferior to these simpler concepts. The same conclusion holds when comparing Logit- and Probit-based models in most conditions, although the latter offers more theoretical coverage. As stated previously, MBC is not well analyzed and most of the comparative studies aim to find one universally superior method for predicting consumer choices (Orme (2011), Pfaff (2021)). However, recent research on generated synthetic data (Neuerburg et al., 2021) suggests that each approach has different properties, thus their performance depends on characteristics of the data, extensiveness of the questionnaire, underlying modeling assumptions, and the purpose of the study. In this paper, we aim to further explore such properties by varying the underlying MBC models, parameter formulations, linearity assumption on the functional form of the price effect and estimation techniques, thus comparing 28 configurations. We focus on investigating how these arrangements affect various performance measures and translate our observations into advantages and disadvantages, which can be used as guidelines for constructing MBC models in the future. We contribute to the literature by incorporating the Modular (MOD) model, an intermediate case combining the SCE and the EA, into our comparative analysis. Additionally, we use the Generic effects formulation for the own-price effects in a novel application assuming respondent-specific product-invariant price sensitivity across all combinations and submodels. Finally, the data is gathered utilizing a high complexity menu tasks allowing over 590 quintillion possible choice combinations, as the majority of comparative studies found in the academic literature is based on much simpler menu situations. This is done not only to more closely mimic menu situations encountered in the real marketplace but also to make our insights equally useful to academic scholars and practitioners.

3 Data

For the analysis we use a dataset gathered in the first quarter of 2022 in the Netherlands utilizing a dedicated experimentally designed Menu-Based Choice questionnaire. The sample includes respondents aged between 16 and 75 years old. We excluded professionals in fields such as market research, advertising and most importantly restaurant business, as they could have disproportional knowledge about the industry, thus not being an accurate representative of the general consumer population. Lastly, in order to not inflate the choice share of the *None* alternative, we preclude from participating people who declared to never visit the restaurant in question.

Each respondent was presented with 12 choice tasks in a form of a restaurant menu, which comprised of the same 69 products. We explicitly asked respondents to compose a meal for themselves and we limited their choices to a single piece per product. We treat a stand-alone product e.g. *Sandwich1*, the value meal including this product, *MealS1*, and enlarged value meal, *LargeMealS1*, as three separate products. Analogously, the same foods offered in different sizes are also treated as different items e.g *Drink1_1, Drink1_2* and *Drink1_3*.

The design includes three price points per product. The second price point is the base value, determined from the market prices offered by the restaurant in question in 2021. The first and third price points are calculated as 90% and 110% of the base case, rounded to €0.05 to accurately mimic the price format of the restaurant. All price points for all menu items have been presented in Table 7 in the Appendix, with large meals prices presented as additional costs added on top of to the standard size meal.

The majority of the respondents are presented with 9 random tasks and 3 holdout tasks: the first with all products offered at minimum price, second at base price and third at the maximum. Only the data from the 9 random tasks is used for the estimation. Every fifth respondent was assigned to the holdout sample, in which each respondent filled 12 fixed tasks. This out-of-sample data is used to obtain the out-of-sample predictive performance measures of the analysed models.

The raw dataset consists of 2160 respondents who completed all 12 tasks. We clean the in-sample and holdout data separately employing two rules. Firstly, we flag and discard the speeders - respondents who perform the conjoint questionnaire in an extraordinary short amount of time, thus assumed not to pay sufficient attention to the trade-offs between presented attributes (Orme, 2020). As a threshold we take 40% of the median of the time spent on the questionnaire across all respondents within a sample. Secondly, we discard respondents who did not select any of the products in 75% of the presented tasks. For both samples we clean out approximately 10% of the respondents and are left with 1937 respondents: 1542 in-sample and 395 out-of-sample. This translates into 23,244 choice tasks in total.

Product	Frequency	Choice share
Saus1	4365	18.78%
Side1_1	2902	12.48%
Side1_2	2259	9.72%
Snack1	2230	9.59%
Snack2	2190	9.42%
Drink1_1	1917	8.25%
MealS3	1818	7.82%
Kids'Meal	1731	7.45%
Dessert3	1677	7.21%
Sandwich3	1632	7.02%

Product	Frequency	Choice share
Dessert6	314	1.35%
MealSalad2	298	1.28%
LargeMealS1	295	1.27%
MealSalad1	283	1.22%
LargeMealF1_2	251	1.08%
LargeMealS11	231	0.99%
LargeMealS2	231	0.99%
LargeMealS6	193	0.83%
LargeMealS7	155	0.67%
LargeMealS9	153	0.66%

Table 1: Most and least chosen products.

We perform the preliminary investigation of the data employing the Counting Analysis, focusing on the frequencies of individual product choices, as well as the most commonly constructed combinations. Table 1 presents the most and the least popular items in the menu. In each choice task 2.52 products were chosen on average and all items were chosen by at least 150 respondents. As could be expected, the majority of the top products are side orders and small products which most likely are chosen in combination with other items. Although *Saus1* was chosen by approximately every fifth respondent obtaining the highest choice share, it is worth noting that *Side1_1* and *Side1_2* are the same product offered in two different sizes - small and medium respectively, which makes *Side1* the most popular food choice. The most popular meal option is the *MealS3*, containing *Sandwich3* also present among the top 10, which leads to a conclusion that it is the favorite among larger menu items. Interestingly, the majority of the least popular products are large value meals indicating that most of the consumers decide not to enlarge their meal options.

Figure 1 presents the product frequencies aggregated based on subsections of the menu, a separation used in one of the modelling approaches utilized in our analysis. *Sandwiches* accumulate most of the chosen products, however the sum of choices captured by two value meal categories, *Meals1* and *Meals2*, is still greater suggesting that most of the consumers prefer making use of the preconfigured rather than composing their meals from individual items. These three subsections are also the ones which include the largest number of products per category, as can be seen in Table 10 in the Appendix. As anticipated, the category with the lowest choice frequency is the *Kids'Meal*, which seems logical because the surveyed respondents were all above 16 years old and because this subsection includes only a single menu item.

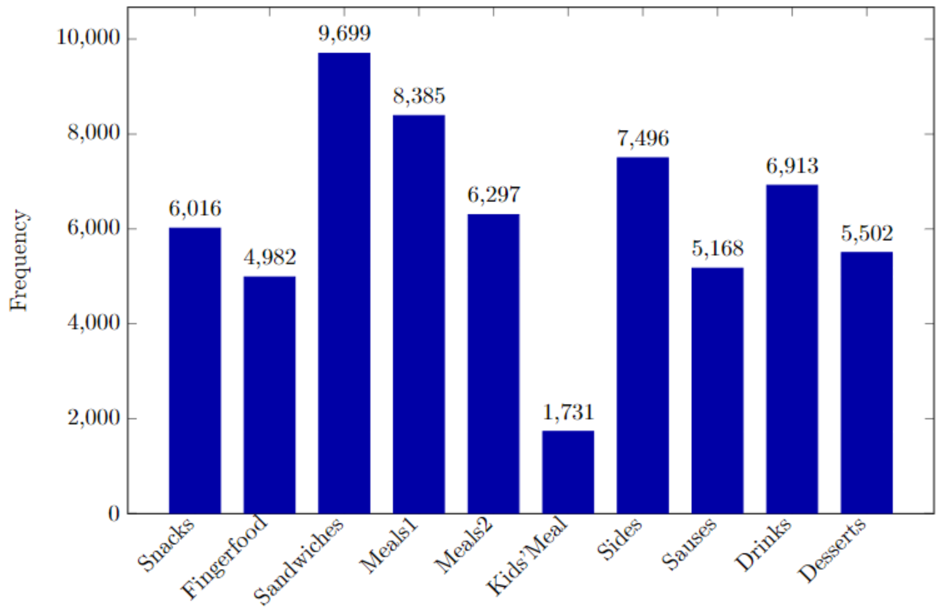

Figure 1: Choice frequencies aggregated by menu subsections.

Subsequently, we investigate the combinations of products chosen in individual tasks. We focus on analyzing the counts of the 50 most popular out of 6,735 different combinations configured by the respondents. The frequencies for the top ten combinations are presented in Table 2, which translates to only 15.69% of all choices. Unsurprisingly, the highest choice share is obtained for the *None*, a combination in which the respondent did not choose any of the menu items. One of the explanations could be that this alternative is usually chosen when the cost of the preferred meal of the respondent exceeds the amount they are willing to spend. Additionally, Orme and Chrzan (2017) suggest that the frequency of *None* choices is positively affected by the task difficulty, which in our case translates to menu complexity. The remaining nine combinations include only a single menu item, which suggests that the majority of respondents is satisfied with ordering a single menu item. The *MealS3* combination is the second most popular, with the other two configurations containing *Sandwich3*-based products also present among the most popular choices. It is also interesting to look into multi-product combinations. Among the 50 most chosen combinations only three contained more than one product, with the leading configuration positioned only at the 25th place. Table 3 showcases, that two of these combinations include a value meal complimented with a side *Saus1*. The remaining composition comprises of two smaller products, which might be an approach taken by some consumers that omit purchasing value meals, but are still unfulfilled with selecting just a single item.

Combination	Frequency	Choice share
None	999	4.30%
MealS3	518	2.23%
MealS4	346	1.49%
Kids'Meal	330	1.42%
LargeMealS3	293	1.26%
Sandwich3	246	1.06%
MealS8	244	1.05%
Snack4	232	1.00%
Snack2	223	0.96%
Snack1	216	0.93%

Table 2: Top ten most chosen product combinations.

Combination		Frequency	Choice share
MealS3	Saus1	102	0.44%
Snack2	Snack3	72	0.31%
MealS4	Saus1	53	0.23%

Table 3: Multi-product combinations present among the 50 most chosen.

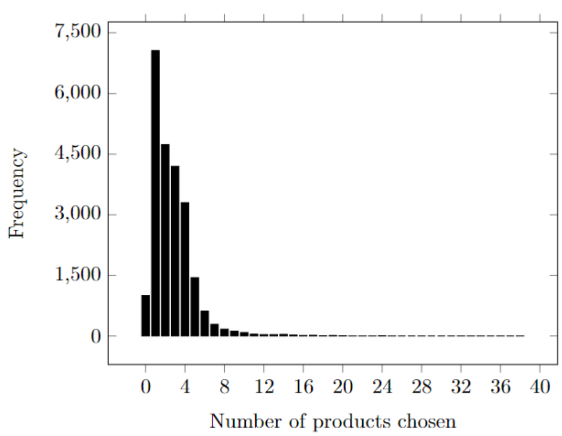

Figure 2: Frequency of number of products chosen per task

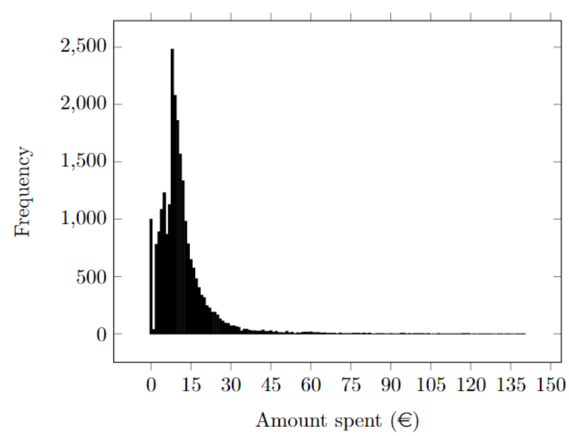

Figure 3: Frequency of the amount in € spent per task

Besides the choice frequencies, we analyze other information contained in the data. Firstly, we observe the number of products chosen in each task, presented in Figure 2. as mentioned before, on average 2.52 products were picked per task. The minimum number of items selected in a single task was zero and the maximum was 38 menu items. In line with the above results, single item choices were the most popular. The majority of respondents chose up to 5 products per task. The choices above nine items seem unrealistic as we explicitly asked the respondents to compose a meal only for themselves, however they only total to 1.38% of all tasks. Secondly, we use a similar approach to investigate the amount of money spent in each task, visualized in Figure 3. The average meal cost was €11.12. The minimum spending was €0 and the maximum was €140, which again is a quite unrealistic. However, it can be seen that the majority of respondents aimed to spend around €10 per meal, with orders worth approximately €8 being the most popular.

In the subsequent section we discuss methods used to model the in-sample choice data, as well as the simulation process allowing us to make predictions about the product choice shares and the individual choices of respondents. The holdout sample will be used to validate the predictive performance of these approaches.

4 Methodology

This chapter considers the methods utilized to analyze the gathered data and is separated into four subsections. Firstly, in Section 4.1 we introduce the multinomial logit (MNL), which is the cornerstone of all presented modelling approaches. Secondly, in Section 4.2 we discuss the analyzed models for conceptualizing MBC situations. Next, Section 4.3 presents two estimation methods: the Aggregated Logit and the Hierarchical Bayes. The final Section 4.4 showcases different performance measurements, used to obtain the predictive performance and to uncover the properties of different techniques.

4.1 Multinomial Logit

Conjoint analysis is a method exploring tradeoffs in the decision behavior and thus needs an appropriate statistical model able to capture these effects. Consider a choice situation, where an individual assigns a latent utility u_{ik} to category k based on perceived properties of X_i, a $1 \times (P+1)$ matrix of the intercept and explanatory variables measuring the observed difference of attributes between category k and a base category (in conjoint task usually none alternative). Moreover, we assume that this latent utility is formulated as a function of X_i and is linear in parameters:

$$u_{ik} = \alpha_k + \beta_{k1} x_{i1} + \ldots + \beta_{kP} x_{iP} + \varepsilon_{ik}. \tag{1}$$

We define the discrete choice y_i as a random variable Y_i, which can take K discrete values and assume the multivariate Bernoulli distribution, which implies the choice probability of category k as $\mathbb{P}[Y_i = k] = \pi_k$, $k = 1, \ldots, K$, with $\pi_1 + \pi_2 + \ldots + \pi_K = 1$ (Franses and Paap, 2001).

Assuming that the error terms of latent utilities ε_{ik} are identically independently distributed and follow an extreme value distribution allows us to utilize the Multinomial Logit model, first proposed by McFadden (1973):

$$\mathbb{P}(Y_i = k \mid X_i) = \frac{exp(u_{ik})}{\sum_{l=1}^{K} exp(u_{il})}. \tag{2}$$

As the category probabilities have to sum to 1, a base category needs to be assigned by restricting its parameters to zero for identification purposes. The choice for the base category does not change the effect of explanatory variables on choice, however typically $\beta_K = 0$ (Franses and Paap, 2001). For $K = 2$ the model reduces to a binomial Logit.

4.2 Models

In this section we introduce three main underlying modelling approaches based on the Multinomial Logit theorem. We separately discuss each of the techniques in detail and describe how we formulate their different variations.

4.2.1 Serial-Cross Effects

The SCE approach converts the menu situation into a series of binary choice models, which predict the likelihood of different products being selected or not. To account for the inter-dependencies between menu items, the separate models are interconnected through the cross-effect terms. We construct a Binomial Logit model for each menu choice, which is made by the consumer when the difference between the preference for the considered product and the base case is positive, thus the dependent variable y_{itj} depends on the continuous latent utility. This rule is quantified by

$$Y_{itj} = \begin{cases} 1, & \text{if } u_{itj} > 0 \\ 0, & \text{if } u_{itj} \leq 0, \end{cases} \tag{3}$$

where Y_{itj} denotes the decision of respondent $i = 1, \ldots, N$, in menu task $t = 1, \ldots, T$, to select item $j = 1, \ldots, J$. Following from (2), the probability of such choice is modeled by

$$\mathbb{P}(Y_{itj} = 1) = \frac{exp(u_{itj})}{1 + exp(u_{itj})}. \tag{4}$$

For the SCE model the latent utility, u_{itj}, is based on the intrinsic desirability, price of the product, P_{itj}, as well as prices of different products in the menu, P_{itl}. We consider utility specifications which differ on the linearity assumption, thus formulation of the own-effects.

Under linearity, we have the Alternative-Specific effects, which follow from the belief that respondents price sensitivity changes from one product to another:

$$u_{itj} = \alpha_{ij} + \beta_{ij} P_{itj} + \sum_{l=2}^{J} \beta_{ilj} P_{itl} + \varepsilon_{itj}. \tag{5}$$

Individual inherent attraction is denoted by α_{ij}, β_{ij} is the individual sensitivity to the price of product j and β_{ilj} represents the individual cross-price effect of product l on product j. The Gumbel distributed error term is denoted by ε_{itj}.

We also consider a non-linear (part-worth) functional form for the own-price effect, which may arise due to respondents having some psychological thresholds not complying to linear effects of price (Orme, 2019). If this is true, the linear formulation may smooth away important insights or hinder the accuracy. The aim is to independently estimate the utility at each discrete price point using $W-1$ effects for a W-level independent variable. In this case the reference level is set to the first level of the attribute. For $w = 1, \dots, W$ price points we have Alternative-Specific effects:

$$u_{itj} = \alpha_{ij} + \sum_{w=1}^{W} \beta_{ijw} P_{itj} \mathbb{I}[P_{itj} = w] + \sum_{l=2}^{J} \beta_{ilj} P_{itl} + \varepsilon_{itj}, \tag{6}$$

where $\mathbb{I}$ is an indicator function which takes value of 1 when considered price takes the value corresponding to the w price point.

When applying the SCE approach a crucial step is to prune the model. Applying (5) or (6) in their full specification, although theoretically complete, would incorporate $J-1$ cross-effects, which could lead to potential overfitting as well as an enormous number of parameters to estimate, especially for very complex MBC tasks. Additionally, inclusion of insignificant variables, which usually constitute to 75% of all cross-effects, is often detrimental to predictive ability of new scenarios (Orme, 2019). Different variable selection methods were studied by Dippold-Tausendpfund and Neuerburg (2018), who found that χ^2-test performs well under different data scenarios, better than more advanced methods they considered. Under null hyphotesis we test if two variables, a dependent choice variable and a explanatory price variable, are independent. The significance level is chosen at 5%, thus all independent variables with a corresponding p-value larger then this threshold can be discarded from the model. Otherwise, we have statistical evidence supporting the significance of the given cross-effect.

4.2.2 Exhaustive Alternatives

The EA approach allows to analyze the menu situation in a single model by assuming that the respondent considers all possible ways that this choice task can be completed and selects the one corresponding to the highest latent utility. In this context, the discrete choice alternatives are different combinations of menu items with an associated total price. Thus, for a menu task with J items there are 2^J possible configurations of the products. The biggest advantage of the EA approach is that it formally recognizes and predicts combinatorial outcomes from menu tasks, making it a more complete model of consumer choices.

For this model we have a categorical response variable Y_{it}, which denotes the configuration of menu items chosen by respondent $i = 1, \ldots, N$, in menu task $t = 1, \ldots, T$. Y_{it} can take any value k corresponding to the combination c_k from the set of possible item combinations $C = c_1, \ldots, c_K$ where $K = 2^J$. The choice probability stems from (2), resulting in

$$\mathbb{P}(Y_{it} = k) = \frac{exp(u_{itk})}{\sum_{l=1}^{K} exp(u_{itl})}, \tag{7}$$

where the restricted base case is the combination including none of the menu items ('None' alternative). In order to obtain the probability of item j being chosen in task t, the probability shares across the combinations including product j need to be summed.

$$\mathbb{P}(Y_{itj} = 1) = \sum_{k=1}^{K} \mathbb{P}(Y_{it} = k)\mathbb{I}\big[j \in c_k\big], \tag{8}$$

where the indicator function $\mathbb{I}$ takes the value of 1 only for combinations containing product j.

For the EA model, the utility u_{itk} is modeled by the intrinsic attractiveness of combination c_k, α_{ik}, and prices of items included in the combination, P_{itj} where $j \in c_k$. We consider different utility specifications, varying them on the linearity assumption and parametrization of the own-price effect.

For the linear case, the utility for a particular combination of menu items under Alternative-Specific effects takes the form

$$u_{itk} = \alpha_{ik} + \sum_{j \in c_k} \beta_{ij} P_{itj} + \varepsilon_{itk}. \tag{9}$$

Here, price sensitivity differs between the products and is denoted by β_{ij} and the error term following the extreme value distribution by ε_{itk}.

Analogously as for the SCE, we also allow for a non-linear functional form, which for the Alternative-Specific effects case with $w = 1, \ldots, W$ price points is denoted by

$$u_{itk} = \alpha_{ik} + \sum_{j \in c_k} \sum_{w=1}^{W} \beta_{ijw} P_{itj} \mathbb{I}[P_{itj} = w] + \varepsilon_{itk}. \tag{10}$$

The second parameter form is the Generic effects, where we generalize (9) assuming that the sensitivity to price changes is unrelated to the product: $\beta_{ij} = \beta_i$ (Orme, 2019). In order to capture this effect across all combinations and prevent the number of estimated parameters from exploding (especially for the non-linear case and the MOD approach) we employ piecewise price coding (Orme and Chrzan, 2017). This approach allows to approximate non-linear function by a flexible function consisting of linear segments, which are defined based on intervals predetermined by the researcher. Firstly, the continuous price variable is recoded to S variables, which is the number of slopes utilized in the function. This is denoted by

$$P_{itjs} = \begin{cases} 0, & \text{if } P_{itj} \leq z_s \\ 1 - \frac{z_{s+1} - P_{itj}}{z_{s+1} - z_s}, & \text{if } z_s < P_{itj} \leq z_{s+1}, \\ 1, & \text{if } P_{itj} > z_{s+1}, \end{cases} \tag{11}$$

where $s = 1, \ldots, S$ stands for the particular slope variable with its interval limited by cut points z_s and z_{s+1}. For the first and last variable the smallest and largest cut points respectively are also end points for the function. The choice of the exact number and value of the cut points is arbitrary, thus quite challenging. However, ideally they should capture the consumers' psychological price thresholds and other points for which there is a significant change in the preference slope. It is important to keep in mind to include enough price points in each interval to ensure that the estimation of all parameters is feasible. Also, the cut points should not overlap and they should not be defined to close to one another.

For our linear version of Generic effects we do not define any cut points which translates to recoding all independent price variables into a single variable, $S = 1$. To approximate the non-linear case well we decide to incorporate $S = 24$ segments into our piecewise function. The utility function for both cases is denoted by

$$u_{itk} = \alpha_{ik} + \sum_{s=1}^{S} \beta_{is} \sum_{j \in c_k} P_{itjs} + \varepsilon_{itk}, \tag{12}$$

where β_{is} is the individual product-invariant price sensitivity in price segment s.

As suggested by Orme (2011), the EA approach works well in practice when the number of possible ways in which the respondents can complete the menu task is relatively small – up to 36 combinations (Orme, 2019). This conclusion is also coherent from the theoretical perspective, as it would be unreasonable to assume that the respondent can accurately consider hundreds of possibilities giving equal attention to each choice. For a large number of possible outcomes the model can become sparse at the individual level, leading to overfitting. To resolve this issue, we restrict subset C to include only most chosen combinations of the products based on the preliminary Counting Analysis. The threshold separating combinations included and excluded from the model is defined based on the cumulative distribution of the information held among all combinations. The aim is to obtain a balance between model complexity and the percentage of all choices made represented by included combinations. In order to reduce the size of the problem, Orme (2011) postulates to discard from the model all combinations which were never chosen by respondents as a preliminary step. This is equivalent to assuming that these have a likelihood of choice equal to zero. The inability to model all possibilities combined with lack of clear guidelines for selecting the most appropriate subset C is one of the biggest disadvantages of the EA approach.

4.2.3 Modular

Finally, we employ the Modular (MOD) model approach, where we combine the SCE and the EA models. We divide the menu task into separate smaller EA models in which we consider combinations of items within a specified subsection, however in order to address inter-dependencies of the menu choices we also include cross-effects in the latent utility estimation. Although this approach is not theoretically as complete as the EA, it offers a compromise between disintegration into binary models and a single model, resolving the concern of model becoming overly complex and the risk of overfitting.

We collapse binary outcomes for every product into M dependent variables representing all feasible combinations of items in each menu subsection. Following from (7), we obtain $m = 1, \ldots, M$ MNL submodels with $C_m = c_{m1}, \ldots, c_{mK}$ feasible combinations in each module.

The latent utility for respondent $i = 1, \ldots, N$ is modeled by the inherent desirability of combination c_{mk}, represented by α_{imk}, prices of items included in the considered combination P_{itj} for $j \in c_{mk}$ and prices of products belonging to other modules P_{itl} for $l \in c_{m'k}$, $m' \neq m$.

Firstly, under assumption of linearity we consider the Alternative-Specific effects for both the price effects within the module, as well as for the cross-price effects.

$$u_{itmk} = \alpha_{imk} + \sum_{j \in c_{mk}} \beta_{ij} P_{itj} + \sum_{l=1}^{J/j \in c_{mk}} \beta_{ilk} P_{itl} + \varepsilon_{itmk}, \tag{13}$$

where β_{ij} is the price sensitivity of product j belonging to the considered module m, β_{ilk} is the cross-price effect of product l belonging to subsection $m' \neq m$ on combination c_{mk} and ε_{itmk} represents the error term. For the part-worth approach we adjust the own-price effect in the utility specification analogously to (6) and (10).

Secondly, for the Generic effects case we utilize the piecewise price function similarly as for the EA. The price of the considered combination is recoded into S variables, however we keep the cross-price effects of products from different modules in the Alternative-Specific form:

$$u_{itmk} = \alpha_{imk} + \sum_{s=1}^{S} \beta_{is} \sum_{j \in c_{mk}} P_{itjs} + \sum_{l=1}^{J/j \in c_{mk}} \beta_{ilk} P_{itl} + \varepsilon_{itmk}, \tag{14}$$

where β_{is} is the individual product-invariant price sensitivity in price segment s, and β_{ilk} is the cross-price effect of product l belonging to subsection $m' \neq m$ on combination c_{mk}. For the linear case $S = 1$ and under non-linear assumption we employ $S = 24$ price intervals.

Similarly to SCE the models need to be pruned of all the insignificant cross-effects to avoid the risk of overfitting.

The separation of the menu items into smaller subsections is performed as suggested by Orme (2019). We separate the items based on logical sections of the menu task. An example of this would be dividing a fast food restaurant menu task into segments including sandwiches, side dishes, desserts, drinks and value meals. We represent each of these modules by a single EA submodel.

4.3 Estimation methods

For every model specification presented in Section 4.2 we obtain the estimates of the model parameters by employing each estimation method presented in this chapter.

4.3.1 Aggregated Logit

The first estimation method, the Aggregated Logit, aims to obtain a vector of pooled utilities that yields the best fit to the data. The technique treats all responses across tasks and individuals as homogeneous - as if coming from a single respondent. In the context of conjoint analysis, the method has proven robust and flexible and allows for results which deliver excellent population-level predictions. However, due to heterogeneity typically present among respondents it provides low fit to the data, with McFadden's $\rho^2 \leq 0.25$ (Orme and Chrzan, 2017). In practice, the estimation of model parameters with the Aggregated Logit takes substantially less time than with the HB and delivers statistical tests for model pruning and selecting the most appropriate functional form of independent variables. For this reasons, a common practice is to estimate the Aggregated Logit to specify the most appropriate model before obtaining the results on individual-level with HB.

Model parameters are estimated utilizing the Maximum Likelihood (MLE) method. For our models we have the likelihood function

$$L(\beta \mid X, y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{k=1}^{K} \mathbb{P}(Y_{it} = k \mid X_{it})^{\mathbb{I}[y_{it}=k]}, \tag{15}$$

where β represents model parameters. The logarithm of the likelihood function is

$$l(\beta \mid X, y) = \sum_{i=1}^{N} \sum_{i=t}^{T} \sum_{k=1}^{K} \mathbb{I}[y_{it} = k] \log \mathbb{P}(Y_{it} = k \mid X_{it}). \tag{16}$$

For the SCE model (15) takes the f The MLE is the parameter value $\hat{\beta}$ corresponding to the maximum of the log-likelihood function over the parameter space, which can be obtained by solving the first-order derivative. However, because the log-likelihood function is nonlinear this solution cannot be obtained analytically (Franses and Paap, 2001).

We employ the Newton-Raphson optimization algorithm, a gradient search procedure which iteratively seeks the first-order condition for a maximum. We obtain the estimates by iterating over

$$\beta_h = \beta_{h-1} - H(\beta_{h-1})^{-1} G(\beta_{h-1}), \tag{17}$$

where h is the current iteration, $G(\beta)$ is the gradient (first-order derivative of the log-likelihood function) and $H(\beta)$ is the Hessian matrix (the second-order derivative if the log-likelihood function). Because the log-likelihood is globally concave, the algorithm will converge to the global optimum for any starting values. We set initial $\beta = 0$. The obtained estimator $\hat{\beta}$ is asymptotically normally distributed $\hat{\beta} \sim N(\beta, (-H(\hat{\beta})^{-1})$ (Franses and Paap, 2001). For conjoint studies, the algorithm usually converges within six iterations.

4.3.2 Hierarchical Bayes

The second estimation technique we consider is the Hierarchical Bayes, which allows to obtain the estimates at the individual-level, thus capturing heterogeneity in consumer preferences. Unlike in the classical methods, where we assume a particular model with specified parameters and analyze the consistency of the data with those assumptions (probability distribution of the data), Bayesian theorem treats data as given and assumes a certain model which parameters are the unknown quantity. This approach allows to regard parameters as random variables and assigning them probability distributions.

Bayesian estimators utilize the information contained in the data, y, with a prior belief concerning the parameter distribution before seeing the data, $\pi(\beta)$. By updating the prior distribution with the information from the data we obtain the posterior distribution of β:

$$\pi(\beta \mid X, y) \;\propto\; l(\beta \mid X, y)\, \pi(\beta), \tag{18}$$

where $\pi(\cdot)$ denotes a prior, $\pi(\cdot \mid X, y)$ the posterior density function of parameters and $l(\cdot)$ is the log-likelihood function given in (16). It is important to note, that both the form and the hyperparameters of the prior can have a major impact on determining the posterior distribution, however this influence diminishes as the sample size increases (Greenberg, 2013, p. 17).

For our application, we have a two level model. At the upper level, we assume that i-th respondent's part-worths, β_i, come from a multivariate normal distribution with population parameters, mean vector γ and covariance matrix D. At the lower level, given respondent's part-worths, we obtain the probability of choosing particular alternatives which is assumed to come from a MNL model. Formally, we have

$$\begin{aligned} &\text{upper model:} && \beta_i|\gamma, D \sim \mathbb{N}(\gamma, D), \\ &\text{lower model:} && \mathbb{P}(Y_{it} = k \mid X_i, \beta_i) = \frac{exp(u_{itk})}{\sum_{l=1}^{K} exp(u_{itl})}. \end{aligned} \tag{19}$$

The hierarchical structure of the model comes from two stages of priors, where the first-stage prior, $\pi(\beta) = \pi(\beta|\gamma, D)$, is proper and parameterized by second-stage priors, $\pi(\gamma)$ and $\pi(D)$, which are conjugate for a multivariate normal distribution. The posterior results for the three parameters are obtained through a simulation algorithm - an iterative process known as Gibbs sampling, summarized in (20). First, given present values of β_i and D we draw γ from a multivariate normal with population parameters. Secondly, given current β_i and the updated γ, we draw D from the Inverted Wishart distribution. Using the updated parameters from previous steps, β_i is drawn employing the Metropolis Hastings algorithm, in which betas are updated in consecutive iterations based on the previous value, providing better fit to the data until convergence.

$$\begin{aligned} &1.\ p(\gamma_{h+1} \mid \beta_{i,h}, D_h) && \pi(\gamma) \sim \mathbb{N}(\bar{\beta}, \tfrac{1}{N} D) \\ &2.\ p(D_{h+1} \mid \gamma_{h+1}, \beta_{ih}) && \pi(D) \sim IW(\nu, \Lambda_\gamma) \\ &3.\ p(\beta_{i,h+1} \mid \gamma_{h+1}, D_{h+1}) && \text{Metropolis Hastings Algorithm,} \end{aligned} \tag{20}$$

where h represents the iteration, $\bar{\beta}$ is the population average, N is the number of respondents, $\nu = N+(P+1)$ are degrees of freedom and Λ_γ is a positive definite $(P+1)$ dimensional matrix obtained by incorporating the prior information with the current estimates of γ and β_i (Alvarez et al. (2014), Sawtooth Software (2021)). As suggested by Orme and Chrzan (2017), we choose the following initial values for the parameters in the Gibbs sampler: $\beta_{i,0} = 0$, $\gamma_0 = 0$, $D_0 = \mathbb{I}_{P+1} = diag(1, \ldots, 1)$.

Drawing the part worth utilities for each respondent is done through the Metropolis Hastings algorithm. We begin with the current estimate of individual's part worths, $\beta_i^{(m)}$, and generate a candidate estimate, β^*, by drawing a random vector d from a jumping distribution, in our case a normal with mean zero and covariance matrix proportional to D. This distribution determines the size of the random jump from $\beta_i^{(m)}$ to $\beta^* = \beta^{(m)} + d$. This candidate is accepted or rejected with a certain probability δ depending on whether it improves the estimate, which is determined based on the ratio of posterior probabilities of two estimates $\beta_i^{(m)}$ and β^*, given the current values of α, D and the data. In case of rejection, the estimate from the previous iteration is used. Rigorously, the algorithm proceeds as follows:

Step 1. Specify the starting value $\beta_i^{(0)} = 0$ and set $m = 0$.

Step 2. Simulate β_i^* from:

$$\begin{aligned}
&\text{Set } \beta_i^{(m+1)} = \beta_i^* \text{ with probability } \delta.\\
&\text{Set } \beta_i^{(m+1)} = \beta_i^{(m)} \text{ with probability } 1-\delta,\\
&\text{where } \delta = min\left(\frac{f(\beta_i^* \mid \gamma, D)\prod_{k=1}^{K}\mathbb{P}(Y_{it}=k \mid X_i, \beta_i^*)}{f(\beta_i^{(m)} \mid \gamma, D)\prod_{k=1}^{K}\mathbb{P}(Y_{it}=k \mid X_i, \beta_i^{(m)})}, 1\right),\\
&\text{and } f(\beta_i \mid \gamma, D) = \exp(-\frac{1}{2}(\beta_i-\gamma)' D^{-1}(\beta_i-\gamma)).
\end{aligned} \tag{21}$$

Step 3. Set $m = m+1$, and go to step 2.

γ, D are the current draws from (20), $f(\beta_i \mid \gamma, D)$ is the relative density of the distribution of β_i serving as a prior in the Bayesian updating, and the probability of an alternative k is calculated according to the Logit model in (19) (Gelman et al., 1995). To ensure convergence, we allow the algorithm to run through m^* burn-in iterations from which the results are not saved or used. Afterwards, we assume that the process has converged and we can use the simulated part worths from subsequent iterations ($m \geq m^*$) as a sample from the distribution of β_i. Based on these sampled draws we calculate the posterior mean, $\hat{\beta}_i$, which is later used to calculate the choice probability.

Determining the appropriate burn-in sample size, m^*, is not straightforward as there are no universal theorems to indicate how large should it be. Usually this value is set to several hundred or thousands iterations. As suggested by (Greenberg, 2013, p. 107), we opt to investigate the trace, a plot of the sampled values, and asses if for the iterations around the chosen m^* show variation around a central value. If this is the case we assume that the process has converged.

We consider two approaches towards obtaining the draws. Firstly, we simply save the required number of draws for iterations after convergence. Secondly, following suggestion of Orme (2019) we employ draws trimming by incorporating a skip factor which allows to compensate for the consecutive draws of β_i not being independent, thus capturing the parameter distribution more completely. The dependence among draws can be detrimental to the precision of inference performed on these draws (Sawtooth Software, 2021). For a skip factor of q we save and use only the draws from every q-th iteration.

4.4 Performance Measures

After obtaining the parameter estimates we transform them into predictions by inserting obtained $\hat{\beta}_i$ ($\hat{\beta}_i = \hat{\beta}$ for the Aggregated Logit) into (2), thus calculating the choice probability of each combination. Because we aim to produce the predictions on menu item level we transform the combinatorial predictions into product choice probabilities by summing the forecasted shares across the combinations containing the product in question, as showcased in (8). For the binary models we input the estimated parameters into (4), thus directly obtaining the probabilities on the item level. We denote the probability of product j being chosen in task t by respondent i by $\hat{y}_{itj} = \mathbb{P}(\hat{Y_{itj}} = 1)$.

To asses the predictive validity of different models we employ two most popular measures used for conjoint methods: Mean Absolute Error (MAE) for out-of-sample predictions and the hit rate for in-sample validity.

4.4.1 Mean Absolute Error

The MAE allows to compare the ability of a particular model to reproduce the observed choice shares on an aggregate level. The measure is calculated as a arithmetic average of absolute errors:

$$MAE = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} |\hat{y}_{itj} - y_{itj}|}{N \times T \times J}, \tag{22}$$

where $\hat{y}_{itj}$ is the predicted choice probability and y_{itj} is the observed choice share. The main advantage of the MAE is that it weights all of the errors on an equal scale, which means that the outliers will not be additionally penalized. Thus, it allows to evenly measure the model's predictive performance. We aim to minimize the MAE, as it indicates higher predictive validity on an aggregate level.

We approach the MAE results in three ways. Besides reporting the overall model score we also obtain the metric for subsections of the menu items. Firstly, based on the Counting Analysis we divide the products into three groups based on their frequency of choice. The first group will contain the most often chosen items (a third of all items) and the third group will contain the rarest choices. We aim to explore if there are differences between the models' performances depending on the frequency buckets. Secondly, we apply the same separation of the menu into modules as we use for the MOD approach and obtain the MAE for each individual subsection. Such decomposition might provide additional insights into the behaviour of different configurations depending on the characteristics of menu items.

4.4.2 Hit rate

The hit rate is a measure of the predictive ability of conjoint methods to predict individual responses to holdout tasks. To prevent determining a hard threshold for defining and recording hits, we average the predicted choice shares for products chosen in the holdout tasks:

$$H = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_h} \sum_{j=1}^{J} \mathbb{I}[Y_{itj} = 1] \hat{y}_{itj}}{N \times T_h \times J \;\; \mathbb{I}[Y_{itj} = 1]} \times 100, H \in [0; 100], \tag{23}$$

where T_h denotes the holdout tasks and $\hat{y}_{itj}$ is the obtained choice probability. This specification allows to better capture the accuracy of the model without being influenced by arbitrary cut-off values chosen by the researcher.

High value of the hit rate indicates high predictive validity. We consider the overall individual level hit rate calculated based on the item choices (Neuerburg et al. (2021), Orme (2020)), followed by a hit rate obtained only based on tasks where more than five products were selected simultaneously. This will allow to explore possible differences in predictive power between choices incorporating less products which make up the majority of cases, as presented in Figure 2, and purchases including many items.

4.4.3 Average number of products estimated

Comparing the average number of menu items that the models predict to be chosen in every task allows to reveal if the technique over- or underestimates the results. To obtain this value we simply aggregate the predicted individual choice shares for all products in a single task. Subsequently, these scores are averaged across all respondents and holdout tasks. This is denoted by:

$$ANP = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_h}\sum_{j=1}^{J}\hat{y}_{itj}}{N \times T_h \times J}. \tag{24}$$

If the reported value is higher than the average number of products per tasks calculated from the data we conclude that the model overestimates. If this value is lower we are dealing with underestimation.

4.4.4 Estimation time

An important aspect of investigating the advantages and disadvantages of different methods is to compare the average computational time. It is important to note that the estimation time may vary between the computational units.

4.4.5 Data fit

We also compare the different models by analyzing the goodness of fit to the data. We employ McFadden's ρ^2, also called Percent Certainty, which is a pseudo R-squared measure. It captures how much better the solution is than for the null log-likelihood expected by chance, in comparison to a perfect solution with the log-likelihood equal to zero. The null corresponds to an uninformative vector of zero utilities. Percent Certainty is calculated by the equation

$$\rho^2 = \frac{l(\hat{\beta} \mid X, y) - l(\beta_{NULL} \mid X, y)}{-l(\beta_{NULL} \mid X, y)}. \tag{25}$$

An advantage of McFadden's ρ^2 is its straightforward interpretation which does not depend on the number of alternatives in a choice task, as opposed to root likelihood (RLH), a popular alternative for measuring fit to the data (Orme and Chrzan, 2017).

5 Results

We estimate and compare the performance of 28 models utilizing three variants of the Sawtooth Software. For the Alternative-Specific effects models we employ Sawtooth Software Menu-Based Choice software version 1.1.1. Because our formulation of the Generic effects is a novel approach it is not supported by the existing MBC solution, thus we utilize the Sawtooth Software CBC/HB System version 5.5.5 for the HB estimations and the Sawtooth Software Latent Class Module version 4.7.0 for the Aggregated Logits. For the HB estimations we use 20000 burn-in iterations, as suggested for MBC models by Orme (2019), and 200 draws saved after convergence. Thus, we perform 20200 iterations for the HB without the skip factor and 40000 for the estimations including the skip factor which is set to 100. The estimations are performed using hardware utilizing a Intel Core i5-4440 3.10GHz processor with four cores and 16GB RAM.

For every model investigated, the utility estimation is followed by construction of a market simulator, which allows to calculate the product choice share conditional on the menu item prices, as described in Section 4.4. We obtain product choice share predictions for the out-of-sample data, as well as the individual choice predictions for the in-sample holdout tasks.

5.1 Model implementation

Serial-Cross Effects

For the SCE approach we consider a linear and non-linear version, each estimated with three different estimation methods which translates to six models in total. We implement the approach by specifying 69 binary models, one for every menu item. As described in Section 4.2.1, for each product choice we investigate which of the 68 cross-price effects are significant and should be included in the model by performing the χ^2-test at 5% significance level. On average 12 variables are found to be significant, with the most extensive model for *Side1_2* including 24 and the most pruned having only 5 cross-prices affect the choice of *LargeMealS1*.

Exhaustive Alternatives

We implement the EA approach with two different effects formulations, Alternative-Specific and Generic, both under the assumption of linearity and non-linearity estimated with three methods. Thus, we obtain 12 different configurations for this method.

The preliminary investigation of the data, detailed in Section 3, uncovers there are 6,735 observed combinatorial choices across our sample, out of 2^{69} ($\simeq$ 590 quintillion) possible combinations. In order to identify the combinations to be included in the modelling, we plot the cumulative sum of choice information captured by combinations ordered from most to least frequent, which is presented in Figure 4. Looking at the distribution, the point where the plot begins to flatten, corresponding to 55% of the cumulative sum, would be the preferred cut-off for the combinations to be included under the criterion of retaining information. However, by selecting this level we would obtain a model of unreasonably high complexity including 422 combinations, which exceeds the suggested number of combinations by more than a tenfold (Orme, 2019). We opt to include 128 combinations translating to 42% of the information being captured by the model, which is visualized in Figure 4 by a vertical line. This cut-off value seems to provide much better balance between the model complexity and encapsulated information. The included combinations comprise of the *None*, 68 individual product choices, with *Drink2_2* being the only product excluded, 42 choices including two menu items, 11 including three items and six four-product combinations.

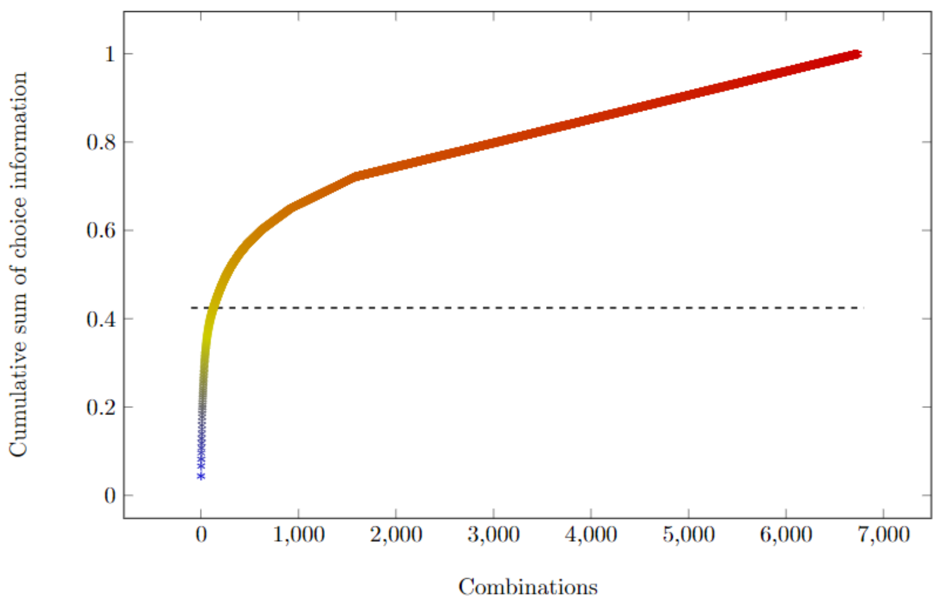

Figure 4: Cumulative distribution of information captured by combinations present in the data. The horizontal line marks the chosen cut-off value.

For the Generic effects formulation we perform an additional step where we recode all the independent price variables utilizing piecewise coding in order to capture the price sensitivity invariant of the products within a particular combination. Under linearity no cut points are included, thus we obtain one slope with end points at 0 and 10.55, the latter being the highest value among the products' price points. For the combinations including multiple items we aggregate the recoded product prices. For the non-linear case we have 24 slopes with the same end points and 22 cut points presented in Table 9 in the Appendix. Such specification allows to capture the utility by a flexible function without inflating the number of parameters and to avoid overfitting for price points with low occurrence.

Modular

Analogously to the EA, we consider the MOD models with Alternative-Specific and Generic effects, in a linear and non-linear version. The Alternative-Specific models are estimated with three methods, where each module is estimated separately. Due to practical limitations we estimate the Generic formulations only with the Aggregated Logit and HB without the skip factor.

Menu subsection	Number of combinations	Cumulative sum of information captured	Average number of cross-price effects
Snacks	4	97%	15.3
Fingerfood	6	98%	10.3
Sandwiches	13	95%	9.6
Meals1	13	96%	9.6
Meals2	17	94%	9
Kids'Meal	2	100%	12
Sides	5	96%	16
Sauses	3	100%	18.3
Drinks	9	96%	12.5
Desserts	7	97%	8.4

Table 4: Characteristics of the menu modules.

Firstly, we assign menu items to 10 modules. The separation is done based on logical subsection (Orme, 2019) and best practices from previous studies conducted by SKIM. The largest module, *Meals2*, includes 16 products and the smallest captures only a single item, *Kids'Meal*. The detailed allocation of products is presented in Table 10 in the Appendix. In contrast to the EA, the MOD approach has the ability to capture much more choice information within the subsection models. Table 4 presents the cumulative sum of information retained by each module, which aggregates to 96.9% for the complete MOD model. For most of the subsections the combinations included represent the *None* and all or some of the individual item choices - 10 products in total were not incorporated to any of the combinations. There are only two combinations comprising of two products which are incorporated in *Meals1* and *Sauses* modules.

Subsequently, we incorporate the cross-price effects between products from different modules. Again, we perform the variable selections employing the the χ^2-test at 5% significance level. On average the choice of each of the 59 products (captured by the combinatorial choices) is affected by 10.8 products' prices. The average number of cross-effects in each module are reported in Table 4.

For the the Generic effects version we recode the independent price variables in the same way as for the EA models. However, in order to capture the price sensitivity independent of the product the estimation for the 10 submodels has to be performed simultaneously. To make this possible, we merge the designs for all modules into a single overall design, which makes the estimation of these models computationally demanding. Because of the size of the design matrix the HB estimation incorporating a skip factor cannot be performed as the CBC/HB System is unable to save the sampled draws, thus we implement the HB without trimming.

5.2 Results

As mentioned earlier, in order to compare and characterize different modelling configurations we investigate in-sample validity, out-of-sample validity and various model characteristics.

5.2.1 Out-of-sample validity

The MAE results obtained based on the holdout sample predictions are presented in Table 5 for all 28 considered model configurations.

Altogether, the SCE approach offers the most predictive power, with the linear Alternative-Specific SCE estimated with HB without the skip factor obtaining the lowest MAE of 1.29%. Once more, there is evidence supporting the simpler disintegrative approach as superior for modelling menus with high complexity.

The EA models predict with the highest absolute error, which is not surprising considering the models make use of only 42% of the choice information from the data. This is inline with the existing research, which suggests EA models perform rather poorly for more complex MBC problems. Unlike for other underlying approaches, the Aggregated Logit estimations for Alternative-Specific effects EA perform similarly or better than the same models calculated employing HB. We cannot investigate if the same is true for the Generic formulation, as both Logit estimations failed terribly in the first iteration of the process. For the linear case we obtain a negative Percent Certainty of -18.1%, presented in Table 6, whereas under non-linearity the calculation stopped without performing any iterations reporting data fit of 100%. We did not identify a clear reason for these failures.

The MOD models redeem the combinatorial modelling approach. We can see that in combination with HB estimations they predict with a substantially better accuracy than the EA, with MAEs approaching the levels captured for the better performing SCE. As mentioned before, in our implementation the MOD includes only two multi-product combinations across 10 submodels, which could be one of the reasons behind this difference. We expect that this performance might change for menus of lower complexity, which would allow to incorporate more combinations of several items. However, we can conclude that the idea of separating the modelling into smaller submodels definitely works in practice. Interestingly, for the MOD we observe the biggest difference in the accuracy comparing the estimations employing the Aggregated Logit and HB. The former produces predictions only slightly better than the EA counterparts with a MAE of 2.4%.

Comparison of the HB estimation with and without the skip factor leads to a very interesting finding. The results across all models consistently show that employing the skip factor impairs their predictive accuracy. This is contrary to what is suggested in the literature, where employing the skip is a way to correct for the dependency among draws and making the inference more precise. Additionally, the skip HB requires much more iterations to be performed, thus is heavier for the hardware and expected to be more time consuming.

Next, we investigate the influence of the effects formulation on the predictive ability of the models. It can be seen that the EA configurations incorporating the Generic effects improve the accuracy of predictions in all of cases. For these models the assumption of item-invariant price sensitivity allows to regain some precision, reducing the MAE by approximately 0.2%. On the other hand, for the MOD the Alternative-Specific effects are superior for the majority of models. The non-linear Generic MOD Logit achieves the highest overall MAE of 2.72%. We suspect that the reason is rooted in the 24 piecewise variables capturing the price sensitivity. However, the most surprising result is captured for the Aggregate Logit version of the linear Generic MOD. Not only is this the best performing Logit estimation and the most accurate MOD configuration but also the offered improvement halves the MAE when compared to the Alternative-Specific equivalent reaching MAE of 1.43%.

Model	Effects	Linearity	Estimation method	Overall (MAE)	Frequency bucket 1	Frequency bucket 2	Frequency bucket 3	Snacks	Fingerfood	Sandwiches	Meals1	Meals2	Kids'Meal	Sides	Sauses	Drinks	Desserts
SCE	Alt Spec	Linear	Logit	1.611%	2.336%	1.371%	1.127%	2.241%	1.407%	1.490%	1.845%	1.121%	2.117%	2.368%	1.968%	1.819%	1.566%
			HB	1.370%	2.154%	1.190%	0.768%	2.412%	1.277%	1.411%	1.124%	0.671%	1.591%	2.557%	2.320%	1.844%	1.427%
			noskip HB	**1.287%**	**2.011%**	**1.126%**	**0.725%**	2.321%	**1.178%**	1.294%	1.081%	**0.632%**	1.377%	**2.342%**	2.244%	**1.755%**	1.350%
		Non-linear	Logit	1.603%	2.346%	1.363%	1.100%	2.252%	1.402%	1.514%	1.797%	1.112%	2.134%	2.447%	1.973%	1.769%	1.548%
			HB	1.395%	2.187%	1.218%	0.780%	2.435%	1.278%	1.447%	1.146%	0.686%	1.590%	2.639%	2.341%	1.875%	1.439%
			noskip HB	1.308%	2.036%	1.148%	0.741%	2.301%	1.221%	1.335%	1.071%	0.655%	1.365%	2.455%	2.204%	1.775%	1.357%
EA	Alt Spec	Linear	Logit	2.551%	4.835%	1.844%	0.973%	5.097%	2.926%	2.073%	1.072%	0.895%	3.774%	6.817%	7.737%	3.779%	2.884%
			HB	2.636%	4.917%	1.953%	1.038%	5.659%	3.086%	2.295%	1.018%	0.935%	2.956%	7.035%	7.917%	3.825%	2.872%
			noskip HB	2.572%	4.823%	1.900%	0.994%	5.456%	2.832%	2.238%	1.058%	0.892%	3.219%	6.965%	7.639%	3.712%	2.847%
		Non-linear	Logit	2.573%	4.856%	1.856%	1.006%	5.008%	2.971%	2.078%	1.094%	0.906%	3.701%	6.903%	7.832%	3.821%	2.921%
			HB	2.643%	4.942%	1.949%	1.038%	5.686%	3.070%	2.325%	1.032%	0.937%	2.819%	7.030%	7.957%	3.831%	2.863%
			noskip HB	2.558%	4.794%	1.903%	0.977%	5.171%	2.965%	2.243%	1.065%	0.882%	3.771%	6.796%	7.311%	3.698%	2.866%
	Generic	Linear	Logit	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			HB	2.405%	4.454%	1.790%	0.971%	4.859%	2.882%	2.076%	**0.915%**	0.879%	**1.167%**	6.883%	7.378%	3.632%	2.413%
			noskip HB	2.332%	4.315%	1.728%	0.953%	4.595%	2.871%	1.966%	0.947%	0.843%	1.873%	6.370%	6.819%	3.580%	2.446%
		Non-linear	Logit	-	-	-	-	-	-	-	-	-	-	-	-	-	-
			HB	2.458%	4.523%	1.848%	1.004%	4.984%	2.967%	2.152%	0.952%	0.904%	1.540%	6.734%	7.241%	3.684%	2.615%
			noskip HB	2.430%	4.581%	1.779%	0.931%	4.863%	2.885%	2.134%	0.964%	0.839%	2.464%	6.777%	7.161%	3.649%	2.498%
MOD	Alt Spec	Linear	Logit	2.397%	3.333%	2.347%	1.509%	2.841%	1.522%	3.617%	3.129%	1.657%	2.117%	3.119%	1.711%	1.964%	1.341%
			HB	1.660%	2.371%	1.495%	1.115%	3.465%	1.414%	1.676%	1.317%	1.013%	1.569%	3.089%	2.410%	2.054%	1.634%
			noskip HB	1.503%	2.146%	1.329%	1.033%	3.245%	1.307%	1.454%	1.173%	0.903%	1.358%	2.853%	2.133%	1.915%	1.516%
		Non-linear	Logit	2.397%	3.342%	2.339%	1.510%	2.855%	1.519%	3.617%	3.129%	1.657%	2.134%	3.168%	1.704%	1.956%	**1.316%**
			HB	1.678%	2.405%	1.509%	1.120%	3.537%	1.420%	1.709%	1.346%	1.023%	1.631%	3.096%	2.371%	2.062%	1.635%
			noskip HB	1.511%	2.167%	1.326%	1.039%	3.272%	1.350%	1.439%	1.170%	0.907%	1.378%	2.909%	2.197%	1.941%	1.485%
	Generic	Linear	Logit	1.431%	2.108%	1.203%	0.984%	2.854%	1.532%	**1.200%**	1.228%	0.843%	2.118%	2.722%	**1.702%**	1.925%	1.353%
			noskip HB	1.657%	2.313%	1.446%	1.211%	3.045%	1.956%	1.375%	1.245%	1.153%	1.941%	3.034%	3.413%	1.826%	1.663%
		Non-linear	Logit	2.723%	4.114%	2.648%	1.406%	**2.008%**	1.308%	3.324%	2.858%	1.586%	12.240%	3.095%	7.288%	3.123%	1.928%
			noskip HB	1.787%	2.663%	1.652%	1.045%	3.300%	1.857%	1.374%	1.201%	0.949%	1.480%	3.253%	9.930%	1.783%	1.569%

Table 5: Models' Mean Absolute Errors including separation between frequency buckets and menu subsections.

Capturing the own-effect with a non-linear formulation does not seem to offer any substantial advantage for any of the configurations. Although it could be expected that the more flexible specification would produce more accurate projections, this observation can be explained by the fact that we consider only three price points per product which seems to work well with the linear assumption.

Capturing the MAE for separate frequency buckets does not seem to offer any additional insights for comparing the performance of different model configurations. In principal the MAE increases with the product frequencies.

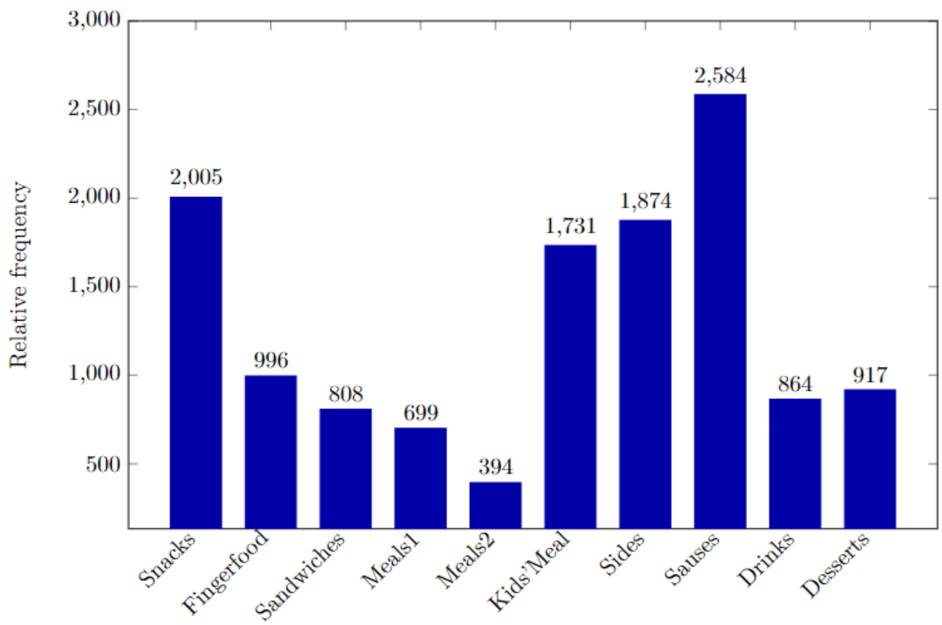

Figure 5: Choice frequencies aggregated by menu subsections relative to the number of products included in each module.

We also analyze how well the models behave for different menu subsections. Firstly, consistently across all underlying models we obtain greater MAEs for modules incorporating add-on products: *Snacks*, *Sides* and *Sauses*. Following the disparity of the metric between the frequency buckets we investigate product frequencies in Figure 1 in order to explain this phenomenon, but no apparent dependency is found. Subsequently, we explore the module frequencies relative to its size and visualize them in Figure 5. The modules in question are characterized by a high relative frequency. Thus, we conclude that the MAE of the MBC models increases with the relative frequency and this effect is more severe for the combinatorial models. Additionally, Table 5 showcases an interesting pattern for these subsection, where the Aggregated Logit configurations perform better than the HB estimations.

Separate module MAEs also allow us to consider the drivers of the improvement offered by the Generic effects. Analyzing the differences between the linear MOD Logit results it appears that the greatest performance gain is attributed to *Sandwiches* and *Meals1*, subsections with highest frequencies. The improvement of EA configurations can be accredited to the *Kids'Meal* module, for which the absolute error decreases across all HB models by approximately 1.4%. On the other hand, the MAE for the worse performing nonlinear MOD Logit is inflated due to two subsections: the prediction errors obtained for *Sauses* are much higher than for the linear case or the Alternative-Specific effects reaching a similar level as for the EA approach, however the main factor is the score for the *Kids'Menu* rising to 12.24%, a magnitude unlike any other MAE reported in this study. For the non-linear MOD HB without the skip we observe that the overall MAE is enlarged by the *Sauses* score of 9.93% exceeding any other result for this module. Interestingly, unlike for the Logit estimation the *Kids'Menu* MAE is reported at a low level.

5.2.2 In-sample validity

The results in Table 6 capture the models' hit rate, allowing to asses the predictive performance for individual respondent choices within the sample.

Looking at the results, it is apparent that the Aggregated Logit estimation offers no predictive power for producing this type of projections, which is in line with our expectation as the method assumes equal utilities across all respondents.

In general, the values obtained for the HB estimations vary around 30% of correctly predicted choices, which is in line with the negative relationship found between the hit rate and menu complexity (Neuerburg, 2015). For the overall hit rate the best performing model is the non-linear SCE estimated without the skip factor, reporting 33.94% of the forecasted item choices to be true.

Considering the underlying models, the SCE is superior to both combinatorial approaches and, similarly to the MAE results, the MOD offers a considerable improvement over the EA. Unlike for the out-of-sample measures, employing the estimation with the skip factor increases the predictive accuracy for the EA and the MOD. The SCE benefits from correcting for the dependency of the draws, however the improvement is marginal.

For the EA models the Generic effects formulation boosts the hit rate by 3.5% on average, which is a considerable improvement. The same cannot be said for the MOD models. Although the Logit models report higher hit rates, their overall level is drastically low. For the HB estimations the metrics are much lower than for the Alternative-Specific effects and even the Generic EA versions estimated with the same approach. Furthermore, we can see that the non-linear own-effects formulation slightly increases the reported metrics across the majority of configurations estimated with HB. The only exception is the Generic effects MOD.

Model	Effects	Linearity	Estimation method	Hit Rate		Avg. number of products estimated	Avg. percent certainty	Adjusted estimation time (sec)	Share of choice information used	Avg. number of parameters
				Overall	Above 5 products					
SCE	Alt Spec	Linear	Logit	6.629%	6.572%	3.05	78.0%	**1.0**	**100.0%**	14
			HB	33.884%	35.618%	1.97	**88.2%**	194.0	**100.0%**	14
			noskip HB	33.910%	**35.736%**	2.05	87.9%	195.5	**100.0%**	14
		Non-linear	Logit	6.604%	6.545%	3.02	78.0%	3.0	**100.0%**	15
			HB	33.859%	35.524%	1.94	**88.2%**	598.3	**100.0%**	15
			noskip HB	**33.937%**	35.667%	2.03	87.6%	106.5	**100.0%**	15
EA	Alt Spec	Linear	Logit	2.726%	2.464%	1.11	12.2%	7560.0	42.0%	195
			HB	28.128%	5.053%	1.04	68.1%	78268.0	42.0%	195
			noskip HB	24.254%	4.839%	1.09	67.6%	39562.0	42.0%	195
		Non-linear	Logit	2.719%	2.451%	1.10	12.2%	18360.0	42.0%	263
			HB	28.435%	5.575%	1.04	69.6%	93846.0	42.0%	263
			noskip HB	24.710%	4.309%	1.10	68.9%	100417.0	42.0%	263
	Generic	Linear	Logit	-	-	-	-18.1%	25.0	42.0%	129
			HB	31.842%	6.284%	1.22	63.9%	52759.0	42.0%	129
			noskip HB	27.853%	4.977%	1.25	63.1%	22888.0	42.0%	129
		Non-linear	Logit	-	-	-	100.0%	54.0	42.0%	152
			HB	32.336%	5.972%	1.16	65.0%	63749.0	42.0%	152
			noskip HB	27.994%	5.351%	1.18	64.7%	32511.0	42.0%	152
MOD	Alt Spec	Linear	Logit	4.017%	4.495%	1.42	41.5%	14.0	96.9%	47.9
			HB	32.435%	24.426%	1.67	77.8%	2311.1	96.9%	47.9
			noskip HB	31.967%	24.200%	1.81	77.7%	1590.6	96.9%	47.9
		Non-linear	Logit	4.008%	4.485%	1.41	41.5%	21.6	96.9%	53.8
			HB	32.525%	24.331%	1.66	77.9%	1778.2	96.9%	53.8
			noskip HB	32.051%	24.316%	1.81	78.0%	1926.6	96.9%	53.8
	Generic	Linear	Logit	5.410%	5.400%	2.18	54.1%	2498.0	96.9%	421
			noskip HB	26.845%	22.102%	2.47	74.4%	718027.0	96.9%	421
		Non-linear	Logit	4.182%	4.957%	1.52	46.5%	715.0	96.9%	444
			noskip HB	26.811%	22.370%	**2.48**	74.8%	765300.0	96.9%	444

Table 6: Models' characteristics and performance measures.

The second version of the hit rate is calculated only across the tasks in which respondents chose more than 5 products. It can be seen that the EA approach fails completely because the largest combinations included in the modelling incorporate only four menu items. The MOD obtains much better results but we still observe a significant decrease of from the overall hit rate. For the Alternative-Specific effects the score differs by approximately 8%, which is double the difference reported for the Generic effects. The decompositional SCE method not only once again proves to be superior but also obtains greater accuracy in predicting choices involving many products.

5.2.3 Other characteristics

Finally, we also analyze characteristics other than the formal validation measures, anticipating they might offer some additional insights into the behaviour of different configurations. They are showcased in Table 6, with the average value reported for approaches comprising of multiple separately estimated models.

Percent certainty

The goodness of fit seems to be closely related to the predictive accuracy. The best performing SCE approach is also characterized by the highest McFadden's ρ^2, achieving 82% for both linear and non-linear versions computed by HB without the skip factor. As expected, the lowest scores are obtained for the EA.

The conclusions concerning varied modelling aspects follow the patterns described in the previous sections. The Aggregated Logit estimations do not fit the data well with the only exception being the SCE models, which partially explains reasonable predictive performance of these configurations. As mentioned before, for the Generic versions of the EA the estimations fail completely reporting very strange levels of Percent Certainty. We find evidence that applying the skip factor for the HB improves the goodness of fit, however once again the difference is minor of approximately 0.5%. The linearity assumption does not have significant influence on McFadden's ρ^2 for the SCE, however for the combinatorial models we see that the non-linear case fits the data better which is in line with our expectations. Irregular behaviour is observed for the MOD with Generic effects where the linear version offers much better fit. Also, the difference between the two configurations is substantial, whereas for all other approaches results are comparable despite the own-effects functional form. The Generic effects formulation lowers the data fit for the HB estimations for both EA and MOD when compared to their Alternative-Specific counterparts. On the other hand, Generic effects seem to better the Percent Certainty obtained for the Aggregated Logit MOD. However, it should be emphasized that for the MOD the number of parameters, also reported in Table 6, explodes when the two effects formulations are contrasted, which might be the driver of the observed improvement for the Logit estimations.

Setting the models' MAE side by side the Percent Certainty levels allows to investigate if there is any indication of overfitting for any of the configurations, which would be anticipated for very good data fit accompanied by extremely poor out-of-sample predictions. Luckily, we do not observe this situation for any of the tested models.

Average number of products estimated

We aggregate the predicted choice shares across all menu items for the in-sample holdout tasks and compare it to the average number of products actually chosen equalling to 2.52 products per task. This simple method provides some insight to asses if a particular model suffers from over- or underestimation.

The majority of the modelling configurations underestimate the number of chosen products, besides two SCE models both estimated with the Aggregated Logit which overestimate the number of selections. This explains the increased individual item prediction accuracy for choices above 5 products reported for these two configurations. In general, we observe that the greater the deviation from the observed average the worse the models predict. Thus, the most underestimating approach is clearly the EA, which predicts only one product being chosen per task. Unsurprisingly, the SCE approach is the most reliable for the Alternative-Specific effects in spite the estimation method. However, the predictions closest to the observed value are obtained for the Generic effects MOD estimated with HB, the better non-linear anticipating 2.48 products selected on average. We accredit such precise score to the exceptionally large number of parameters when contrasted with the other models.

Estimation time

The last property we discuss is connected to the practical side of the modelling, which is the computational time. In Table 6 we report the adjusted estimation time in seconds. We correct for the fact that for the models comprising of multiple submodels the software can perform up to four simulations at once, thus for these instances we divide the total time by four.

It is apparent, that the Aggregated Logit trade-offs the predictive power for the calculation speed as the convergence happens considerably faster than for the HBs, often taking only few seconds. The quickest model is the linear Alternative-Specific SCE, which produces results in a single second and still retains a reasonable level of predictive power for the out-of-sample product share. Estimations with HB take considerably slower, however in general they produce lower MAEs and most importantly they allow to some degree predict the item choices on an individual level. The slowest method, the non-linear Generic effects MOD HB without the skip factor, took 765,300 seconds to estimate which translates to 212 hours and 35 minutes.

We observe substantial differences between three underlying models compared based on the Alternative-Specific effects formulation. The amount of time required to obtain the estimates for the EA seems unreasonably high and impractical, especially when its performance is also taken into account. Although, the SCE estimations are much faster than for the MOD, they both seem to offer a fair balance between the predictive power and computational feasibility for the Alternative-Specific formulation. We cannot draw explicit conclusions about the Generic effects. This formulation is an experimental approach for which the implementation process can be further tested and optimized. We applied the method with the Sawtooth Software which was not designed for such applications hence the lengthy estimation time due to large number of parameters in the design matrix for the MOD. Notably, for the EA HB estimations performed significantly faster under the Generic effects formulation than its Alternative-Specific equivalent.

Interestingly, the reported results uncover some inconsistencies with our expectations concerning the time difference between the HB estimations with and without the skip factor. The former requires 19,800 iterations more than the latter for the chosen skip of 100. Thus, we logically assume that performing almost twice as many repetitions would consistently prolong the convergence time regardless of the model configuration. In Table 6 we identify that in three out of eight cases the HB without the skip took longer to estimate that the more independent version of the sampler. This is very odd and we did not find any apparent reason for such situation. One of the reasons could be a momentary issue with the hardware or the software. Similarly, we see that in three out of 13 instances (where the failed estimations are discarded) the linear model is more time consuming than its non-linear equivalent, which is counterintuitive as the latter contains more parameters to be estimated and therefore more computationally demanding. The only similarities between these pairs is that in two of them the reported goodness of fit to the data is higher for the linear case than for the non-linear, in the third both Percent certainties are nearly equal.

6 Conclusions

In time when product customization becomes more prevalent on the market and the menu-based sales become the predominant strategy for many industries the ability to understand and choose the most appropriate modelling approach is more essential then ever. In this research we compare multiple modelling arrangements with an aim to reveal their advantages and disadvantages and formulate useful recommendations for the modelling choices.

We employ three different underlying models for the MBC analysis. The SCE approach decomposies the menu situation into a series of binary choice models interconnected with product cross-price effects. The combinatorial EA captures the choices in a single model considering all possible choices and the MOD separates the menu into modules, each analyzed by separate EA submodel and incorporating cross-price effects of products belonging to other menu subsections. Each model is tested with a linear and non-linear functional form of the main price-effect. For the combinatorial models two assumptions about price sensitivity are investigated, product dependent and item invariant versions. Finally, we estimate the proposed model configurations with three estimation techniques including the Aggregated Logit and the HB with and without the skip factor. The modelling configurations are examined by the means of in-sample and out-of-sample predictive performance, the goodness of fit to the data and estimations time. In order to obtain the accuracy measures we incorporate holdout respondents and holdout tasks for the estimation sample during the data gathering process.

We also introduce an innovative application of the Generic effects formulation build on the idea of Orme (2019), where we assume that the price sensitivity is invariant of the product in question. Additionally, as the customizable products and services are becoming more complex and most of the existing MBC literature tackles simplified menus, we perform the analysis on a dataset based on a real-life menu of very high complexity gathered for the purpose of this study. It includes 69 products, which translates to approximately 590 quintillion combinations. This is done so our research can contribute not only to academic literature but also to be meaningful for practical applications.

6.1 Recommendations

Considering the above results, we formulate recommendations about the performance and behaviour of tested configurations and their varying components. It should be emphasized that our reasoning is based strictly on a high complexity menu example and we do not believe there exists single best technique for MBC modelling, but rather that the methods should be chosen and adjusted depending on the data, characteristics of menu tasks and aim of the analysis.

Firstly, in line with existing research we find evidence that the simpler decompositional SCE method performed the best both for the in-sample and out-of-sample validity measures as well as achieved the best goodness of fit to the data. Importantly, in the majority of tasks only a single item was chosen which might have been advantageous for this technique, when compared with the combinatorial models as most of the combinations also incorporated a single product. For this reason we consider the performance of the MOD models (estimated with HB) very reasonable, especially that the underlying approach is different and more theoretically complete by allowing to account for multiple choices to be made simultaneously. Because of this characteristic, we expect that for less complex models and for menu situations in which respondents typically choose more than a single item this approach might deem preferable and more insightful. Both techniques require to identify the cross-price effects influencing the choice of particular menu items, which provides additional understanding of consumers' decision-making process and seems to better the forecast quality. For research aiming to make in-sample item predictions on respondent level the SCE proved to be the model of choice not only because of the highest hit rate scores but also because it is the only model for which this accuracy improves for multi-product choices.

Secondly, we find that the EA approach in the current implementation is completely unreliable. The configurations take extremely long to estimate and produce the worst predictions of the three underlying models tested. Additionally, considering the practical application only a limited part of the gathered data is used for producing the projections which not only is the reason behind the poor model performance but also is inefficient when data acquisition costs are taken into account. The main advantage of the EA is that it simulates the decision process of the consumers who compare all viable options, which can be useful for determining products to be included and offered as bundles. However, for such complex menu situations the assumption that the respondents would consider all the possible 590 quintillion combinations is completely unrealistic. It is worth mentioning that Pfaff (2021) discovers an alternative way of constructing the combination subset, which significantly improved the predictive performance of the EA approach, thus the technique should not be discarded but rather further investigated.

Thirdly, we trial the Generic effects formulation across multiple models which has never been done before. Although the assumption of product-invariant price sensitivity seems improbable it turns out it works well for the combinatorial EA models, when put together with the HB, improving the predictive performance for both observed measures. We strongly discourage implementations of the Generic EA with the Aggregated Logit. Analyzing the average number of products estimated it can be seen that the underestimation for Generic models is lower than for the Alternative-Specific counterparts for both the EA and MOD, however for the latter approach the predictive accuracy is worse for the HB configurations. We do not suggest utilizing this specification for MOD HB estimations without the skip factor. However, for the Generic linear MOD estimated with the Aggregated Logit the reported predictions are not only superior to Alternative-Specific effects but are the best scores reported for any of the models employing this estimation method. This is a promising result which we recommend to be further investigated.

The main advantage of the Aggregate Logit estimation is the considerably shorter computational time when compared to HB. As mentioned before, for the majority of specifications this decrease comes from giving up some of the predictive power and goodness of fit to the data. We find that for the SCE models this technique performs relatively well when producing out-of-sample product share predictions. However, the HB results for the SCE are available after only few minutes with much more accurate out-of-sample forecasts and some ability to predict on the individual level. On the other hand, for the linear Generic MOD the Aggregated Logit performs very well and the time gain equals to dozens of hours, thus we find it very practical for this application. We also uncover that the Aggregated Logit allows to obtain similar or lower MAE for product subsections with high relative frequency despite the underlying model. It is unclear if this dependency is generalizable, but it might mean that Aggregated Logit could be a preferable estimation method for enormous datasets. Unfortunately, in practice acquiring this amount of choice data is highly infeasible.

For the more accurate HB technique we consider two versions - with and without a skip factor, and make a very unexpected discovery. Opposite to what is suggested in the literature, the correction for dependency between draws does not prevent from precision loss, but rather impairs the accuracy of out-of-sample predictions. Additionally, for most configurations the estimation is much faster as it requires less iterations to be performed. Thus, we would not suggest implementing the skip factor unless the focus is on in-sample predictions measured with the hit rate for which the skip proves advantageous.

We do not discover any apparent advantages for using the non-linear own-effects functional form, except a marginal improvement in the hit rate scores. For the out-of-sample predictions the linear case performs much better across all models and for the majority of configurations the estimation was considerably faster. Thus, we suggest employing the linear functional form for the price of menu items when three or less price points are investigated. It should be noted, that we expect this conclusion might change for studies including more price points per product and the non-linear models may prove to be more appropriate for such applications.

Finally, we formulate some remarks concerning the model comparison. Percent Certainty seems to be related to the models' predictive power and can be used as a preliminary indicator of its performance, however it is imperative to keep in mind that the scores also depend on other modelling choices. Based on reported values we observe that the configurations which obtained goodness of fit of 75% or higher where the models which were predicting reasonably well. We also found that analyzing the MAE for different frequency buckets did not provide any additional insight into model behaviour, thus we suggest not to employ this approach. On the other hand, the investigation of the MAE for menu subsections proved useful and allowed us to uncover the relationship between MAE and the relative frequency.

6.2 Limitations and future research

During our research we encountered a number of limitations. The investigated application of Generic effects for the EA model failed for the Aggregated Logit estimation for reasons we did not identify. The MOD version required to merge the design matrix externally which limited its size, thus restricting the number of combinations incorporated in the submodels. Consequently, the Alternative-Specific MOD had to be adjusted for a fair comparison of performance between these two assumptions.

We recommend extending our research to cover a wider scope of configurations and make the comparison more thorough. The Generic effects SCE models can be investigated, as well as the Substitute-Specific effects formulation. The latter is an idea of a novel intermediate case based on the strong substitutes determined using Counting Analysis, which assumes that the price sensitivity does change between products, but is equal among substitutes. We denote this by $\beta_{ij} = \beta_{iSub}$ where $j \in Sub$, a set of substitutes.

Furthermore, we encourage incorporating different approaches to constructing the modules for the MOD approach. One of the approaches could be based on groups of products which are strong complements. The intuition is that if there are strong complements to a product the respondent is determined to choose, he will consider all the possible choices which include that product and might increase his preference. We would represent such a situation by a single EA submodel.

The non-linear version of the models could be tested for more price points than three in order to discover if there exists a threshold above which the linear case becomes inferior.

We strongly endorse a closer comparison of the SCE and MOD models and exploring if incorporating more combinations in the submodels would further improve the MOD predictive performance. The behaviour of this model could be also tested with the Stratified Importance Sampling proposed by Pfaff (2021) for the EA, possibly for both the Alternative-Specific and Generic effects.

Finally, the drawn conclusions are based on a single MBC dataset, thus we would find it reasonable to investigate if our insights are also consistent for other high complexity menu situations and if the type of industry represented by the menu may change the characteristics of consumers' choices.

References

Alvarez, I., Niemi, J., and Simpson, M. (2014). Bayesian inference for a covariance matrix. In *26th Annual Conference on Applied Statistics in Agriculture*.

Bakken, D. and Bond, M. K. (2004). Estimating preferences for product bundles vs. a la carte choices. In *Proceedings of the Sawtooth Software Conference*, pages 125–133.

Ben-Akiva, M. and Gershenfeld, S. (1998). Multi-featured products and services: Analysing pricing and bundling strategies. *Journal of Forecasting*, 17(3-4):175–196.

Cohen, S. H. and Liechty, J. (2007). Have it your way: Menu-based conjoint analysis helps marketers understand mass customization. *Marketing Research*, 19(3):28–34.

Cordella, P., Borghi, C., van der Wagt, K., and Loosschilder, G. (2012). Menu-based choice modeling (MBC): An empirical comparison of two analysis approaches. In *Proceedings of the Sawtooth Software Conference*, pages 247–256.

Dippold-Tausendpfund, K. and Neuerburg, C. (2018). Variable selection for mbc cross-price effects. In *Proceedings of the Sawtooth Software Conference*, pages 189–198.

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

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (1995). *Bayesian data analysis.* Chapman and Hall/CRC, third edition.

Green, P. E. and Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8(3):355–363.

Greenberg, E. (2013). *Introduction to Bayesian Econometrics.* Cambridge University Press, second edition.

Kamakura, W. A. and Kwak, K. (2020). Menu-choice modeling with interactions and heterogeneous correlated preferences. *Journal of Choice Modelling*, 37:100214.

Liechty, J., Ramaswamy, V., and Cohen, S. H. (2001). Choice menus for mass customization: An experimental approach for analyzing customer demand with an application to a web-based information service. *Journal of Marketing Research*, 38(2):183–196.

Luce, R. D. and Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1):1–27.

Magretta, J. (1998). The power of virtual integration: an interview with dell computer's michael dell. *Harvard Business Review*, 76(2):72–84.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*, pages 105–142.

Neuerburg, C. (2015). Menu-based choice: Probit as an alternative to logit? In *Proceedings of the Sawtooth Software Conference*, pages 285–298.

Neuerburg, C., Koschate-Fischer, N., and Pescher, C. (2021). Menu-based choice models for customization: On the recoverability of reservation prices, model fit, and predictive validity. *Journal of Interactive Marketing*, 53(1):1–14.

Orme, B. (2014). Including holdout choice tasks in conjoint studies.

Orme, B. K. (2011). Menu-based choice modeling using traditional tools. In *Proceedings of the Sawtooth Software Conference*, pages 37–57.

Orme, B. K. (2019). *Menu-Based Choice (MBC) for Multi-Check choice experiments*. Sawtooth Software, Inc., v1.1.1 edition.

Orme, B. K. (2020). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Research Publishers LLC, fourth edition.

Orme, B. K. and Chrzan, K. (2017). *Becooming an expert in Conjoint Analysis: Choice modeling for Pros*. Sawtooth Software, Inc.

Pfaff, L. (2021). Comparison of menu-based conjoint methods for different questionnaire designs. Master's thesis.

Sawtooth Software, I. (2021). The cbc/hb system technical paper v5.6.

Wind, J. and Mahajan, V. (1997). Issues and opportunities in new product development: An introduction to the special issue. *Journal of Marketing Research*, 34(1):1–12.

7 Appendix

	Product	Price 1	Price 2	Price 3
1	Snack1	1.45	1.60	1.75
2	Snack2	1.15	1.30	1.45
3	Snack3	1.35	1.50	1.65
4	Fingerfood1_1	2.00	2.25	2.50
5	Fingerfood1_2	3.15	3.50	3.85
6	Snack4	2.50	2.80	3.10
7	Sandwich1	3.50	3.90	4.30
8	Sandwich2	3.95	4.40	4.85
9	Sandwich3	3.95	4.40	4.85
10	Sandwich4	3.95	4.40	4.85
11	Sandwich5	3.95	4.40	4.85
12	Fingerfood1_3	6.70	7.45	8.20
13	Sandwich6	4.70	5.20	5.70
14	Sandwich7	5.35	5.95	6.55
15	Sandwich8	5.75	6.40	7.05
16	Sandwich9	5.35	5.95	6.55
17	Sandwich10	5.60	6.20	6.80
18	Fingerfood2_1	3.80	4.20	4.60
19	Fingerfood2_2	5.75	6.40	7.05
20	Sandwich11	4.15	4.60	5.05
21	MealS1	7.15	7.95	8.75
22	LargeMealS1	0.65	0.70	0.75
23	MealS3	7.15	7.95	8.75
24	LargeMealS3	0.65	0.70	0.75
25	MealS4	7.15	7.95	8.75
26	LargeMealS4	0.65	0.70	0.75
27	MealS5	7.15	7.95	8.75
28	LargeMealS5	0.65	0.70	0.75
29	MealF1_2	7.15	7.95	8.75
30	LargeMealF1_2	0.65	0.70	0.75
31	MealS2	7.15	7.95	8.75
32	LargeMealS2	0.65	0.70	0.75
33	MealS6	7.75	8.60	9.45
34	LargeMealS6	0.65	0.70	0.75
35	MealS7	8.35	9.30	10.25
36	LargeMealS7	0.65	0.70	0.75
37	MealS8	8.65	9.60	10.55
38	LargeMealS8	0.65	0.70	0.75
39	MealS9	8.35	9.30	10.25
40	LargeMealS9	0.65	0.70	0.75
41	MealS10	8.65	9.60	10.55
42	LargeMealS10	0.65	0.70	0.75
43	MealS11	7.05	7.85	8.65
44	LargeMealS11	0.65	0.70	0.75
45	Salad1	4.95	5.50	6.05
46	MealSalad1	7.45	8.30	9.15
47	Salad2	4.95	5.50	6.05
48	MealSalad2	7.75	8.60	9.45
49	Kids'Meal	4.05	4.50	4.95
50	Drink1_1	2.05	2.30	2.55
51	Drink1_2	2.45	2.70	2.95
52	Drink1_3	2.75	3.05	3.35
53	Drink2_1	1.60	1.75	1.95
54	Drink2_2	2.00	2.20	2.40
55	Drink3	2.25	2.50	2.75
56	Drink4	2.25	2.50	2.75
57	Dessert1	3.25	3.60	3.95
58	Dessert2	3.05	3.40	3.75
59	Drink5	2.50	2.75	3.00
60	Dessert3	2.90	3.25	3.60
61	Dessert4	2.15	2.40	2.65
62	Dessert5	1.35	1.50	1.65
63	Dessert6	3.60	4.00	4.40
64	Saus1	0.60	0.65	0.70
65	Saus2	0.60	0.65	0.70
66	Side1_1	2.05	2.30	2.55
67	Side1_2	2.45	2.70	2.95
68	Side1_3	2.75	3.05	3.35
69	Side2	2.45	2.70	2.95

Table 7: All menu items with their corresponding price points (in €).

	Product	Price 1	Price 2	Price 3	Total	Choice share
1	Snack1	807	741	682	2230	9.59%
2	Snack2	785	688	717	2190	9.42%
3	Snack3	555	562	479	1596	6.87%
4	Fingerfood1_1	579	509	543	1631	7.02%
5	Fingerfood1_2	436	397	401	1234	5.31%
6	Snack4	586	533	487	1606	6.91%
7	Sandwich1	268	276	240	784	3.37%
8	Sandwich2	157	157	165	479	2.06%
9	Sandwich3	621	499	512	1632	7.02%
10	Sandwich4	446	392	377	1215	5.23%
11	Sandwich5	310	269	282	861	3.70%
12	Fingerfood1_3	463	402	399	1264	5.44%
13	Sandwich6	217	221	191	629	2.71%
14	Sandwich7	126	135	110	371	1.60%
15	Sandwich8	209	197	158	564	2.43%
16	Sandwich9	127	139	123	389	1.67%
17	Sandwich10	214	206	197	617	2.65%
18	Fingerfood2_1	128	112	111	351	1.51%
19	Fingerfood2_2	180	171	151	502	2.16%
20	Sandwich11	185	190	177	552	2.37%
21	MealS1	217	209	216	642	2.76%
22	LargeMealS1	98	105	92	295	1.27%
23	MealS3	656	617	545	1818	7.82%
24	LargeMealS3	344	355	349	1048	4.51%
25	MealS4	434	412	409	1255	5.40%
26	LargeMealS4	181	194	164	539	2.32%
27	MealS5	275	260	226	761	3.27%
28	LargeMealS5	116	129	113	358	1.54%
29	MealF1_2	200	197	194	591	2.54%
30	LargeMealF1_2	95	77	79	251	1.08%
31	MealS2	178	187	167	532	2.29%

32	LargeMealS2	74	78	79	231	0.99%
33	MealS6	204	167	159	530	2.28%
34	LargeMealS6	57	76	60	193	0.83%
35	MealS7	138	112	111	361	1.55%
36	LargeMealS7	54	53	48	155	0.67%
37	MealS8	285	259	203	747	3.21%
38	LargeMealS8	123	114	122	359	1.54%
39	MealS9	135	130	128	393	1.69%
40	LargeMealS9	53	46	54	153	0.66%
41	MealS10	261	240	221	722	3.11%
42	LargeMealS10	144	144	128	416	1.79%
43	MealS11	169	163	139	471	2.03%
44	LargeMealS11	76	84	71	231	0.99%
45	Salad1	155	161	159	475	2.04%
46	MealSalad1	97	84	102	283	1.22%
47	Salad2	195	199	180	574	2.47%
48	MealSalad2	106	101	91	298	1.28%
49	Kids'Meal	617	580	534	1731	7.45%
50	Drink1_1	681	613	623	1917	8.25%
51	Drink1_2	451	408	396	1255	5.40%
52	Drink1_3	221	180	199	600	2.58%
53	Drink2_1	349	347	322	1018	4.38%
54	Drink2_2	127	95	93	315	1.36%
55	Drink3	193	216	204	613	2.64%
56	Drink4	146	150	144	440	1.89%
57	Dessert1	359	311	319	989	4.25%
58	Dessert2	227	245	258	730	3.14%
59	Drink5	264	254	237	755	3.25%
60	Dessert3	604	533	540	1677	7.21%
61	Dessert4	390	401	355	1146	4.93%
62	Dessert5	219	236	191	646	2.78%
63	Dessert6	110	103	101	314	1.35%
64	Saus1	1491	1466	1408	4365	18.78%

65	Saus2	288	253	262	803	3.45%
66	Side1_1	946	987	969	2902	12.48%
67	Side1_2	773	762	724	2259	9.72%
68	Side1_3	597	527	469	1593	6.85%
69	Side2	224	264	254	742	3.19%

Table 8: Product frequencies and choice shares

	slope1	slope2	slope3	slope4	slope5	slope6	slope7	slope8
End	0.65	0.7	0.75	1.5	1.95	2.4	2.9	3
Start	0.6	0.65	0.7	1.15	1.5	2	2.4	2.9

	slope9	slope10	slope11	slope12	slope13	slope14	slope15	slope16
End	3.4	3.9	4	4.4	4.95	5.5	5.95	6.4
Start	3	3.4	3.9	4	4.4	4.95	5.5	5.95

	slope17	slope18	slope19	slope20	slope21	slope22	slope23	slope24
End	6.8	7.05	7.95	8.2	8.75	9.15	9.6	10.55
Start	6.4	6.8	7.05	7.95	8.2	8.75	9.15	9.6

Table 9: Piecewise price coding cut points for the non-linear case.

	Snacks	Fingerfood	Sandwiches	Meals1	Meals2	Kids'Meal	Sides	Sauses	Drinks	Desserts
1	Snack1	Fingerfood1_1	Snack4	MealS3	MealS1	Kids'Meal	Side1_1	Saus1	Drink1_1	Dessert1
2	Snack2	Fingerfood1_2	Sandwich1	LargeMealS3	LargeMealS1		Side1_2	Saus2	Drink1_2	Dessert2
3	Snack3	Fingerfood1_3	Sandwich2	MealS4	MealS5		Side1_3		Drink1_3	Dessert3
4		Fingerfood2_1	Sandwich3	LargeMealS4	LargeMealS5		Side2		Drink2_1	Dessert4
5		Fingerfood2_2	Sandwich4	MealS6	MealF1_2				Drink2_2	Dessert5
6			Sandwich5	LargeMealS6	LargeMealF1_2				Drink3	Dessert6
7			Sandwich6	MealS8	MealS2				Drink4	
8			Sandwich7	LargeMealS8	LargeMealS2				Drink5	
9			Sandwich8	MealS10	MealS7					
10			Sandwich9	LargeMealS10	LargeMealS7					
11			Sandwich10	Salad1	MealS9					
12			Sandwich11	MealSalad1	LargeMealS9					
13					MealS11					
14					LargeMealS11					
15					Salad2					
16					MealSalad2					

Table 10: Menu items separated into subsections.