Journal of Revenue and Pricing Management
https://doi.org/10.1057/s41272-024-00486-6

RESEARCH ARTICLE

# Group-constrained assortment optimization under the multinomial logit model

**Julia Heger[1] · Robert Klein[1]**

Received: 23 September 2023 / Accepted: 11 March 2024


**Abstract**
We study an assortment problem under the multinomial logit model with two new types of group constraints that are motivated by a joint project with the German car manufacturer BMW. Under group constraints, products are either attributed to exactly one group or to several groups at once and there is either a bound on the number of products offered per group or on the number of groups from which products are offered. We formulate both optimization problems as binary fractional linear program and provide reformulations that can be solved using state-of-the-art solvers. Finally, we conduct a numerical study and find that all instances of the products-per-group constrained problem as well as small to medium size instances of the number-of-offered-groups constrained problem can be solved within fractions of a second, whereas large instances of the latter problem might take some seconds to be solved.

**Keywords** Assortment optimization · Group constraint · MNL model

## Introduction

This research is motivated by an ongoing project with the German car manufacturer BMW. In particular, the project concerns its online sales platform—also known as BMW new car locator—which is depicted in Fig. 1. The new car locator website offers all BMW vehicles that are directly available for sale. The vehicles become visible in batches of size six by clicking on the 'Mehr anzeigen' (engl.: 'show more') button displayed on the bottom of Fig. 1.

There are various vehicles available that can be offered to the customer via the new car locator website. The decision on which of these vehicles should be offered is of significant importance since it directly impacts the success of the business (see Jena et al. 2020). However, selecting the optimal offer set of vehicles is challenging since according to Jena et al. (2020) it is well known that too large assortments jeopardize total sales, but offering more vehicles might increase the conversion, i.e., the number of sold vehicles. This becomes more apparent when considering the extreme cases of offering all available vehicles and offering only one of the available vehicles, respectively.

Offering all possible vehicles has three major drawbacks. First, the offer set would be extremely large such that it would take a significant amount of time to browse through the whole assortment, which could annoy the customer and thus harms the customer experience. Second, such a large offer set might overwhelm the customer, such that the customer decides to directly leave the platform without purchase instead of browsing through the offer set. Third, the vehicles offered to the customer might cannibalize each other (see Jena et al. 2020). To see this, think about two vehicles that both satisfy the customer's requirements, whereby the price of one vehicle is significantly lower than the price of the other. The customer would purchase the cheaper vehicle which would result in less revenue for BMW. If the cheaper vehicle had not been offered, the customer would have purchased the more expensive one, resulting in higher revenue for BMW.

Considering the other extreme, offering only one vehicle would be suboptimal as well. When a customer arrives on the new car locator website, it is uncertain which vehicle is desired by this customer. When offering only one vehicle, the probability is high that this vehicle is not of interest to the customer resulting in a direct leave of the

✉ Robert Klein
Robert.Klein@uni-a.de

Julia Heger
Julia.Heger@uni-a.de

[1] Chair of Analytics & Optimization, University of Augsburg, Universitätsstr. 16, 86159 Augsburg, Germany

Published online: 27 May 2024

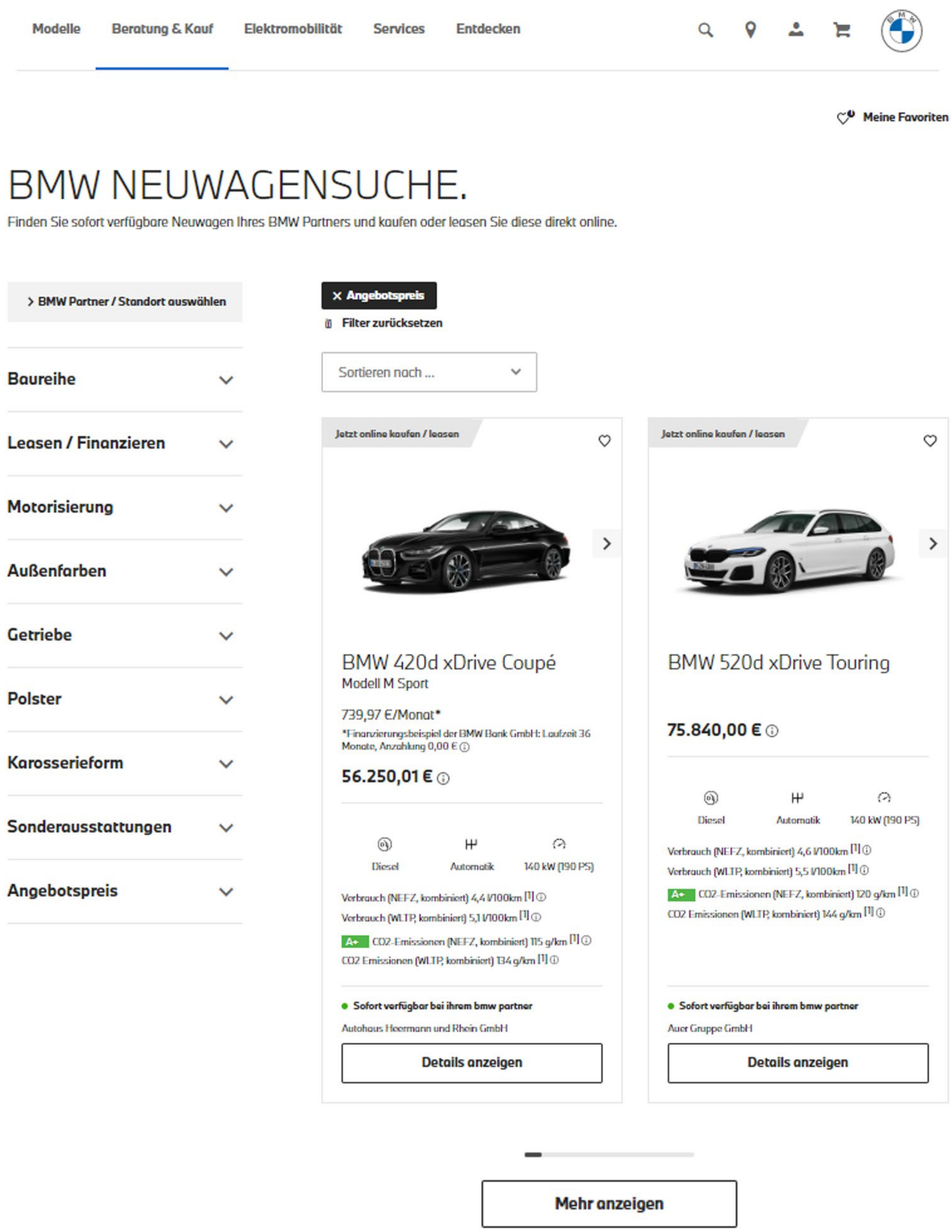



**Fig. 1** BMW new car locator website

new car locator website. The same reasoning holds for small offer sets. Hence, an optimal assortment consisting of a subset of all available vehicles should be offered to the customer that maximizes the expected revenue for BMW, provides sufficient choice options, but neither overwhelms the customer nor results in cannibalization effects.

The goal of the joint project with BMW is to develop a system to automatically determine revenue-maximizing offer sets for the BMW new car locator website while adhering to selected business requirements using mathematical optimization. Such business requirements are typically defined by the sales department but occasionally also result from legal requirements.

The business requirements for the BMW new car locator platform include—but are not limited to—the following demands:

- **no dealer discrimination:** Since the vehicles are sold via dealers, none of them should be discriminated by the offer set.
- **vehicle class diversity:** The offer set should provide sufficient diversity regarding the different BMW vehicle classes called UKL, KKL, MKL, and GKL. These vehicle classes are based on the vehicle size and each consist of multiple vehicle series as visualized in Fig. 2.
- **vehicle series diversity:** It is desired that the offer set covers sufficient diversity regarding the different BMW

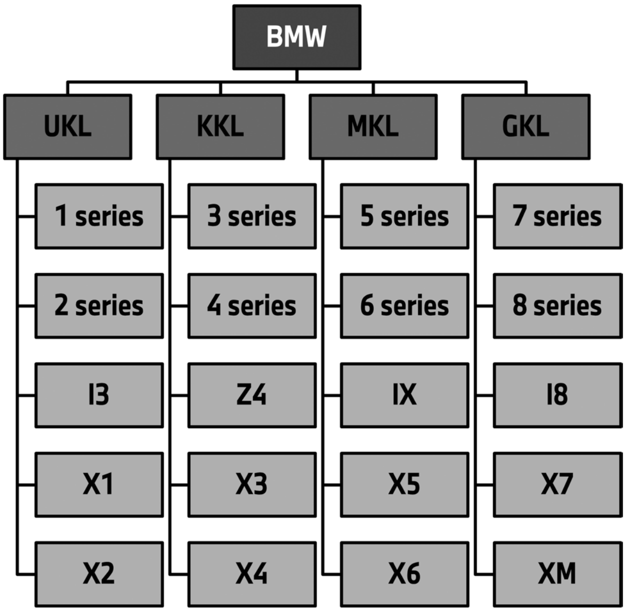



**Fig. 2** BMW vehicle tree depicting the hierarchical structure between the different BMW vehicle classes (UKL, KKL, MKL, GKL) and the corresponding BMW series

vehicle series such as 1 series, 2 series, and so on. A complete overview of all BMW series is provided in the vehicle tree in Fig. 2.

- **vehicle attribute diversity:** The offer set should comprise sufficient diversity regarding the different BMW vehicle attributes such as body type (roadster, touring, convertible, coupé, ...), power train (petrol, diesel, BEV, ...), gear (automatic, manual), color (black sapphire metallic, skyscraper grey metallic, san francisco red metallic, portimao blue metallic, isle of man green metallic, alpine white, ...), and much more.
- **price level diversity:** It is requested that the offer set captures a certain split of the BMW vehicle price levels. These vehicle price levels consist of the classes S, M, and L based on the sales price and the selection of optional equipment the vehicle is equipped with.

When considering only a single of the above business requirements, some of them can be addressed by making use of existing variants of an assortment optimization problem as further detailed in section "Assortment optimization under group constraints". However, other requirements can not yet be handled in this way. Moreover, BMW is not interested in determining an assortment that satisfies only one of the business requirements but all of them at once. When combining multiple of these business requirements, the resulting assortment problem has not yet been examined and analyzed with respect to its efficient solution using in particular standard solvers.

To address all requirements of BMW, we study the assortment optimization problem under the well-known multinomial logit (MNL) choice model and propose a new general type of constraint that we refer to as group constraint. The basic idea behind a group constraint is a mapping between products—in our BMW project the vehicles—and groups—in our BMW project the dealers, the vehicle classes, the vehicle series, the vehicle attributes, or the vehicle price levels, respectively—and there is an individual constraint per group.

To the best of our knowledge, the only type of group constraint studied in the literature on assortment optimization so far incorporates a limit on the maximum number of products to be offered per group when the products are disjointly attributable to the groups, i.e., when every product can be mapped to exactly one group. In this case, the constraint is typically referred to as partition constraint.

In contrast to the existing literature, in our study the mapping between products and groups is arbitrary, i.e., it can but does not necessarily need to be disjoint. Having the mapping between products and groups available, we propose two types of group constraints.

- First, we assume that the mapping between products and groups is arbitrary and the number of products per group that can be offered is lower, upper, or interval bounded. We refer to this case as **products-per-group constraint (PPG)**.
- Second, we assume that the mapping between products and groups is arbitrary and the number of groups from which products can be offered is lower, upper, or interval bounded. This case is referred to as **number-of-offered-groups constraint (NOG)**.

To illustrate both types of constraints, let us consider the BMW requirement of vehicle series diversity, which states that the offer set should cover sufficient diversity regarding the different BMW vehicle series such as 1 series, 2 series, 3 series, and so on. To formulate this requirement as a group constraint, we identify the products as the vehicles that can be sold via the new car locator website and the groups as the different BMW series, whereby the mapping which vehicle belongs to which series is known. Using the *products-per-group constraint*, we can then address the vehicle series diversity requirement by demanding that at least a certain number of vehicles per series—e.g., three 1 series vehicles, one 2 series vehicle, and four 3 series vehicles—are selected. Alternatively, to address the requirement of vehicle series diversity using the *number-of-offered-groups constraint*, we can demand that vehicles from at least a certain number of different series—e.g., from three different series—are offered.

Note that in this example, the formulation of the vehicle series diversity requirement using the products-per-group constraint necessitates the specification of the exact number of vehicles that should be selected per series and is thus more restrictive regarding possible assortments to be offered compared to the formulation using the number-of-offered-groups constraint which only requires the specification of the number of different series from which vehicles should be chosen.

We formulate both the products-per-group constrained and the number-of-offered-groups constrained optimization problems as binary fractional linear program (BFLP) that are known to be NP-hard. Due to the constantly changing requirements on the optimization problem in terms of business demands, a highly flexible, easily adjustable solution approach is needed to continually change the optimization problem accordingly by, e.g., adding further constraints or removing existing ones. Moreover, the instance needs to be quickly solvable and the solution approach must be easy to integrate into the BMW IT infrastructure. Hence, we provide mixed integer linear program (MILP) reformulations of both BFLP problem formulations which can be solved using common state-of-the-art solvers such as CPLEX or Gurobi. For the number-of-offered-groups constrained setting, we additionally provide a two-step solution approach that is typically more efficient in terms of computation time compared to our MILP reformulation.

Finally, to provide evidence of the practical applicability of our approach, we conduct an extensive numerical study for both group-constrained assortment problem settings using synthetic data of various realistic problem set sizes.

This paper is organized as follows: In section "Introduction", we motivate our research and mention our main contributions. In section "Related literature", we provide a brief overview of the existing literature that is related to the assortment problem setting studied by us. We introduce the preliminaries including relevant notation, group constraints in general, the considered MNL model, and the resulting assortment problem in section "Preliminaries". The two assortment problem settings under different group constraints are motivated, formalized, and reformulated in section "Assortment optimization under group constraints". We describe our numerical experiments including data generation and results evaluation in section "Numerical experiments". Finally, we conclude in section "Conclusion".

## Related literature

Research on assortment optimization starts with the seminal paper of Talluri and van Ryzin (2004). The authors study the unconstrained version of the assortment problem and prove that the optimal assortment is revenue-ordered. That is to say, the optimal assortment consists of all products priced above a certain price threshold and can be obtained by greedily adding products into the offered assortment in the order of decreasing revenues until the price threshold is reached.

Since then, research in the field of assortment optimization has experienced a considerable boost in attention—particularly initialized by the transition from independent demand to choice-based revenue management; see Strauss et al. (2018). During this time, various assortment problems have been studied under diverse choice models, so that by now there exists a wide range of literature on assortment optimization. A comprehensive review of the existing literature is, e.g., provided by Heger and Klein (2024).

In our work, we assume that customer demand follows the MNL model and incorporate a new type of constraint—the group constraint—into the retail assortment problem. Therefore, in section "Literature considering retail applications of assortment optimization" we first provide a brief overview of selected retail applications of assortment optimization that have been investigated in the literature so far. Subsequently, we focus on the studies on constrained assortment optimization that are most closely related to our proposed group constraints. Doing so, we first review literature on totally unimodular constraints in section "Literature considering totally unimodular constraints" before briefly commenting on literature covering partition constraints in section "Literature considering partition constraints".

### Literature considering retail applications of assortment optimization

The problem of finding the optimal assortment is omnipresent in various retail applications. These applications range from online and offline to omni-channel settings, from single-period to multi-period considerations, from single-purchase to multi-purchase assumptions, from pure assortment planning tasks to integrating inventory, price, or positioning decisions, from general to personalized recommendations, and from frequently purchased, reusable, or short-lived products to infrequently purchased, non-reusable, or long-lived products. In the following, we provide a brief overview of selected retail applications addressed in the literature on assortment optimization so far.

Qiu et al. (2020) study the task of determining the optimal store assortment, where different products from one store brand and one national brand can be offered and the customer is assumed to first choose which brand to buy before deciding on one product within that brand.

In contrast, Hübner and Schaal (2017) study the task of maximizing a retailer's profit by selecting the optimal store assortment and allocating limited shelf space to its items. Hübner et al. (2020) extend this setting by considering two-dimensional shelves such as those for offering, e.g., meat, bread, fish, cheese, or clothes.

As opposed to such offline settings, the optimal assortment offered to a customer via the online channel can be personalized using the customer's personal information and purchase history. This is, e.g., studied by Bernstein et al. (2015), Bernstein et al. (2019), and Golrezaei et al. (2014). Since the amount of customer information that is available in online settings is typically of high dimensionality, which results in significant computational challenges, Miao and Chao (2022) propose a way of personalizing online assortments while accounting for the high dimensionality of available customer information.

Moreover, Ettl et al. (2019) study the task of recommending a whole personalized discounted product bundle to an online shopper that considers the trade-off between profit maximization and inventory management, while selecting products that are relevant to the consumer's preferences.

Besides the personalization of the offer set, online retail also allows for the frequent introduction and removal of products. To address this, Agrawal et al. (2019) study the assortment problem while accounting for short selling horizons of the considered products. Caro et al. (2014) extend this setting by assuming that, once introduced, a product's attractiveness lasts only a few periods and vanishes over time and that the retailer needs to decide in advance on the release date of each product in a given collection during a selling season to maximize the total profit over the selling season.

In addition, unlike offline settings, in online settings the offered assortment is not necessarily visible all at once. E.g., Liu et al. (2020), Gao et al. (2021), and Feldman and Segev (2022) study the assortment problem of an online retailer whose offered assortment is incrementally shown over multiple results pages.

Apart from solely optimizing the offered assortment, it is also common to study integrated decisions. For example, Chen and Jiang (2020) consider the task of finding the revenue-maximizing subset of products as well as their corresponding display positions. Furthermore, Aouad et al. (2018) study the joint assortment and inventory planning problem where the retailer needs to select an assortment of products along with their initial inventory levels, given a capacity constraint on the total number of units to be stocked. Beyond that, Katsifou et al. (2014) study the joint assortment, inventory, and price optimization problem while accounting for a combined product assortment consisting of both regular 'standard' products and more fashionable and short-lived 'special' products. Such an assortment structure should increase sales due to cross-selling effects—customers attracted by special products might purchase standard products and vice versa.

Instead of purely considering either the online or the offline sales channel, some studies on assortment optimization also deal with omni-channel settings. Among them, Lo and Topaloglu (2021) consider the assortment problem of a retailer that operates a physical store as well as an online store. In the online store, the full assortment is offered and the retailer needs to decide on the assortment to be offered in the physical store to maximize the total expected revenue. Similarly, Miller et al. (2010) study the task of developing an operational methodology for choosing retail assortments for infrequently purchased products such as consumer electronics, appliances, and home furnishings. The authors propose to design the company's website in a way to collect customer preference information, which can then be incorporated into the assortment decision. In contrast, Hense and Hübner (2022) study joint assortment, space, and inventory decisions for an omni-channel retailer operating with interconnected bricks-and-mortar stores and an online shop.

## Literature considering totally unimodular constraints

The study that is most closely related to our research is Davis et al. (2013), who consider the assortment optimization problem under the MNL model with totally unimodular (TU) constraints. TU constraints are a general type of constraint that can be formalized by $A\boldsymbol{x} \leq \boldsymbol{b}$. Here, $\boldsymbol{x}$ denotes the decision variable, $\boldsymbol{b}$ represents a vector that is assumed to be integral, and $A$ is a TU matrix, i.e., a matrix with every square submatrix having determinant $\pm 1$ or 0. These total unimodularity constraints capture a broad range of constraint settings. E.g., the authors argue that a partitioning of products into disjoint groups can be represented by an interval matrix which is totally unimodular. As a key result, Davis et al. (2013) show that the assortment problem under such TU constraints can be formulated as binary fractional linear program and solved as an equivalent linear program.

Following Davis et al. (2013), further studies on assortment optimization under totally unimodular constraint structures appeared. Among them, Sumida et al. (2020) examine the revenue-utility assortment problem under the MNL model while adhering to totally unimodular constraints on the offered assortment. The goal of the revenue-utility assortment problem is to determine an assortment that maximizes a linear combination of the expected revenue of the firm and the expected utility of the customer. The authors provide evidence that the revenue-utility assortment problem can be solved by identifying the assortment that maximizes the expected revenue after adjusting each product's revenue by the same constant.

In contrast to Sumida et al. (2020), who show that the classic assortment problem under the MNL model with TU constraints admits a linear programming-based polynomial time algorithm, Bai et al. (2023) provide evidence that this is not the case when considering the multi-purchase setting where customers can make multiple simultaneous purchases. To be precise, the authors show that the assortment problem under the multi-purchase MNL model is NP-hard to approximate under TU constraints and propose a polynomial time approximation scheme. Likewise, Aouad et al. (2022) prove that the assortment problem with TU constraints under the exponomial choice model is NP-hard as well.

Moreover, literature on TU-constrained optimization also arises in the area of product line design. To tackle this type of problem, typically a two-step approach is followed. At first, a set of candidate products is determined. Subsequently, in the second step, a product line selection—i.e., assortment optimization—and pricing problem is solved on the set of candidate products to determine the optimal product line. Chen and Hausman (2000) study this problem under the MNL model. In their study, the price of each product is chosen from a predetermined set of discrete values and it is ensured that every product profile is offered at only one price to the customer. The length of the product line is controlled by imposing bounds on the number of offered products. Doing so, the authors show that this problem can be formulated as binary fractional linear program with totally unimodular constraints, which is solvable using standard methods. Later on, Schön (2010) extends this approach to determine a profit-maximizing product line under a personalized or group pricing strategy in markets with multiple heterogeneous consumers.

### Literature considering partition constraints

The first study dealing with partition constraints as special case of group constraints is Agrawal et al. (2019). They consider a dynamic assortment problem with partial knowledge about the parameters of the MNL model. In every round the retailer offers a subset of products to a customer. Doing so, the products can be partitioned into disjoint segments, and the retailer can offer at most a specified number of products from each segment. The customer then selects one of these offered products according to the MNL model or leaves the market without purchase. The retailer observes the customers' choice and aims at dynamically learning the MNL model parameters while optimizing cumulative revenues over the whole selling horizon. The authors propose an algorithm that is able to simultaneously explore and exploit.

The second study dealing with a special case of group constraints is Ghuge et al. (2021). The authors study the assortment problem under the paired combinatorial logit model with group constraints where the products are partitioned into disjoint categories and there is a limit on the number of selected products per category. Ghuge et al. (2021) propose a binary-search-based approximation framework combined with a local-search algorithm.

All of the studies introduced above only cover a special case of group constraint where the products are partitioned into disjoint groups and there is a limit on the number of products to be offered per group. Besides those studies, to the best of our knowledge there is no further literature on group-constrained assortment optimization. Particularly, there are no studies capturing the more general settings considered by us, i.e., the case when the products can belong to multiple groups at once or when the number of groups from which products are offered is limited.

## Preliminaries

In section "Products and assortments", we start by defining the notation for products and assortments. Subsequently, in section "Group constraints" we formally introduce our proposed group constraints before briefly presenting the MNL model in section "MNL model". We end the section by expounding the related assortment problem in section "Assortment optimization".

### Products and assortments

Let $N = \{1, ..., n\}$ be the set of available products and denote the no-purchase option by $\{0\}$. The retailer needs to select an assortment of the available products to be offered. This subset—also referred to as offer set—is represented as a binary decision variable $x_j$ for each product $j \in N$ that indicates whether this product is offered or not by setting $x_j = 1$ and $x_j = 0$, respectively. More formally, define

$$x_j = \begin{cases} 1, \text{ if product } j \text{ is offered} \\ 0, \text{ else} \end{cases} \qquad \forall j \in N.$$

Note that the no-purchase option is always offered, implying that $x_0 = 1$ holds for any assortment. The number of options that are offered in an assortment can be obtained by summing over all $x_j$, i.e., $\sum_{j \in N} x_j$.

The expected demand for any product depends on the substitution behavior of customers and is captured by the MNL model specifying the probability that a customer selects a particular option from a given offer set as detailed in section "MNL model". In general, the customer selects alternative $j \in N$ with probability $p_j(\boldsymbol{x})$ given that assortment $\boldsymbol{x}$ is offered and decides to not purchase anything with probability $p_0(\boldsymbol{x})$. Finally, assume that product $j \in N$ is sold at revenue $r_j$ and that the market is of size 1 without loss of generality.

This general setting can be transferred to our BMW project by identifying the products $N = \{1, ..., n\}$ to be the vehicles that can be offered via the new car locator website at revenues $r_j$, $j \in N$. A vehicle $j \in N$ is then purchased with probability $p_j(\boldsymbol{x})$ given that the assortment $\boldsymbol{x}$ of vehicles is offered and the customer decides to leave the new car locator without purchasing a vehicle with probability $p_0(\boldsymbol{x})$.

## Group constraints

Group constraints are a type of constraint under which the products are attributed to groups; more formally, assume that the set of products $N$ can be attributed to $k$ groups $K = \{1, ..., k\}$. This can be represented by a matrix $A \in \{0,1\}^{k \times n}$ whose rows correspond to the $k$ groups and the columns to the $n$ products. The entries $a_{ij}$ of the matrix $A$ are either 0 or 1 depending if product $j$ is contained in group $i$ or not, i.e., more formally:

$$a_{ij} = \begin{cases} 1, \text{ if } j \in i \\ 0, \text{ if } j \notin i \end{cases} \quad \forall j \in N, i \in K.$$

Each product can be either assigned to exactly one or to multiple groups:

- In the former case, thinking about our project with BMW, the resulting disjoint groups can, e.g., be thought of as the different BMW vehicle classes UKL, KKL, MKL, and GKL, whereby each vehicle belongs to exactly one of these four vehicle classes. This results in a totally unimodular matrix $A$. As mentioned in section "Related literature", the assortment problem under such totally unimodular constraint structures is, e.g., studied by Davis et al. (2013) and Sumida et al. (2020).
- Transferring the latter case to our project with BMW, the groups might, e.g., correspond to the various different BMW vehicle attributes such as body type roadster, body type touring, manual gearing, automatic gearing, color black sapphire metallic, and color alpine white, whereby each vehicle can have multiple such attributes at once—think, e.g., about an alpine white touring with automatic gearing. Hence, intuitively speaking the matrix $A$ can have multiple 1's per column and is thus not necessarily totally unimodular.

Having the assignment of products to groups available, one can either bound the number of offered products per group or the number of groups from which products are offered.

For the former setting, one can make use of the fact that the multiplication of matrix $A$ with the assortment vector $\boldsymbol{x}$ yields the number of offered products per group. This number can be bound from below and/or above by using a constraint of the form $A\boldsymbol{x} \geq \boldsymbol{b}$ and/or $A\boldsymbol{x} \leq \boldsymbol{b}$ respectively to limit the number of offered products per group to $\boldsymbol{b}$. This setting is described in further detail in section "Products-per-group constraint".

To incorporate the latter setting into a constraint, an additional decision variable needs to be introduced that checks whether at least one product per group is offered. This setting is discussed in further detail in section "Number-of-offered-groups constraint".

## MNL model

Intuitively, choice models aim at capturing the demand behavior of customers and can thus be used to model which of the offered options might be purchased by a customer. In recent years, the assortment problem has been studied under a variety of choice models that can be divided into parametric and non-parametric approaches as detailed by Heger and Klein (2024) and Strauss et al. (2018).

Non-parametric choice models are typically designed as ranked lists of preferences, also referred to as customer types, whereby demand is modeled by a probability distribution over all customer types. Under such rank list-based models, the customer then chooses the highest-ranking offered item in the list or leaves without purchase if none of the available items ranks higher than the no-purchase option.

In contrast, parametric choice models are based on random utility theory, where it is assumed that customers associate a certain utility with every product, and decide on the alternative that maximizes their utility. This framework is referred to as random utility maximization (RUM). Within this framework, the utility $U_j = u_j + \epsilon_j$ of product $j$ is composed of the deterministic part $u_j$ and a random component $\epsilon_j$. Using this, the probability $p_j(\boldsymbol{x})$ that product $j$ is chosen among the offered assortment is given by the probability that this product is associated with the highest utility, i.e.,

$$p_j(\boldsymbol{x}) = P\Big(U_j \geq U_i \; \forall i : x_i = 1\Big).$$

The deterministic component $u_j$ can be expressed as a linear function $u_j = \beta^T \boldsymbol{d_j}$ of an attribute vector $\boldsymbol{d_j}$ that influences the purchase probabilities. The different parametric choice models result from different assumptions made on the distribution of the random component $\epsilon_j$.

One of the most famous parametric demand models that is particularly popular due to its simplicity is the MNL model of Luce (1959) and McFadden (1973). This model can be identified as member of the RUM framework by choosing the random components $\epsilon_j$ to be iid. random variables that follow the Gumbel distribution with a common scale parameter, typically normalized to one, and location parameters $u_j$, $j \in N$ with $u_0 := 0$. $e^{u_j}$ is referred to as preference weight $v_j$ of alternative $j \in N$. Under the MNL model, the probability to select a product $j$ from the offer set is determined by its utility relative to the total utility of the offer set; more formally:

$$p_j(\boldsymbol{x}) = \frac{e^{u_j} x_j}{1 + \sum_{i \in N} e^{u_i} x_i} = \frac{v_j x_j}{1 + \sum_{i \in N} v_i x_i}.$$

In our project with BMW, we assume that customer demand follows the above described MNL model. This decision is made for three reasons – 1) the lack of availability of customer segmentation on BMW websites, 2) the MNL model's intuitive appeal, and 3) its mathematical tractability. We briefly detail on each of the three reasons in the following.

1. When aiming to purchase a vehicle, typically, a customer already has some vehicle characteristics in mind. Think, e.g., about a parent with several children. This customer would likely be interested in big, secure Sports Activity Vehicles (that is what BMW calls SUVs) rather than in a roadster with only two seats. Hence, there might exist customer segments covering the diverse interests and needs of the different customers and thus specifying the purchase dynamics of the customers belonging to the respective segments. This, in turn, implies that it would make sense to model the customer's choice behavior using, e.g., a mixed logit or a nested logit model. However, by now, there is no such segmentation of customers entering the BMW new car locator website available. In contrast to other choice models, the MNL model does not require information on different customer segments.
2. As detailed before, under the MNL model, the utility of each vehicle is composed of a random component following a Gumbel distribution and a deterministic part that can be expressed as a linear function of attributes influencing the purchase probabilities. This clear dependence on the utilities and, thus, of the purchase probabilities on certain known attributes, such as product characteristics, makes the MNL model both intuitive and easily interpretable, which is of high importance for companies when it comes to decision-making.
3. Under the MNL model, each vehicle's purchase probability can be described in closed form as its utility relative to the total utility of all offered vehicles. This formulation results in a fractional linear objective function of the considered assortment problem, which can be linearized using adequate transformation techniques. That way, the assortment problem can be reformulated into a tractable mixed integer linear program that can be solved efficiently. This is not necessarily the case when considering other, more complex choice models.

## Assortment optimization

Assortment optimization refers to the problem of analytically determining a selection of products to be offered to customers in order to maximize the expected revenues. Using the notation introduced in section "Products and assortments", the classic assortment problem can be formalized by

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \sum_{j \in N} p_j(\boldsymbol{x}) \cdot r_j \\ \text{subject to} \quad & x_j \in \{0, 1\} \qquad \forall j \in N \end{aligned} \tag{AOP}$$

In the above optimization problem (AOP), $\boldsymbol{x} \in \{0,1\}^n$ denotes the decision variable indicating which of the products $j \in N$ are included in the offer set. The objective represents the expected revenue and sums up the expected revenue obtained by each product $j$. The expected revenue per product is obtained by multiplying the purchase probability $p_j(\boldsymbol{x})$ of the product with its revenue $r_j$. The optimization problem formulation is completed by a binary constraint ensuring that all values of $x_j$ are either 0 or 1.

# Assortment optimization under group constraints

In this section, we introduce two types of group constraints, both of which are motivated by our joint project with BMW. In sections "Products-per-group constraint" and "Number-of-offered-groups constraint", we motivate the constraint settings, formulate the assortment problems as binary fractional linear programs and reformulate them as MILP that can be solved using common standard solvers. Note that we do not explicitly incorporate cardinality constraints—i.e., constraints limiting the total number of offered products—in our assortment problems. However, extending the problem formulations by a cardinality constraint on the assortment size is straight forward and does not affect our proposed MILP reformulations.

## Products-per-group constraint

We consider a type of group constraint where all products from the whole universe of products can be assigned to multiple groups at once and the number of products per group that can be offered is limited.

### Motivation

Recall our joint project with BMW introduced in section "Introduction". As mentioned, the goal of this project is to determine a revenue-maximizing offer set for the BMW new car locator website while adhering to multiple business requirements. These requirements include the prevention of dealer discrimination as well as the diversity of vehicle classes, vehicle series, vehicle attributes, and price levels.

Some of these business demands on the offer set of the new car locator website—such as, e.g., the requirements regarding the diversity of vehicle classes, vehicle series, or price levels—can be individually addressed by making use of the AOP setting under totally unimodular constraint structures introduced in Davis et al. (2013). That is, the corresponding AOP could be solved efficiently when exactly one of the requirements had to be considered.

To see this, let us consider the case of vehicle series diversity. First, we identify the products as vehicles that can be attributed to groups—in our example the different BMW series. Note that at BMW, every vehicle is defined to belong to exactly one series. Hence, the mapping of vehicles to series is disjoint and can thus be represented by a TU matrix $A$ where the rows correspond to the different BMW series and the columns represent the vehicles. Requiring that at least one vehicle per series is contained in the offer set already satisfies the vehicle series diversity requirement. This can be done by adding a constraint of the form $A\boldsymbol{x} \geq \boldsymbol{b}$ to the assortment problem, which might look as follows:

$$\begin{matrix} \text{1 series} \\ \text{2 series} \\ \vdots \\ \text{8 series} \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geq \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

The same holds for the diversity requirements regarding vehicle classes or price levels. Instead of identifying the groups as BMW series as done in our previous example, the groups can be chosen to be the different vehicle classes or price levels, respectively. Again, the attribution of vehicles to classes or price levels is disjoint, resulting in a totally unimodular matrix.

However, as mentioned before, BMW is not interested in determining an assortment that satisfies only one of their business requirements but all of them at once. To combine multiple business requirements that can be individually formulated as group constraints, one can row-wise concatenate their respective matrices $A$ and the vectors $\boldsymbol{b}$ to obtain a new constraint of the form $A\boldsymbol{x} \geq \boldsymbol{b}$. However, when concatenating multiple totally unimodular matrices, the resulting matrix is not necessarily totally unimodular anymore. Likewise, concatenating matrices where at least one of them is not totally unimodular does not result in a TU matrix. Hence, the case of adhering to multiple business requirements at once is not addressable by making use of the assortment problem with totally unimodular constraint structures proposed in Davis et al. (2013).

In addition, some of the introduced business requirements such as the diversity requirement regarding product attributes and the prevention of dealer discrimination can not be formulated as TU constraints—even when addressing them individually. To see this, let us consider the requirement of product attribute diversity. First, we again identify the products as vehicles. These vehicles can be characterized by the specification of their product attributes. Examples of product attributes (and their specifications) include body type (roadster, touring), power train (petrol, diesel, BEV), gear (automatic, manual), and color (black sapphire metallic, skyscraper grey metallic, san francisco red metallic, portimao blue metallic, isle of man green metallic, alpine white). We identify the groups to be the specifications of the product attributes, i.e., for example body type roadster, body type touring, power train petrol and so on. To ensure sufficient diversity regarding these vehicle attributes, we can simply require that at least a certain number of vehicles per attribute is offered. That is to say, e.g., 10 roadsters, 15 tourings, 20 petrols, 10 diesel, 30 BEVs, 25 vehicles with automatic gearing, and 5 alpine white vehicles. This can be done by adding a constraint of the form $A\boldsymbol{x} \geq \boldsymbol{b}$ to the assortment problem, which might look as follows:

$$\begin{matrix} \text{body type roadster} \\ \text{body type touring} \\ \text{power train petrol} \\ \text{power train diesel} \\ \text{power train BEV} \\ \text{automatic gearing} \\ \vdots \\ \text{color alpine white} \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{pmatrix} \geq \begin{pmatrix} 10 \\ 15 \\ 20 \\ 10 \\ 30 \\ 25 \\ \vdots \\ 5 \end{pmatrix}.$$

The mapping between vehicles and vehicle attributes is represented by the attribute-vehicle mapping matrix $A$, whose rows correspond to the attributes and the columns to the vehicles. Clearly, this mapping is not disjoint, since a vehicle can be an alpine white petrol roadster with automatic gearing and thus belongs to multiple groups at once. Hence, the mapping matrix $A$ is not necessarily totally unimodular.

The requirement regarding the prevention of dealers can be treated analogously to this example by replacing the vehicle attributes with the different dealers. Since a vehicle can be sold via multiple dealers, the mapping between vehicles and dealers is again not disjoint such that the matrix $A$ is not totally unimodular.

This implies that even simple individual business requirements such as the diversity regarding product attributes or the prevention of dealer discrimination—let alone the combination of multiple requirements—can not be addressed by making use of the assortment problem with totally unimodular constraint structures proposed in Davis et al. (2013).

However, it is possible to incorporate both, the business requirements that can not be formulated using totally unimodular mapping matrices as well as the combination of multiple business cases into the assortment problem formulation. This can be achieved by adding our proposed products-per-group constraint of the form $A\boldsymbol{x} \geq \boldsymbol{b}$ to the assortment problem formulation, where no special assumptions are to be made regarding the structure of the group-product mapping matrix $A$. Adding this constraint results in a feasible region of the form $F = \{\boldsymbol{x} \in \{0,1\}^n \mid A\boldsymbol{x} \geq \boldsymbol{b}\}$.

Note that not only 'greater-or-equal' (geq) constraints can be incorporated. Instead, it is of course also possible to formulate a 'less-or-equal' (leq) constraint by limiting the feasible region to $F = \{\boldsymbol{x} \in \{0,1\}^n \mid A\boldsymbol{x} \leq \boldsymbol{b}\}$. By combining both constraints, an equality requirement can be formulated, giving a feasible region of the form $F = \{\boldsymbol{x} \in \{0,1\}^n \mid A\boldsymbol{x} \geq \boldsymbol{b}, A\boldsymbol{x} \leq \boldsymbol{b}\}$.

Moreover, note that the settings where the number of products per group should be at most one ($A\boldsymbol{x} \leq \mathbf{1}$) or exactly one ($A\boldsymbol{x} = \mathbf{1}$) can also be used to solve joint assortment and pricing problems by identifying each product-price combination as a new individual product and applying the assortment problem to this extended product set.

## Problem formulation

In general, using $p_j(\boldsymbol{x}) = \frac{v_j x_j}{1+\sum_{i \in N} v_i x_i}$ under the MNL model, the assortment problem under the products-per-group constraint motivated in section "Motivation" can be formulated as a Binary Fractional Linear Program (BFLP):

$$\text{(BFLP PPG)} \quad \max_{x} \quad \frac{\sum_{j \in N} r_j v_j x_j}{1 + \sum_{i \in N} v_i x_i} \tag{1}$$

$$\text{subject to} \quad \sum_{j \in N} a_{ij} x_j \geq b_i \quad \forall i \in K \tag{2}$$

$$x_j \in \{0,1\} \quad \forall j \in N \tag{3}$$

Due to its fractional objective, this problem formulation is not directly solvable using a common solver such as CPLEX or Gurobi and must thus be reformulated for this purpose. However, in contrast to the problem setting considered in Davis et al. (2013), in this case the matrix $A$ is not totally unimodular and thus, the LP-reformulation of the optimization problem introduced in Davis et al. (2013) is not valid anymore. Though, application of the modified Charnes and Cooper (1962) transformation for linear-fractional programming with $w_j = w_0 x_j$ gives:

$$\text{(MILP-v0 PPG)} \quad \max_{w, w_0} \quad \sum_{j \in N} r_j v_j w_j \tag{4}$$

$$\text{subject to} \quad w_0 + \sum_{j \in N} v_j w_j = 1 \tag{5}$$

$$\sum_{j \in N} a_{ij} w_j \geq b_i w_0 \quad \forall i \in K \tag{6}$$

$$w_j \in \{0, w_0\} \quad \forall j \in N \tag{7}$$

$$0 \leq w_0 \leq 1 \tag{8}$$

Here, $w_j = \frac{x_j}{1+\sum_{i \in N} v_i x_i}$ can be interpreted as the probability that a customer purchases product $j$ when assortment $\boldsymbol{x}$ is offered, whereas $w_0 = \frac{1}{1+\sum_{i \in N} v_i x_i}$ denotes the probability that the customer does not purchase anything. Constraint (7) enforces $w_j$ to take values from the discrete set $\{0, w_0\}$. Such a constraint is not directly implementable in solvers such as Gurobi but can be handled as follows. Let $w_j' w_0 = w_j$ to get the equivalent formulation of the above optimization problem:

$$\text{(MIQP PPG)} \quad \max_{w', w_0} \quad \sum_{j \in N} r_j v_j w_j' w_0 \tag{9}$$

$$\text{subject to} \quad w_0 + \sum_{j \in N} v_j w_j' w_0 = 1 \tag{10}$$

$$\sum_{j \in N} a_{ij} w_j' w_0 \geq b_i w_0 \quad \forall i \in K \tag{11}$$

$$w_j' \in \{0,1\} \quad \forall j \in N \tag{12}$$

$$0 \leq w_0 \leq 1 \tag{13}$$

The problem formulation (MIQP PPG) is a mixed integer quadratic program (MIQP) since its objective consists of the product of the decision variables $w_j'$ and $w_0$. Replacing $z_j = w_j' w_0$ and adding the McCormick (1976) envelope inequalities yields the following MILP reformulation of the original problem:

$$\text{(MILP-v1 PPG)} \quad \max_{w', w_0, z} \quad \sum_{j \in N} r_j v_j z_j \tag{14}$$

$$\text{subject to} \quad w_0 + \sum_{j \in N} v_j z_j = 1 \tag{15}$$

$$\sum_{j \in N} a_{ij} z_j \geq b_i w_0 \quad \forall i \in K \tag{16}$$

$$z_j \geq 0 \quad \forall j \in N \tag{17}$$

$$z_j \geq w_0 + w_j' - 1 \quad \forall j \in N \tag{18}$$

$$z_j \leq w_0 \quad \forall j \in N \tag{19}$$

$$z_j \leq w'_j \quad \forall j \in N \tag{20}$$

$$w'_j \in \{0,1\} \quad \forall j \in N \tag{21}$$

$$0 \leq w_0 \leq 1 \tag{22}$$

Note that (MIQP PPG) can already be solved using standard solvers such as CPLEX or Gurobi. However, as tested by the authors, this would take between 2.4 times and 6.4 times—and on average 4 times—as long as solving (MILP-v1 PPG). Hence, in terms of computation time it is preferable to execute the last transformation step to make use of our final MILP reformulation (MILP-v1 PPG).

## Number-of-offered-groups constraint

In this section, we consider a type of group constraint where all products from the whole universe of products are assigned to groups and the number of groups from which products can be offered is limited.

### Motivation

Again recall our joint project with BMW with the goal of determining a revenue-maximizing offer set for the BMW new car locator website while adhering to multiple business requirements. These business requirements include the prevention of dealer discrimination as well as the diversity of vehicle classes, vehicle series, vehicle attributes, and price levels. In section "Products-per-group constraint", it is shown that all of these requirements as well as their combination can be addressed by incorporating a products-per-group constraint into the assortment problem formulation. However, for some of these requirements, a products-per-group constraint might be a little too strong.

To see this, imagine that BMW requires the offer set to contain sufficient diversity regarding the vehicle colors, as it is assumed that the customer's attention to the vehicle set displayed on the website persists longer if different colors are shown. Using a products-per-group constraint to ensure that this requirement is met, we can, e.g., demand that at least one vehicle per color is offered. However, as the number of stock vehicles for uncommon colors can be quite low, this requirement might be overly restrictive. To ensure adequate color diversity, it would already be sufficient to offer vehicles in, e.g., at least three different colors.

To incorporate such a requirement of limiting the number of colors in which vehicles are offered, the mapping of vehicles to colors is needed. As before, this mapping can be represented by a color-vehicle matrix $A$ where the rows denote the groups—i.e., the colors—and the columns correspond to the vehicles. Note that in this setting, we again do not impose any structure on the mapping matrix $A$, i.e., in particular this matrix must not necessarily be TU.

Next, one can verify whether at least one vehicle per color is offered. This is achieved by checking if $a_i\boldsymbol{x} \geq 1$ for $i \in K$ where $a_i$ denotes row $i$ of matrix $A$. Finally, the number of colors for which at least one vehicle is offered are summed up; more formally, $\sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}}$. The resulting number can then be bounded from above by $m$ to limit the maximum number of colors from which vehicles can be offered. This gives the feasible region $F = \{\boldsymbol{x} \in \{0,1\}^n \mid \sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}} \leq m\}$. Likewise, the total number of colors in which vehicles are offered can also be bounded from below, giving the feasible region $F = \{\boldsymbol{x} \in \{0,1\}^n \mid \sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}} \geq m\}$. Finally, combining both feasible regions defines the exact number of colors in which vehicles can be offered, i.e., $F = \{\boldsymbol{x} \in \{0,1\}^n \mid \sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}} = m\}$.

To incorporate the indicator function into the optimization problem, another decision variable $\boldsymbol{y} \in \{0,1\}^k$ needs to be introduced that indicates whether at least one product per group is offered or not, i.e., more formally

$$y_i = \begin{cases} 1, \text{ if at least one product from group } i \text{ is offered} \\ 0, \text{ else} \end{cases} \quad \forall i \in K.$$

Summing up the elements of $\boldsymbol{y}$ yields the number of groups from which products are offered, i.e., in our example the number of colors in which at least one vehicle is offered. This number can then be restricted using a leq, geq, or equality constraint. For our example, the case of a geq constraint on the number of colors in which vehicles should be offered might look as follows, where $D$ is a constant with a value of at least $n$:

$$\begin{matrix} \text{black sapphire metallic} \\ \text{skyscraper grey metallic} \\ \vdots \\ \text{isle of man green metallic} \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq D \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix},$$

$$\begin{matrix} \text{black sapphire metallic} \\ \text{skyscraper grey metallic} \\ \vdots \\ \text{isle of man green metallic} \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geq \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix},$$

$$\sum_{i \in K} y_i \geq m.$$

Note that the same approach can be taken to incorporate a number-of-offered-groups constraint for the other business requirements such as the vehicle class diversity, the vehicle series diversity, the price level diversity, or the prevention of dealer discrimination. To do so, in our example the vehicle colors must be replaced by the vehicle classes, series, price levels, or dealers, respectively.

### Problem formulation

In general, the optimization problem can be formulated as

$$\text{(BFLP NOG)} \quad \max_{x,y} \quad \frac{\sum_{j\in N} r_j v_j x_j}{1 + \sum_{i\in N} v_i x_i} \tag{23}$$

$$\text{subject to} \quad \sum_{j\in N} a_{ij} x_j \leq D \cdot y_i \quad \forall i \in K \tag{24}$$

$$\sum_{j\in N} a_{ij} x_j \geq y_i \quad \forall i \in K \tag{25}$$

$$\sum_{i\in K} y_i \geq m \tag{26}$$

$$x_j \in \{0, 1\} \quad \forall j \in N \tag{27}$$

$$y_i \in \{0, 1\} \quad \forall i \in K \tag{28}$$

In the above optimization problem, $D$ denotes a constant that is at least as large as the largest possible value of $a_i x$ $\forall i \in K$, i.e., $n$ in our case.

We again aim at reformulating this problem as MILP. To this end, we first apply the modified (Charnes and Cooper 1962) transformation for linear-fractional programming with $w_j = w_0 x_j$. This gives

$$\text{(MILP-v0 NOG)} \quad \max_{w,w_0,y} \quad \sum_{j\in N} r_j v_j w_j \tag{29}$$

$$\text{subject to} \quad w_0 + \sum_{j\in N} v_j w_j = 1 \tag{30}$$

$$\sum_{j\in N} a_{ij} w_j \leq D \cdot w_0 \cdot y_i \quad \forall i \in K \tag{31}$$

$$\sum_{j\in N} a_{ij} w_j \geq w_0 \cdot y_i \quad \forall i \in K \tag{32}$$

$$\sum_{i\in K} y_i \geq m \tag{33}$$

$$w_j \in \{0, w_0\} \quad \forall j \in N \tag{34}$$

$$0 \leq w_0 \leq 1 \tag{35}$$

$$y_i \in \{0, 1\} \quad \forall i \in K \tag{36}$$

Constraint (34) enforces $w_j$ to take values from the discrete set $\{0, w_0\}$. Such a constraint is not directly implementable in solvers such as Gurobi but can be handled as follows. Let $w'_j w_0 = w_j$ to get the equivalent formulation of the above optimization problem:

$$\text{(MIQP NOG)} \quad \max_{w',w_0,y} \quad \sum_{j\in N} r_j v_j w'_j w_0 \tag{37}$$

$$\text{subject to} \quad w_0 + \sum_{j\in N} v_j w'_j w_0 = 1 \tag{38}$$

$$\sum_{j\in N} a_{ij} w'_j \leq D \cdot y_i \quad \forall i \in K \tag{39}$$

$$\sum_{j\in N} a_{ij} w'_j \geq y_i \quad \forall i \in K \tag{40}$$

$$\sum_{i\in K} y_i \geq m \tag{41}$$

$$w'_j \in \{0, 1\} \quad \forall j \in N \tag{42}$$

$$0 \leq w_0 \leq 1 \tag{43}$$

$$y_i \in \{0, 1\} \quad \forall i \in K \tag{44}$$

The problem formulation (MIQP NOG) is again a mixed integer quadratic program since its objective consists of the product of the decision variables $w'_j$ and $w_0$. Replacing $z_j = w'_j w_0$ and adding the McCormick (1976) envelope inequalities yields the following MILP reformulation of the original problem:

$$\text{(MILP-v1 NOG)} \quad \max_{w',w_0,z,y} \quad \sum_{j\in N} r_j v_j z_j \tag{45}$$

$$\text{subject to} \quad w_0 + \sum_{j\in N} v_j z_j = 1 \tag{46}$$

$$\sum_{j\in N} a_{ij} w'_j \leq D \cdot y_i \quad \forall i \in K \tag{47}$$

$$\sum_{j\in N} a_{ij} w'_j \geq y_i \quad \forall i \in K \tag{48}$$

$$\sum_{i \in K} y_i \geq m \tag{49}$$

$$z_j \geq 0 \quad \forall j \in N \tag{50}$$

$$z_j \geq w_0 + w'_j - 1 \quad \forall j \in N \tag{51}$$

$$z_j \leq w_0 \quad \forall j \in N \tag{52}$$

$$z_j \leq w'_j \quad \forall j \in N \tag{53}$$

$$w'_j \in \{0, 1\} \quad \forall j \in N \tag{54}$$

$$0 \leq w_0 \leq 1 \tag{55}$$

$$y_i \in \{0, 1\} \qquad \forall i \in K \tag{56}$$

Note that it might be the case that the optimal solution of the unconstrained assortment problem (AOP) already satisfies the number-of-offered-groups constraint that products from at least (at most) a certain number of groups are offered. In this case, the optimal assortment under the number-of-offered-groups constraint equals the optimal solution of the unconstrained problem and is thus revenue-ordered as shown in Talluri and van Ryzin (2004). In contrast, in case the number-of-offered-groups constraint requires that products from more (less) groups than in the optimal unconstrained assortment are offered, the optimal constrained assortment is not necessarily revenue-ordered anymore. This is in line with the finding of Wang (2013), who shows that the consideration of a simple cardinality constraint already implies that revenue-ordered assortments are not optimal. The above finding allows for the introduction of a two-step solution approach. This two-step approach proceeds as outlined in Algorithm 1.

## Numerical experiments

In this section, we conduct an extensive numerical study for the two group-constrained assortment problem settings introduced in sections "Products-per-group constraint" and "Number-of-offered-groups constraint" using synthetic data of various realistic problem set sizes. We start by commenting on the data generation process in section "Data generation" before discussing the results of our numerical experiments in section "Numerical results".

### Data generation

Since we are not allowed to disclose real BMW data, we execute our numerical experiments using synthetic data. Nevertheless, we base the selection of our parameters $n, k$, and $m$ on the experience from our project with BMW. The structure underlying the synthetic data is based on suitable economic reasoning. In the following, we first describe the process of generating the products, the groups, the group-product matrix, the preference weights, and the revenues before expounding on how we set the products-per-group constraint and the number-of-offered-groups constraint, respectively.

#### Products, groups, group-product matrix, preference weights, and revenues

We vary the number of **products** over $n \in \{250, 500, 1000\}$ and the number of **groups** over $k \in \{10, 50, 100\}$. These problem set sizes correspond to the real-world sizes known from our project with BMW. The products and groups are generated as integer vectors from 1 to $n$ and 1 to $k$, respectively.

The **group-product matrix** $A$ is constructed by randomly assigning zeros or ones to its entries, whereby the number of 1's per column is randomly selected from

**Algorithm 1** Two-step approach for solving the number-of-offered-groups constraint problem (Two-step NOG)

1: Solve the unconstrained problem of Talluri and van Ryzin (2004).
2: Evaluate $\sum_{i=1}^{k} 1_{\{a_i \boldsymbol{x} \geq 1\}}$.
3: **if** $\sum_{i=1}^{k} 1_{\{a_i \boldsymbol{x} \geq 1\}} \geq m$ (or $\leq m$) is already satisfied **then**
4: Optimal constrained solution := optimal unconstrained solution.
5: **else**
6: Solve (MILP-v1 NOG).
7: **end if**

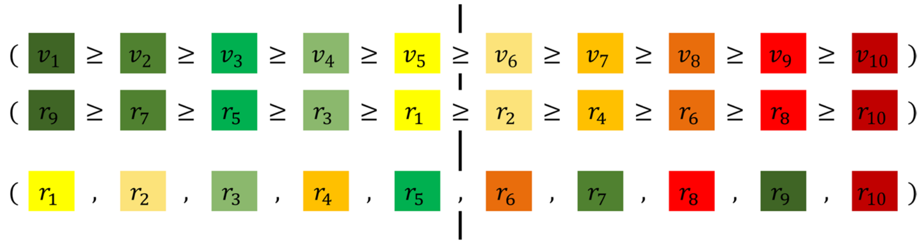



**Fig. 3** Exemplary visualization of the assumed relationship between the preference weights $v_j$, $j \in \{1, ..., 10\}$ and the corresponding revenues $r_j$, $j \in \{1, ..., 10\}$. The black vertical line each indicates the middle of the preference weight vector and the revenue vectors, respectively. The colors indicate the magnitude of the preference weights and revenues from high (green) over medium (yellow) to low (red)

values between 1 and $\max(2, \lceil k/20 \rceil)$. That way, every product belongs to between 1 and 2 ($k = 10$), 3 ($k = 50$), or 5 ($k = 100$) groups, which is realistic.

To generate the **preference weights** $v_j$, we first sample the utilities $u_j$ uniformly at random over the interval $[-1, 4]$, whereby the utility of the no-purchase option is set to 0. Doing so, 20% of the products are modeled to be less preferred than the no-purchase option, whereas the others are more preferred. Based on this, the preference weights are obtained by applying the exp function, i.e., $v_j = e^{u_j}$, such that the preference weight $v_0$ of the no-purchase option equals 1.

We follow Ghuge et al. (2021) and generate the **revenues** $r_j$ uniformly over the interval [0, 1]; the revenue $r_0$ associated with the no-purchase option is set to 0. According to Bechler et al. (2021), there exists a well-known relationship between product revenues and customer preferences—the so-called price compromise according to which customers tend to purchase neither the most expensive nor the cheapest alternative but prefer the mid-priced one whereby their preferences decrease when prices become higher or lower than average. To capture this real-world behavior, we sort both the preference weights and the revenues in descending order and then re-sort the revenues in a way such that the middle highest revenue corresponds to the highest preference weight, the second middle highest revenue corresponds to the second highest preference weight and so on.

This procedure is exemplarily visualized in Fig. 3 for $n = 10$ products. The preference weights $v_j$ depicted in the first row and the revenues $r_j$ shown in the second row of the figure are sampled as described above. The revenues provided in the last row of the figure are obtained by resorting the generated revenues as described in the previous paragraph.

#### Products-per-group constraint

For generating the constraint vector $\boldsymbol{b}$, we study a selection of four different business strategies. As mentioned in section "Introduction", companies typically aim at enhancing assortment diversity—to, e.g., incorporate business requirements – or at limiting the size of the offer set—be it due to space constraints, to not overwhelm the customer, or to satisfy certain business decisions such as not to disclose the number of available products (per group) to competitors. Our selection of business strategies covers a variety of realistic approaches for setting the minimum (geq) or maximum (leq) number of products to be offered per group that might be interesting for companies in practice since these approaches, e.g., enforce to offer products to the customer that would not have been provided otherwise (geq), or prevent the inclusion of products to the assortment that would have been contained otherwise (leq). We present our selection of business strategies in the following:

1. First, we implement the business strategy that at least (geq) / at most (leq) one product per group is offered by defining $\boldsymbol{b}$ to be a vector of ones. In the former case, this strategy allows to raise the diversity of the offer set by offering at least one product from every group and thereby showcasing every group. In contrast, in the second case, this strategy enables to provide an overview of the relevant product portfolio by offering one product from revenue relevant groups and not showing products from all other groups. We refer to this strategy as 'one per group'.
2. Second, we model the business strategy that more products are required (geq) / allowed (leq) to be offered from groups for which lots of products are available. To do so, we set $\boldsymbol{b}$ to a percentage $\rho$ of the number of products available per group according to matrix $A$. For our study, we vary $\rho \in \{2.5\%, 5\%, 10\%, 20\%\}$. This setting is referred to as 'avail. $\rho$' (short for availability with percentage $\rho$).
3. Third, we incorporate the popularity of the groups into the 'avail.' strategy such that we do not require at least/at most $\rho$ percent of the products to be offered from every group, but relax this constraint the more preferred a

**Table 1** Exemplary percentage values of products to be offered per group under the business strategies 'avail. 20%', 'avail. pref. incr. 20%', and 'avail. pref. decr. 20%' for $k = 10$ groups

| | Avail. 20% | Avail. pref. incr. 20% | | | Avail. pref. decr. 20% | | |
|---|---|---|---|---|---|---|---|
| Group | product pct. (geq/leq) | pct. preference rank | product pct. (geq) | product pct. (leq) | pct. preference rank | product pct. (geq) | product pct. (leq) |
| 1 | 20% | 100% | 20% | 38% | 10% | 2% | 20% |
| 2 | 20% | 90% | 18% | 36% | 20% | 4% | 22% |
| 3 | 20% | 80% | 16% | 34% | 30% | 6% | 24% |
| 4 | 20% | 70% | 14% | 32% | 40% | 8% | 26% |
| 5 | 20% | 60% | 12% | 30% | 50% | 10% | 28% |
| 6 | 20% | 50% | 10% | 28% | 60% | 12% | 30% |
| 7 | 20% | 40% | 8% | 26% | 70% | 14% | 32% |
| 8 | 20% | 30% | 6% | 24% | 80% | 16% | 34% |
| 9 | 20% | 20% | 4% | 22% | 90% | 18% | 36% |
| 10 | 20% | 10% | 2% | 20% | 100% | 20% | 38% |

For the strategy 'avail. pref. incr. 20%', a group is more preferred the higher its percentage preference rank is, whereas for the 'avail. pref. decr. 20%' strategy, the reverse relationship applies, i.e., the lower a group's percentage preference rank, the more it is favored

group is. To implement this business strategy, we first evaluate the average preference weight per group and rank the groups in increasing order of their average preference weights to obtain an indication on how preferred each group is. Next, we determine the percentage preference rank per group by dividing each group's rank by the total number of groups. Hence, the higher the percentage preference rank, the more preferred the group. For the geq constraint, this percentage preference rank per group is then multiplied with $\rho$ to determine the percentage value of products to be at least offered per group. For the leq constraint, we multiply $\rho$ with the sum of the percentage preference rank per group and the difference between maximum and minimum percentage preference rank to determine the percentage value of products to be at most offered per group. Finally, we define $\boldsymbol{b}$ by multiplying the percentage value of products to be at least / at most offered per group with the number of products available for this group. As for the strategy 'avail', we vary $\rho \in \{2.5\%, 5\%, 10\%, 20\%\}$. This strategy is referred to as 'avail. pref. incr. $\rho$' (short for availability and preference increasing with percentage $\rho$).

4. Fourth, we consider the opposite strategy of 'avail. pref. incr. $\rho$', i.e., we relax the requirement of offering at least/ at most $\rho$ percent of the products per group the less preferred a group is. To do so, we proceed as described for the strategy 'avail. pref. incr. $\rho$'. However, in this setting, we rank the groups in decreasing order of their average preference weights, implying that higher percentage preference ranks correspond to less preferred groups and vice versa. Again, we study $\rho \in \{2.5\%, 5\%, 10\%, 20\%\}$. This setting is referred to as 'avail. pref. decr. $\rho$' (short for availability and preference decreasing with percentage $\rho$).

The percentage values of products to be offered per group under the strategies 'avail. $\rho$', 'avail. pref. incr. $\rho$', and 'avail. pref. decr. $\rho$' are exemplarily showcased for $k = 10$ groups and $\rho = 20\%$ in Table 1.

For all four business strategies, we each consider both, the greater-or-equal constraint $A\boldsymbol{x} \geq \boldsymbol{b}$ and the less-or-equal constraint $A\boldsymbol{x} \leq \boldsymbol{b}$.

### Number-of-offered-groups constraint

For the number-of-offered groups constrained problems (MILP-v1 NOG) and (Two-step NOG), we study two different settings of choosing $m$:

1. First, we set $m = 0.8 \cdot k$ for the geq constraint and $m = 0.4 \cdot k$ under the leq constraint for all combinations of $n$ and $k$, implying that products from at least 80% (geq) or at most 40% (leq) of the existing groups should be offered, which is realistic.
2. Second, to analyze the impact of $m$ on both solution approaches, we fix $n = 500$ and $k = 50$ and vary $m$ across all possible values, i.e., between 1 and 50 in steps of one.

For both approaches of defining $m$, we choose $D = n$ and study the greater-or-equal constraint $\sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}} \geq m$ as well as the less-or-equal constraint $\sum_{i=1}^{k} 1_{\{a_i\boldsymbol{x} \geq 1\}} \leq m$.

## Numerical results

We compare the studied products-per-group constrained problem (PPG) and the number-of-offered-groups constrained problem (NOG) to the unconstrained assortment problem (AOP) as well as to two possible extreme assortment policies—offering all products which we refer to as (Offer All), and offering only one product which we refer to as (Offer One). All solution approaches are implemented in a Jupyter Notebook using the programming language Python.

The problem (Offer One) is solved by evaluating the expected revenue for each possible assortment of size one and offering the one that maximizes the expected revenue, whereas the problem (Offer All) is solved by offering all available products. All other problems are solved using the standard Gurobi LP and MILP solvers on a common Fujitsu Lifebook U Series Laptop with an Intel Core i5 7th Gen processor and 8GB RAM.

We consider 100 samples for each problem set ($n$-$k$). Simulated data that only depends on the number of products $n$—i.e., revenues and preference weights—are kept the same for every fixed $n \in \{250, 500, 1000\}$. Thereby, for fixed $n$, we study the same set of products—defined by their revenues and preference weights – while varying the number of groups and the assignment of these products to the groups. We assess the average computation time, the average expected revenue, as well as the average size of the optimal assortment for all considered solution approaches across the 100 samples.

For brevity, the results regarding the computation time are relegated to the Supplementary Material. Overall, we find that all problems besides NOG can be solved efficiently within fractions of a second despite the usage of Jupyter Notebooks. For NOG, instances of small to medium size can be solved efficiently as well; the computation time of larger instance sizes depends on the respective instance whereby seemingly more complicated instances might take some seconds to be solved to optimality though near-optimal solutions are found very fast.

Since we are interested in how much expected revenue is lost by restricting the solution space in order to account, e.g., for business requirements, we do not report the average expected revenues, but provide the average percentage expected revenue losses (APERL) of (Offer All), (Offer One), (PPG), and (NOG) compared to the unconstrained problem (AOP). The APERL values are obtained by i) subtracting the expected revenue of (Offer All), (Offer One), (PPG), or (NOG), respectively, from the expected revenue under the unconstrained problem (AOP) and ii) dividing the resulting difference through the expected revenue under (AOP) for every instance to obtain the percentage expected revenue loss per instance and iii) taking the average across all 100 instances to obtain the APERL.

Moreover, it is intriguing to determine how the average size of the assortment changes compared to the optimal unconstrained assortment by adding a products-per-group constraint or a number-of-offered-groups constraint to the assortment problem. For this purpose, we report the average change in assortment size (ACAS) of (PPG) respectively (NOG) compared to (AOP). The reported values are obtained by i) subtracting the assortment size under (AOP) from the assortment size under (PPG) respectively (NOG) for every instance and ii) taking the average across all 100 instances. That way, a positive value indicates that adding a constraint on average increases the assortment size by this value; negative values imply that the assortment size is on average reduced by this number of products. Obviously, increasing or decreasing the size of the offer set compared to the optimal unconstrained assortment by imposing a constraint results in a loss in expected revenue.

To structurally present our results, we first comment on our insights concerning (Offer All) and (Offer One), followed by analyzing the numerical results for the products-per-group-constrained problem (PPG). Finally, we report our findings for the number-of-offered-groups-constrained problem (NOG).[1] Doing so, some effects follow directly from the theoretical properties of the MNL model and the structure of the constraints. For brevity, we only report the empirical results and do not discuss their theoretical foundation.

### (Offer All) and (Offer One)

Clearly, the average expected revenue is always highest for the unconstrained problem (AOP), since in this case, the solution space is not restricted. In our study, the optimal unconstrained assortment on average consists of 27.88, 43, and 64.95 products for $n \in \{250, 500, 1000\}$, respectively.

Note that the average expected revenue as well as the average size of the optimal assortment under (AOP) increase with the number of products $n$ but do not depend on the number of groups $k$.

Compared to (AOP), the strategy of offering all products, i.e., (Offer All), results in average percentage expected revenue losses of 44.10%, 45.16%, and 46.53% for $n \in \{250, 500, 1000\}$, respectively. These significant losses result from the fact that way more, namely 222.12 ($n = 250$), 457 ($n = 500$), and 935 ($n = 1000$) additional products are offered on average compared to the optimal assortment. Obviously, the APERL and the (average) assortment size increase with $n$ and do not depend on $k$.

Likewise, the strategy of offering only one product, i.e., (Offer One), yields average percentage expected revenue losses of 29.42%, 31.57%, and 33.45% for $n \in \{250, 500, 1000\}$, respectively, compared to the unconstrained problem (AOP). Under this policy, the significant losses result from offering way fewer—namely 26.88 ($n = 250$), 42 ($n = 500$), and 63.95 ($n = 1000$) less—products compared to the optimal assortment. Again, the APERL

[1] Note that we do not compare (PPG) and (NOG) since they can not be reasonably compared because of their different constraint types.

increases with the number of products and does not depend on the number of groups, whereas the (average) assortment size by definition remains equal to one across all problem sets.

Comparing the APERLs of (Offer All) and (Offer One) shows, that (Offer One) performs slightly better for all considered problem sets though both approaches lead to revenue losses between around one third and almost half of the expected revenue compared to the case when the optimal offer set is obtained by solving an unconstrained assortment problem.

#### (PPG)

We evaluate the APERL as well as the ACAS under the assortment problem with a products-per-group-constraint (PPG) compared to the unconstrained problem (AOP) for all business strategies described in section "Products-per-group constraint" under both the geq and the leq constraint. The results under the greater-or-equal constraint are documented in Table 2 and visualized in Fig. 4. The results regarding the less-or-equal constraint are summarized in Table 3 and depicted in Fig. 5. We start by presenting the results under the geq constraint before commenting on the leq constraint.

***Greater-or-equal constraint***

Overall, taking a look at Table 2 shows, that the APERL ranges between 0.00% and 21.37% across all considered business strategies when imposing a greater-or-equal constraint on the number of products to be offered per group. This implies, that adding such a constraint ensures the diversification of the offer set at the cost of at most 21.37% loss in average percentage expected revenue. The revenue loss can be explained by the fact that the constraint enforces a certain number of products to be *at least* offered per group. Thereby, additional products need to be offered compared to the optimal unconstrained assortment. The amount of additional products, i.e., the average change in assortment size depends on the strength of the constraint and the number of instances for which the constraint is hit. This in turn depends on the size of the problem set as well as on the business strategy which defines how the constraint is set. As can be seen in Table 2, the average assortment size increases by between 0 and 184.80 products compared to the size of the optimal unconstrained offer set across all considered business strategies.

According to Table 2 and Fig. 4, the increase in assortment size and thus in APERL is on average lowest for the strategy 'one per group'. Clearly, the requirement to offer at least one product per group is rather easily met—particularly for a low number of groups and a high number of products. However, fulfilling the requirement becomes more difficult the higher the number of groups is, since more products then need to be added to the assortment compared to the optimal offer set. This is also reflected by Table 2 and Fig. 4, which show that the ACAS, and thus also the APERL increase with the number of groups $k$ for fixed $n$. Interestingly, the APERL per one unit increase in assortment size (i.e., adding one product to the assortment) increases with $k$ for fixed $n$, implying that increasing the offer set becomes more costly in terms of expected revenue loss when the number of groups is higher.

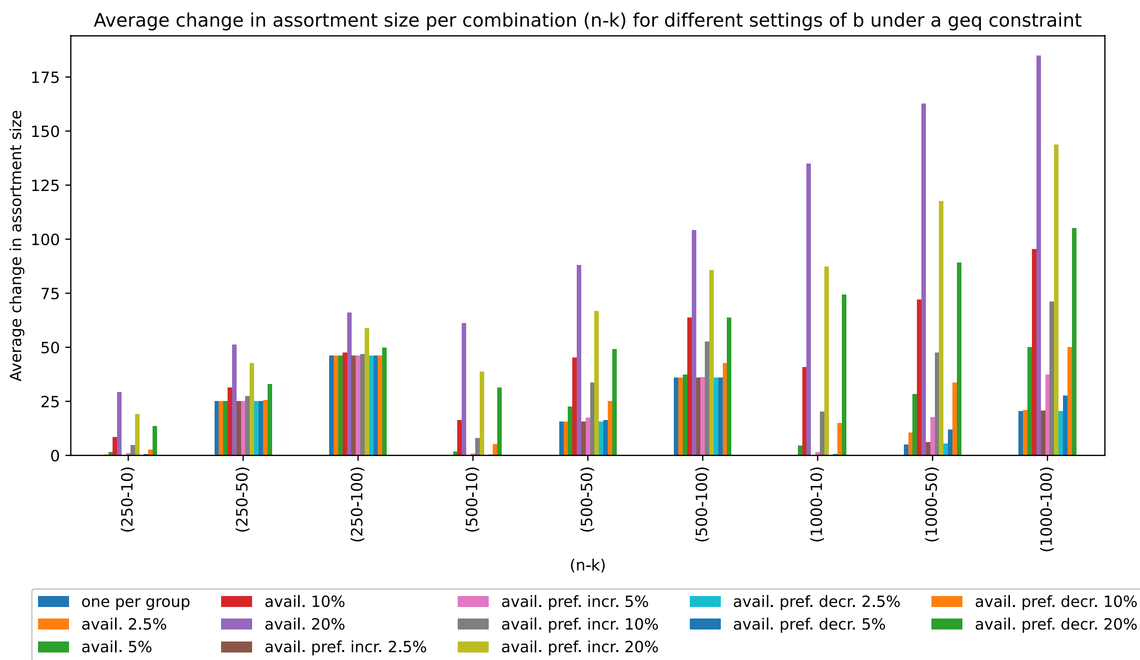
Besides, the fulfillment of the requirement to offer at least one product per group also gets more difficult the lower the number of products is, or put differently, becomes easier for a higher number of products, which is reflected by the fact that both the average change in assortment size and the APERL decrease with the number of products $n$ for fixed $k$ (see Table 2 and Fig. 4). This finding also holds for the APERL per one unit increase in assortment size, i.e., adding one additional product to the assortment becomes less costly in terms of expected revenue loss when the number of products is higher. On average, every one unit increase in assortment size results in an average percentage expected revenue loss of 0.16% when restricting the offer set according to the business strategy 'one per group' compared to the optimal unconstrained assortment.

Regarding the strategies 'avail. $\rho$', 'avail. pref. incr. $\rho$', and 'avail. pref. decr. $\rho$', obviously enlarging $\rho$ means that a higher number of products is required to be at least offered per group, which in turn implies that the change (here: growth) in average assortment size and thus also the APERL raise with $\rho$ (see Table 2). Interestingly, doubling $\rho$ neither results in a doubling of the APERL nor of the ACAS, signifying that the APERL per one unit increase in assortment size grows with $\rho$, although the growth is not linear. This means, that increasing the assortment by one product becomes more costly in terms of expected revenue loss when a more restrictive bound—i.e., a higher value of $\rho$—is imposed on the number of products to be offered per group.

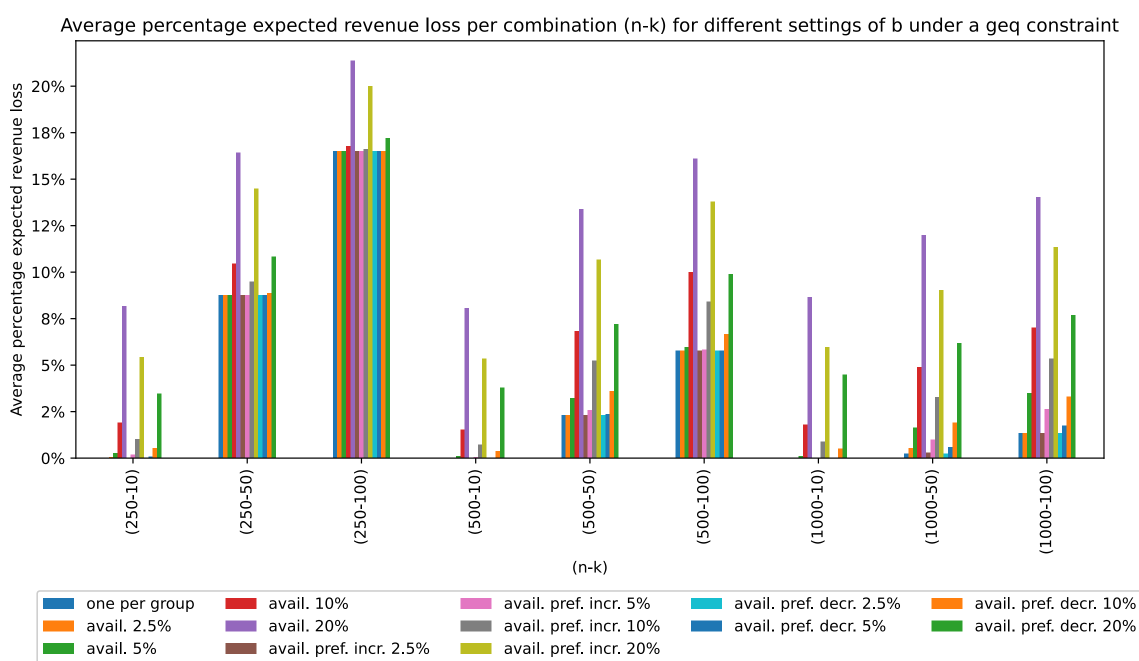
Moreover, increasing the number of groups $k$ for fixed $n$ translates in a higher number of products that need to be added to the assortment to satisfy the constraint, which, according to Table 2 and Fig. 4, results in a larger assortment and higher average percentage expected revenue losses for all of the business strategies 'avail. $\rho$', 'avail. pref. incr. $\rho$', and 'avail. pref. decr. $\rho$'. In line with this, the APERL per additional product added to the assortment increases with the number of groups $k$, implying that adding products to the assortment becomes more costly in terms of expected revenue loss when facing a larger number of groups. Vice versa, as per Table 2 and Fig. 4, a larger number of products $n$ yields a lower APERL for all business strategies except for those with $\rho = 20$ for $k = 10$ in which case no consistent structure is observable. Likewise, the APERL per one unit increase in assortment size

**Table 2** APERL (upper half of the table) and ACAS (lower half of the table) of (PPG) under a geq constraint

| (*n*-*k*) | One per group | Avail. 2.5% | Avail. 5% | Avail. 10% | Avail. 20% | Avail. pref. incr. 2.5% | Avail. pref. incr. 5% | Avail. pref. incr. 10% | Avail. pref. incr. 20% | Avail. pref. decr. 2.5% | Avail. pref. decr. 5% | Avail. pref. decr. 10% | Avail. pref. decr. 20% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (250-10) | 0.04% | 0.04% | 0.28% | 1.91% | 8.18% | 0.04% | 0.20% | 1.01% | 5.42% | 0.04% | 0.07% | 0.54% | 3.47% |
| (250-50) | 8.75% | 8.75% | 8.76% | 10.45% | 16.42% | 8.75% | 8.75% | 9.48% | 14.50% | 8.75% | 8.75% | 8.86% | 10.83% |
| (250-100) | 16.49% | 16.49% | 16.49% | 16.77% | 21.37% | 16.49% | 16.49% | 16.62% | 19.99% | 16.49% | 16.49% | 16.50% | 17.20% |
| (500-10) | 0.00% | 0.01% | 0.11% | 1.52% | 8.07% | 0.00% | 0.04% | 0.72% | 5.34% | 0.00% | 0.02% | 0.38% | 3.78% |
| (500-50) | 2.31% | 2.31% | 3.21% | 6.82% | 13.40% | 2.31% | 2.58% | 5.25% | 10.68% | 2.31% | 2.36% | 3.60% | 7.20% |
| (500-100) | 5.79% | 5.79% | 5.96% | 10.01% | 16.12% | 5.79% | 5.83% | 8.41% | 13.80% | 5.79% | 5.79% | 6.67% | 9.90% |
| (1000-10) | 0.00% | 0.00% | 0.10% | 1.80% | 8.65% | 0.00% | 0.04% | 0.88% | 5.97% | 0.00% | 0.01% | 0.50% | 4.50% |
| (1000-50) | 0.24% | 0.53% | 1.63% | 4.89% | 12.00% | 0.30% | 1.00% | 3.28% | 9.03% | 0.25% | 0.59% | 1.92% | 6.18% |
| (1000-100) | 1.34% | 1.36% | 3.50% | 7.02% | 14.03% | 1.34% | 2.62% | 5.34% | 11.34% | 1.34% | 1.76% | 3.31% | 7.68% |
| (250-10) | 0.24 | 0.31 | 1.64 | 8.51 | 29.24 | 0.25 | 1.07 | 4.67 | 19.16 | 0.24 | 0.56 | 2.79 | 13.46 |
| (250-50) | 25.04 | 25.04 | 25.06 | 31.32 | 51.35 | 25.04 | 25.04 | 27.42 | 42.72 | 25.04 | 25.04 | 25.57 | 32.90 |
| (250-100) | 46.25 | 46.25 | 46.25 | 47.66 | 66.06 | 46.25 | 46.25 | 46.78 | 58.80 | 46.25 | 46.25 | 46.27 | 49.95 |
| (500-10) | 0.01 | 0.11 | 1.79 | 16.22 | 61.13 | 0.03 | 0.78 | 7.94 | 38.81 | 0.05 | 0.38 | 5.21 | 31.31 |
| (500-50) | 15.70 | 15.70 | 22.68 | 45.26 | 87.98 | 15.70 | 17.43 | 33.66 | 66.70 | 15.70 | 16.24 | 25.21 | 49.09 |
| (500-100) | 35.98 | 35.98 | 37.36 | 63.83 | 104.16 | 35.98 | 36.25 | 52.63 | 85.70 | 35.98 | 35.99 | 42.81 | 63.76 |
| (1000-10) | 0.00 | 0.10 | 4.45 | 40.76 | 134.89 | 0.05 | 1.63 | 20.36 | 87.27 | 0.03 | 0.68 | 14.91 | 74.33 |
| (1000-50) | 5.11 | 10.62 | 28.35 | 72.03 | 162.82 | 6.15 | 17.70 | 47.63 | 117.62 | 5.46 | 12.04 | 33.57 | 89.25 |
| (1000-100) | 20.54 | 20.89 | 50.06 | 95.33 | 184.80 | 20.62 | 37.33 | 71.03 | 143.70 | 20.54 | 27.70 | 50.07 | 105.23 |




(a)




(b)

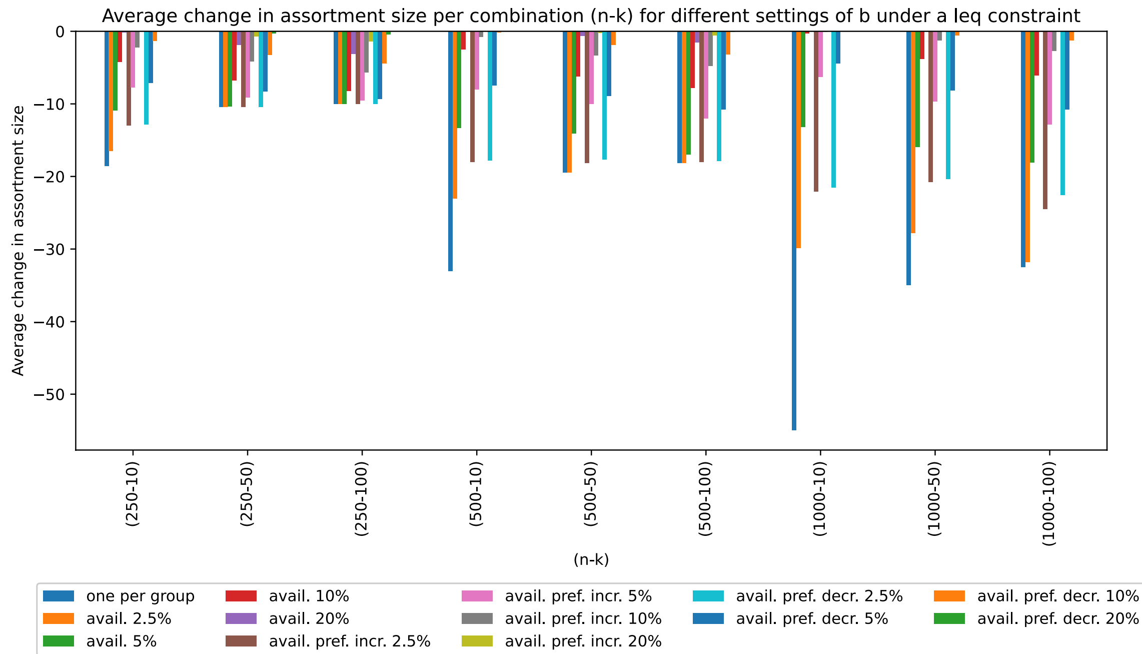
**Fig. 4** **a** Average change in assortment size and **b** average percentage expected revenue loss of (PPG) compared to (AOP) under the greater-or-equal constraint

decreases with $n$, which means that extending the assortment in general becomes less costly when a large amount of products is considered.

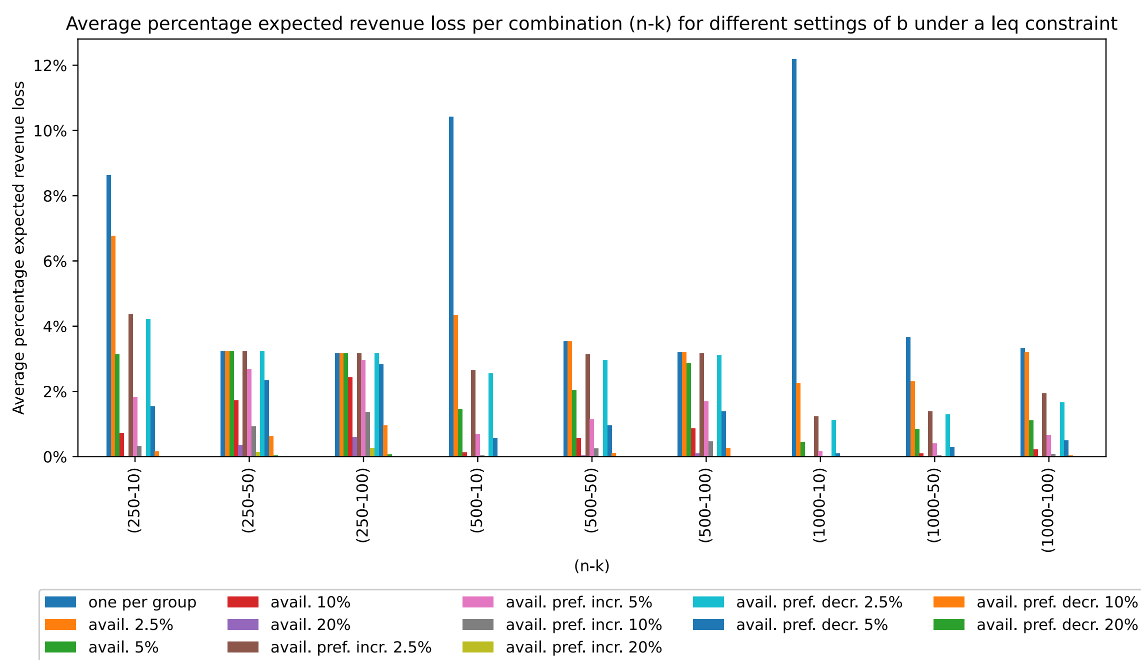
Comparison of the strategies 'avail. pref. incr. $\rho$' and 'avail. pref. decr. $\rho$' with 'avail. $\rho$' shows, that taking preferences into account leads to lower increases in average assortment size and lower APERL values compared to the pure

**Table 3** APERL (upper half of the table) and ACAS (lower half of the table) of (PPG) under a leq constraint

| (*n*-*k*) | One per group | Avail. 2.5% | Avail. 5% | Avail. 10% | Avail. 20% | Avail. pref. incr. 2.5% | Avail. pref. incr. 5% | Avail. pref. incr. 10% | Avail. pref. incr. 20% | Avail. pref. decr. 2.5% | Avail. pref. decr. 5% | Avail. pref. decr. 10% | Avail. pref. decr. 20% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (250-10) | 8.62% | 6.77% | 3.14% | 0.73% | 0.02% | 4.38% | 1.84% | 0.34% | 0.01% | 4.22% | 1.54% | 0.16% | 0.00% |
| (250-50) | 3.25% | 3.25% | 3.24% | 1.72% | 0.35% | 3.25% | 2.69% | 0.93% | 0.14% | 3.25% | 2.34% | 0.64% | 0.04% |
| (250-100) | 3.17% | 3.17% | 3.17% | 2.44% | 0.61% | 3.17% | 2.96% | 1.37% | 0.27% | 3.17% | 2.83% | 0.96% | 0.07% |
| (500-10) | 10.43% | 4.35% | 1.46% | 0.13% | 0.00% | 2.66% | 0.69% | 0.04% | 0.00% | 2.55% | 0.57% | 0.01% | 0.00% |
| (500-50) | 3.54% | 3.54% | 2.05% | 0.58% | 0.03% | 3.13% | 1.14% | 0.26% | 0.01% | 2.98% | 0.96% | 0.11% | 0.00% |
| (500-100) | 3.21% | 3.21% | 2.87% | 0.87% | 0.10% | 3.17% | 1.69% | 0.47% | 0.03% | 3.11% | 1.39% | 0.26% | 0.01% |
| (1000-10) | 12.19% | 2.26% | 0.46% | 0.00% | 0.00% | 1.24% | 0.18% | 0.00% | 0.00% | 1.13% | 0.10% | 0.00% | 0.00% |
| (1000-50) | 3.66% | 2.31% | 0.86% | 0.10% | 0.00% | 1.39% | 0.41% | 0.04% | 0.00% | 1.29% | 0.31% | 0.01% | 0.00% |
| (1000-100) | 3.32% | 3.20% | 1.12% | 0.22% | 0.00% | 1.95% | 0.67% | 0.08% | 0.00% | 1.66% | 0.49% | 0.03% | 0.00% |
| (250-10) | −18.57 | −16.51 | −10.93 | −4.26 | −0.18 | −13.01 | −7.72 | −2.25 | −0.05 | −12.88 | −7.11 | −1.32 | −0.05 |
| (250-50) | −10.41 | −10.41 | −10.39 | −6.77 | −1.86 | −10.41 | −9.15 | −4.16 | −0.72 | −10.41 | −8.31 | −3.27 | −0.28 |
| (250-100) | −9.99 | −9.99 | −9.99 | −8.25 | −3.12 | −9.99 | −9.54 | −5.70 | −1.38 | −9.99 | −9.32 | −4.46 | −0.46 |
| (500-10) | −33.04 | −23.04 | −13.30 | −2.53 | 0.00 | −18.04 | −7.99 | −0.80 | 0.00 | −17.80 | −7.46 | −0.18 | 0.00 |
| (500-50) | −19.50 | −19.50 | −14.10 | −6.26 | −0.62 | −18.19 | −10.01 | −3.32 | −0.24 | −17.68 | −8.95 | −1.87 | −0.03 |
| (500-100) | −18.17 | −18.17 | −16.95 | −7.83 | −1.54 | −17.99 | −12.02 | −4.77 | −0.56 | −17.85 | −10.78 | −3.22 | −0.09 |
| (1000-10) | −54.95 | −29.85 | −13.18 | −0.29 | 0.00 | −22.07 | −6.27 | −0.10 | 0.00 | −21.53 | −4.41 | 0.00 | 0.00 |
| (1000-50) | −34.95 | −27.83 | −15.94 | −3.80 | −0.05 | −20.81 | −9.68 | −1.26 | −0.01 | −20.35 | −8.19 | −0.55 | 0.00 |
| (1000-100) | −32.49 | −31.84 | −18.08 | −6.11 | −0.11 | −24.48 | −12.88 | −2.72 | −0.05 | −22.58 | −10.80 | −1.23 | 0.00 |




(a)




(b)

**Fig. 5** **a** Average change in assortment size and **b** average percentage expected revenue loss of (PPG) compared to (AOP) under the less-or-equal constraint

'avail. $\rho$' strategy across all considered problem sets (see Table 2). Overall, this implies that additionally taking preferences into account by following one of the strategies 'avail. pref. incr. $\rho$' or 'avail. pref. decr. $\rho$' is always beneficial over 'avail. $\rho$', since in comparison, those two strategies guarantee a diversification of the offer set while reducing the average percentage expected revenue loss as well as the increase in average assortment size.

Comparing the strategies 'avail. pref. incr. $\rho$' and 'avail. pref. decr. $\rho$' with each other indicates, that it is more favorable to require that a higher percentage of products is offered from less preferred groups than following the strategy to offer a higher percentage of products from more preferred ones. As can be seen in Table 2 and Fig. 4, putting a stronger constraint on less preferred groups results in a smaller increase in average assortment size as well as in lower APERLs compared to the opposite strategy. Interestingly, the strategy 'avail. pref. decr. $\rho$' also yields a lower APERL per one unit increase in assortment size compared to 'avail. pref. incr. $\rho$' for fixed $\rho \in \{2.5\%, 5\%, 10\%, 20\%\}$. This signifies, that increasing the assortment is less costly when selecting the offer set in accordance with the strategy 'avail. pref. decr. $\rho$'.

In summary, according to Table 2 and Fig. 4, the increase in average assortment size and thus in APERL is on average lowest for the strategy 'one per group', followed by all strategies with $\rho = 2.5\%$, $\rho = 5\%$, $\rho = 10\%$, and $\rho = 20\%$ in this order, whereby for fixed values of $\rho$, 'avail. pref. decr.' always yields lower increases in average assortment size and lower APERLs compared to 'avail. pref. incr.', which in turn always results in lower increases in the size of the offer set and lower APERLs than 'avail.'.

Comparing all APERL and ACAS values with those of the trivial solution approaches (Offer All) and (Offer One) yields, that (PPG) outperforms both trivial policies for all considered business strategies across all analyzed problem sets when setting a greater-or-equal constraint on the number of products to be offered per group. This indicates that requiring to offer at least a certain amount of products per group is not only favorable over the policies to offer all products or only one product in terms of revenue performance, but also ensures a good balance between the diversification and the size of the offer set.

### *Less-or-equal constraint*

When imposing a less-or-equal constraint on the number of products to be offered per group, the APERL of (PPG) varies between 0% and 12.19% across all considered business strategies as can be seen in Table 3. Hence, applying such a constraint ensures that the size of the offer set becomes significantly limited in a meaningful way at a comparably low expected revenue loss of at most 12.19% relative to the unconstrained assortment. The expected revenue loss results from the fact that the constraint enforces that *at most* a certain number of products is offered per group. Thereby, products that would have been contained in the optimal unconstrained assortment need to be removed from the offer set. The number of products that need to be removed, i.e., the average change (here: decrease) in assortment size depends on the strength of the constraint and the number of instances for which the constraint is hit, which in turn depends on the way the constraint is set, i.e., on the size of the problem set and the applied business strategy. Table 3 shows, that the assortment size is on average decreased by between 54.95 and 0 products in comparison to the optimal unconstrained offer set across all considered business strategies.

As exactly opposed to the geq constraint, the strongest decrease in assortment size and thus the highest percentage expected revenue loss is suffered under the strategy 'one per group' (see Table 3 and Fig. 5). Obviously, the request to offer at most one product per group is quite strong—especially when facing a large number of products and a low number of groups since in this case only a very small percentage of products can be offered.

Fulfilling the limitation to offer at most one product per group becomes easier the higher the number of groups is, since fewer products then need to be removed from the optimal offer set. This is also reflected by Table 3 and Fig. 5, which indicate that the deviation from the average optimal unconstrained assortment size and thus the APERL decrease with the number of groups $k$ when keeping $n$ fixed. In line with this, the APERL per one unit decrease in assortment size (i.e., removing one product from the assortment) decreases with $k$, implying that reducing the size of the offer set becomes less costly in terms of expected percentage revenue loss when the number of groups is high.

Apart from that, the fulfillment of the desire to offer at most one product per group also gets easier the smaller the number of products is. This is also reflected by the fact that both the deviation from the size of the optimal offer set as well as the APERL are lower for a smaller number of products (see Table 3 and Fig. 5). However, the APERL per removed product decreases with the number of products, implying that a reduction of the size of the offer set becomes less costly the higher the number of products is.

Applying the strategies 'avail. $\rho$', 'avail. pref. incr. $\rho$', and 'avail. pref. decr. $\rho$' under a less-or-equal-constraint, clearly, increasing $\rho$ means that a higher number of products is allowed to be at most offered per group. Therefore, the average change (here: reduction) of the assortment size and thus also the APERL decrease with $\rho$ (see Table 3). Interestingly, as for the geq constraint, doubling $\rho$ does not lead to APERL or ACAS being halved; the APERL per one unit reduction in assortment size decreases with $\rho$ though the decrease is not linear. This means, that removing a product from the assortment becomes less costly in terms of expected revenue loss when a less restrictive bound—i.e., a higher value of $\rho$—is imposed on the number of products to be at most offered per group.

As opposed to the geq constraint, according to Table 3 and Fig. 5, there is no consistent impact on the change in average assortment size nor on the APERL when increasing the number of groups $k$ while fixing the number of products

$n$. To be precise, for 'avail. $\rho$', 'avail. pref. incr. $\rho$', and 'avail. pref. decr. $\rho$', a larger number of groups mostly comes with a lower number of products that need to be removed from the assortment compared to the optimal offer set except for the strategies with $\rho = 2.5\%$ for $k = 10, 50, 100$. Regarding the APERL, a higher number of groups results in a higher average percentage expected revenue loss for all business strategies besides the strategies with $\rho = 2.5\%$ for $k = 10$, and the strategy 'avail. 2.5%' for $k = 50$. In those exceptional cases, a reduction in APERL is observable. Overall, the APERL per one unit decrease in assortment size mostly increases with $k$, implying that reducing the assortment by one product comes at a higher cost in terms of expected revenue loss the higher the number of groups is.

Vice versa, a higher number of products typically leads to an APERL decrease for all considered business strategies except for the strategy 'avail. 2.5%' for $k = 50, 100$. However, there is no consistent pattern regarding the average change in assortment size observable across the different business strategies (see Table 3 and Fig. 5). The APERL for reducing the size of the offer set by one product decreases with the number of products, meaning that reducing the assortment size becomes less costly when a large amount of products is available.

Comparison of the strategies 'avail. pref. incr. $\rho$' and 'avail. pref. decr. $\rho$' with 'avail. $\rho$' indicates, that taking preferences into account leads to a slightly less severe reduction of the average assortment size and lower values of APERL compared to the pure 'avail. $\rho$' strategy across all considered problem sets (see Table 2). This implies, that additionally taking preferences into account by following one of the strategies 'avail. pref. incr. $\rho$' or 'avail. pref. decr. $\rho$' is always favorable over the strategy 'avail. $\rho$', since those two strategies still allow for a meaningful reduction of the size of the offer set while significantly reducing the average percentage expected revenue loss compared to the strategy 'avail. $\rho$'.

As for the geq constraint, comparing the strategies 'avail. pref. incr. $\rho$' and 'avail. pref. decr. $\rho$' with each other yields, that it is more beneficial to require that a higher percentage of products is allowed to be offered from less preferred groups than following the strategy to allow offering a higher percentage of products from more preferred ones. As per Table 3 and Fig. 5, putting a stronger constraint on less preferred groups results in only mildly less severe reductions in average assortment size but significantly lower APERLs compared to the opposite strategy. Interestingly, again the strategy 'avail. pref. decr. $\rho$' yields a lower APERL per one unit decrease in assortment size compared to 'avail. pref. incr. $\rho$' for fixed $\rho \in \{2.5\%, 5\%, 10\%, 20\%\}$. This means that, just as in case of the geq constraint, decreasing the assortment is less costly when selecting the offer set in accordance with the strategy 'avail. pref. decr. $\rho$'.

In summary, taking a look at Table 3 and Fig. 5 shows, that under a leq constraint, the decrease in average assortment size and thus the APERL is on average lowest for the three strategies 'avail. pref. decr.', 'avail. pref. incr.', and 'avail.' with $\rho = 20\%$, followed by those strategies with $\rho = 10\%, \rho = 5\%$, and $\rho = 2.5\%$, whereas ACAS and APERL are highest under the strategy 'one per group'. Again, for fixed values of $\rho$, the decrease in assortment size as well as the APERL are lowest under the strategy 'avail. pref. decr.', followed by 'avail. pref. incr.', whereas the strategy 'avail.' leads to the strongest reduction in the size of the offer set along with the highest APERL.

Comparison of all APERL and ACAS values with those of the trivial solution approaches (Offer All) and (Offer One) shows, that (PPG) outperforms both trivial policies for all considered business strategies across all analyzed problem sets when imposing a less-or-equal constraint on the number of products to be offered per group. This signifies, that requiring to offer at most a certain amount of products per group not only significantly outperforms the policies to offer all products or only one product in terms of revenue performance, but also ensures a diversification of the assortment at an appropriate size of the offer set.

**(NOG)**

We evaluate the APERL as well as the change in the average size of the optimal offer set under the assortment problem with a number-of-offered-groups-constraint (NOG) for two different settings of choosing the bound $m$ on the number of groups from which products should be offered.[2] The results when fixing $m = 0.8 \cdot k$ under the geq constraint respectively $m = 0.4 \cdot k$ under the leq constraint for all combinations of $n$ and $k$ are documented in Table 4 and visualized in Fig. 6 and Fig. 7. To be precise, in Table 4, the column 'APERL' denotes the average percentage expected revenue loss of (MILP-v1 NOG) and (Two-step NOG), the column 'ACAS' indicates the average change in assortment size compared to the unconstrained optimal offer set, the column '$\varnothing \sum_{i \in K} y_i$' documents the average number of groups from which products are offered, the column '$m$' shows the bound imposed by the number-of-offered-groups constraint, and the column 'constraint active' provides the percentage of active constraints. The results on varying $m$ between 1 and 50 in steps of one while fixing $n = 500$ and $k = 50$ are depicted in Fig. 8.

[2] Note that we impose a time limit of 10 seconds per instance for solving (MILP-v1 NOG) using Gurobi. Extensive tests have shown that even significantly increasing the time limit does not really impact the results, implying that a (near) optimal solution is found extremely fast using Gurobi, though it then takes a while to slightly improve this solution or to 'prove' it to be optimal.

**Table 4** Computational results of (NOG) when choosing $m = 0.8 \cdot k$ under the geq constraint and $m = 0.4 \cdot k$ under the leq constraint

| (n-k) | Greater-or-equal | | | | | Less-or-equal | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | APERL | ACAS | $\varnothing \sum_{i \in K} y_i$ | $m$ | Constraint active | APERL | ACAS | $\varnothing \sum_{i \in K} y_i$ | $m$ | Constraint active |
| (250-10) | 0.00% | 0.00 | 9.86 | 8 | 0% | 4.23% | −11.40 | 4.00 | 4 | 100% |
| (250-50) | 0.89% | 5.25 | 40.02 | 40 | 96% | 1.89% | −7.50 | 20.00 | 20 | 100% |
| (250-100) | 3.42% | 15.38 | 80.00 | 80 | 100% | 1.15% | −6.06 | 40.00 | 40 | 100% |
| (500-10) | 0.00% | 0.00 | 9.99 | 8 | 0% | 3.99% | −19.09 | 4.00 | 4 | 100% |
| (500-50) | 0.01% | 0.42 | 41.82 | 40 | 25% | 2.29% | −15.15 | 20.00 | 20 | 100% |
| (500-100) | 0.24% | 5.18 | 80.17 | 80 | 90% | 1.84% | −14.01 | 40.00 | 40 | 100% |
| (1000-10) | 0.00% | 0.00 | 10.00 | 8 | 0% | 3.32% | −29.56 | 4.00 | 4 | 100% |
| (1000-50) | 0.00% | 0.00 | 46.38 | 40 | 0% | 2.44% | −26.65 | 19.99 | 20 | 100% |
| (1000-100) | 0.00% | 0.08 | 86.12 | 80 | 6% | 2.31% | −27.27 | 39.99 | 40 | 100% |

We start by presenting the results when fixing $m$ to a certain percentage of $k$ before outlining our findings when varying $m$ between 1 and 50. In both cases, we first comment on the greater-or-equal constraint before addressing the results under the less-or-equal constraint.

Note that by construction, (MILP-v1 NOG) and (Two-step NOG) always yield the same APERL and ACAS values, which is why we only report both results once. Moreover, note that it holds in general that in case the number-of-offered-groups constraint is inactive in the optimal solution of an instance, the average assortment size of (NOG) and (AOP) is the same such that the average change in assortment size equals zero since the optimal assortment under (NOG) equals the one under (AOP). Therefore, also the average expected revenue of (NOG) is equal to the optimal average expected revenue of the unconstrained assortment problem (AOP), resulting in an APERL of 0%. In contrast, if the number-of-offered-groups constraint is active, the average expected revenue of (NOG) is lower than the one of (AOP) since the solution space becomes restricted and the optimal offer set becomes larger (under a geq constraint) or smaller (under a leq constraint) compared to the optimal unconstrained assortment. This results in an APERL larger than 0%.

### *Greater-or-equal constraint when setting* $m = 0.8 \cdot k$

Examining the left hand side of Table 4 demonstrates that the APERL of (NOG) under a geq constraint ranges between 0.00% and 3.42% and becomes greater than 0% as soon as at least one instance with an active constraint is considered (see, e.g., problem sets (250-10) vs. (250-100)). Likewise, the size of the optimal assortment is increased by between 0.42 and 15.38 products compared to the optimal unconstrained offer set as soon as at least one constraint is active, and remains unchanged otherwise.

Hence, even imposing a presumably strong greater-or-equal constraint of offering products from at least 80% of the groups only results in mild increases in average assortment size as well as in mild average percentage expected revenue losses. This means, that the diversification of the optimal unconstrained offer set is typically already high such that it is not necessary to artificially enforce assortment diversity by adding a constraint on the number of groups from which products should at least be offered. This finding particularly holds when facing a low number of groups along with a large number of products since in this case, most groups are contained in the optimal assortment anyway such that the constraint is always inactive (see left hand side of Table 4).

The requirement of offering products from at least a certain number of groups becomes more difficult to be satisfied the larger the number of groups is, implying that the percentage of constraints that need to be actively enforced increases with the number of groups (see column 'constraint active' on the left hand side of Table 4). As per column '$\varnothing \sum_{i \in K} y_i$', this is also reflected by the fact that the average number of groups from which products are offered in the optimal assortment under (NOG) gets closer to its bound $m$ when increasing the number of groups $k$ while fixing the number of products $n$. In addition, both the increase in average assortment size as well as the APERL grow with the number of groups $k$ when keeping $n$ fixed in case the constraint is active for at least one instance; otherwise, the APERL and the average change in assortment size equal zero. Likewise, the APERL per one unit increase in assortment size increases with $k$, meaning that adding an additional product to the assortment becomes more costly in terms of expected revenue loss the more groups exist.

Besides, the requirement of offering products from at least a certain number of groups also becomes more difficult to be satisfied the smaller the number of considered products is—or formulated the other way round, becomes easier the larger the number of products is. This is in line with the finding that the percentage of constraints that need to be actively enforced decreases with $n$ (see column 'constraint active' on the left hand side of Table 4). Moreover, taking a look at column '$\varnothing \sum_{i \in K} y_i$' indicates, that the number of groups from which products are offered in the optimal assortment under (NOG) gets further away from its bound $m$ when increasing the number of products $n$ for fixed $k$. Likewise, both the average increase in assortment size as well as the APERL become lower when facing a larger amount of products. In accordance, also the APERL per one unit increase in assortment size decreases with $n$, implying that adding one product to the assortment becomes less costly in terms of loss in expected revenue the higher the number of products is.

Overall, as can be seen in Fig. 6, the APERL of (NOG) strongly relates to the average change in assortment size. That is to say, the stronger the assortment is adjusted by adding additional products compared to the optimal unconstrained assortment, the higher the average percentage expected revenue loss. In general, across the considered problem sets, every one unit increase in assortment size (i.e., adding one product to the assortment) on average results in an average percentage expected revenue loss of 0.09% compared to the optimal unconstrained assortment.

Comparison of all APERL values with those of the trivial solution approaches (Offer All) and (Offer One) shows, that (NOG) significantly outperforms both trivial policies for all considered problem sets when imposing a greater-or-equal constraint on the number of groups from which products are offered. This implies, that setting a constraint on the number of groups from which products should at least be offered is strongly favorable over the polices of offering all products or only one product in terms of expected revenue loss and ensures an appropriate diversification of the offer set while keeping the number of offered products at a reasonable level.

***Less-or-equal constraint when setting $m = 0.4 \cdot k$***

The right hand side of Table 4 shows that the APERL of (NOG) under a leq constraint ranges between 1.15% and 4.23% and is greater than 0% in case at least one instance with an active constraint is considered—i.e., for all studied problem sets. As opposed to the geq constraint, the optimal assortment is decreased by between 6.06 and 29.56 products compared to the optimal unconstrained offer set.

Hence, imposing a less-or-equal constraint of offering products from at most 40% of the groups can significantly reduce the complexity of the assortment by strongly lowering its size, whereby this significant complexity reduction comes at a comparably low cost of at most 4.23% loss in average percentage expected revenue. Even when only demanding that products from at most 40% of the groups are selected, this requirement needs to be actively enforced for all considered instances, which is reflected by the fact that the constraint is always active (see column 'constraint active' on the right hand side of Table 4). This implies, that the optimal unconstrained offer set is typically more complex than desired such that it makes sense to artificially reduce the assortment complexity by adding a leq constraint on the number of groups from which products can be offered.

The requirement of presenting products from at most a certain number of groups becomes easier to fulfill the larger the number of groups is. This is also reflected by the fact that the number of groups from which products are offered in the optimal assortment under (NOG) departs further away from the bound $m$ when increasing the number of groups $k$ while fixing $n$. Likewise, the average change in assortment size and thus the APERL decrease with $k$ when keeping $n$ fixed. In line with this, the APERL per one unit reduction in assortment size also decreases with $k$, implying that removing one product from the assortment becomes less costly in terms of expected revenue loss when facing a larger number of groups.

Besides, offering products from at most a certain number of groups is also easier to be satisfied the smaller the number of products is—or formulated the other way round, becomes more difficult the larger the amount of considered products is. Accordingly, as documented on the right hand side of Table 4, enlarging the number of products $n$ increases the average change in assortment size and thus the APERL for all considered problem sets except for $k = 10$, in which case we observe a decrease in APERL. However, as per column '$\varnothing \sum_{i \in K} y_i$', the number of groups from which products are offered in the optimal assortment under (NOG) departs further from its bound $m$ when increasing $n$ for fixed $k$. Likewise, the APERL per one unit reduction in assortment size decreases with $n$, meaning that removing one product from the assortment becomes less costly the larger the number of considered products is.

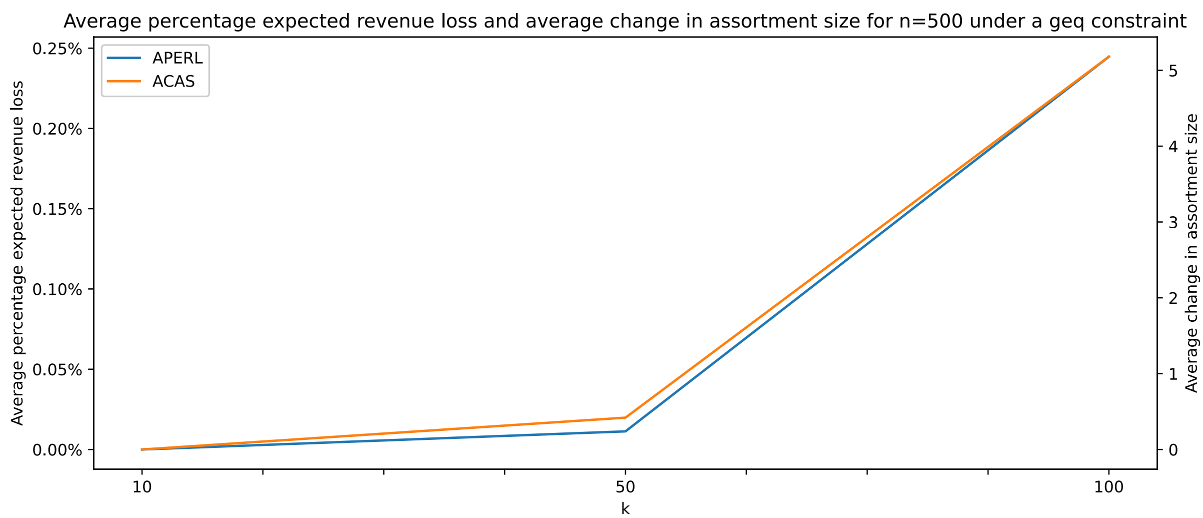
Figure 7 shows, that there is a strong relationship between the APERL and the average change in assortment size.[3] That is to say, the stronger the assortment is adjusted by removing products from the optimal unconstrained assortment, the higher the average percentage expected revenue loss. Overall, across the considered problem sets, every one unit decrease in assortment size (i.e., removing one product from the assortment) on average results in an average percentage expected revenue loss of 0.18% compared to the optimal unconstrained assortment.

Comparing all APERL values with those of the trivial solution approaches (Offer All) and (Offer One) indicates, that (NOG) also significantly outperforms both trivial policies for all considered problem sets when imposing a less-or-equal constraint on the number of groups from which products are offered. Hence, imposing a constraint on the

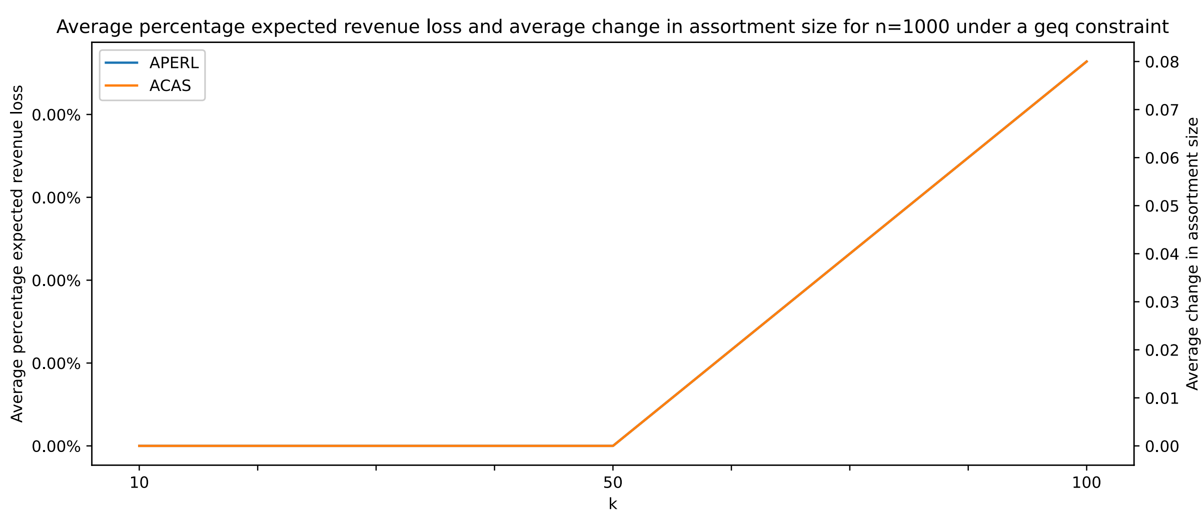
[3] Note that in this figure, the absolute average change in assortment size is visualized to better emphasize this relationship since the actual average change in assortment size is realized as a decrease in the average number of offered products compared to the optimal unconstrained assortment.


Average percentage expected revenue loss and average change in assortment size for n=250 under a geq constraint
APERL
ACAS
Average percentage expected revenue loss
Average change in assortment size
3.50%
3.00%
2.50%
2.00%
1.50%
1.00%
0.50%
0.00%
16
14
12
10
8
6
4
2
0
10
50
100
k

(a)


Average percentage expected revenue loss and average change in assortment size for n=500 under a geq constraint
APERL
ACAS
Average percentage expected revenue loss
Average change in assortment size
0.25%
0.20%
0.15%
0.10%
0.05%
0.00%
5
4
3
2
1
0
10
50
100
k

(b)


Average percentage expected revenue loss and average change in assortment size for n=1000 under a geq constraint
APERL
ACAS
Average percentage expected revenue loss
Average change in assortment size
0.00%
0.00%
0.00%
0.00%
0.00%
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
10
50
100
k

(c)

◂ **Fig. 6** Average percentage expected revenue loss and average change in assortment size of (NOG) compared to (AOP) under the geq constraint for **a** $n = 250$, **b** $n = 500$, and **c** $n = 1000$

number of groups from which products should at most be offered is strongly beneficial over the polices of offering all products or only one product in terms of expected revenue loss while ensuring an appropriate reduction of the assortment complexity despite maintaining a reasonable diversity of the offer set.

***Greater-or-equal constraint when varying*** $m \in [1, 50]$

Taking a look at part (a) of Fig. 8 shows, that the average change in assortment size and thus the APERL of (NOG) under a geq constraint equals 0% despite increasing $m$ as long as no constrained instances are considered. This is the case until $m = 34$, which corresponds to the requirement of offering products from at least 68% of all groups. Clearly, this implies that the optimal unconstrained offer set is already highly diversified.

Further strengthening this diversity requirement by enlarging $m$, the diversification needs to be actively enforced which results in an increasing percentage of active constraints as well as in a growing number of products that need to be added to the assortment to satisfy the requirement, whereby the number of groups from which products are offered in the optimized offer set gets closer to its bound $m$. Hence, the APERL monotonously increases with $m$, since the constraint is hit stronger and for a rising fraction of the considered 100 instances. In line with that, also the APERL per one unit increase in assortment size increases with $m$ (see line APERL/ACAS[4]), implying that adding one additional product to the assortment becomes more costly in terms of expected revenue loss the stronger the requirement regarding the number of groups from which products need to be offered is chosen.

Requiring to offer products from all groups (i.e., setting $m = 50$) only results in adding on average 15.7 products to the optimal unconstrained offer set at the cost of an average percentage expected revenue loss of 2.31%. Note that this is exactly the same as demanding to offer at least one product per group and thus also yields the same ACAS and APERL values as the strategy 'one per group' under a geq constraint (see problem set (500-50) in Table 2). In comparison, the policy of offering all products increases the assortment by 457 products at an APERL of 45.16%. This signifies, that requiring to offer products from every group is far from implying that every product is offered. Under NOG with $m = 50$, on average 58.7 products are contained in the optimal offer set, meaning that mostly one product is offered per group, though there are some groups from which more products are selected. Overall, this indicates that enhancing the diversity of the offer set by imposing a geq constraint on the number of groups from which products need to be offered instead of instead of simply offering all products is way more meaningful in terms of expected revenue loss.

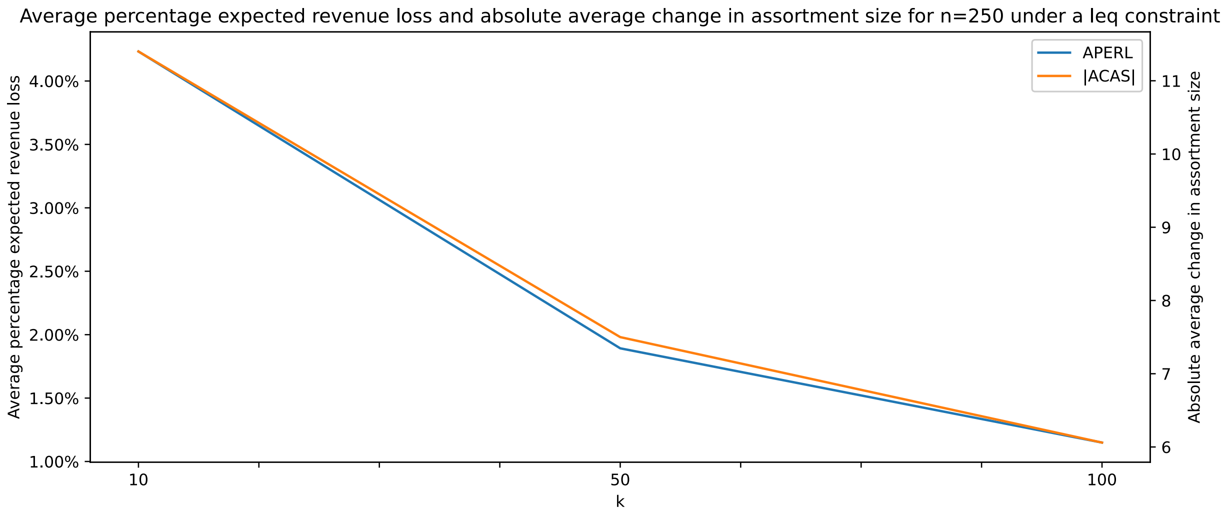
***Less-or-equal constraint when varying*** $m \in [1, 50]$

Under the less-or-equal constraint, the number of groups from which products can at most be offered increases with $m$, implying that the requirement gets easier to fulfill the higher the value of the constraint $m$ is chosen. However, part (b) of Fig. 8 indicates that 100% of the constraints are active until $m = 33$, implying that the requirement of offering products from at most $m$ groups needs to be actively enforced for all instances until $m = 33$, i.e., until demanding to present products from at most 66% of the groups. This in turn indicates, that the optimal unconstrained offer set is highly complex in terms of the number of presented groups, such that it indeed makes sense to reduce the complexity of the assortment by limiting the number of groups from which products can be selected. Further relaxing the constraint, i.e., increasing the value of $m$ beyond 33 results in a decreasing percentage of active constraints until no constraint needs to be actively enforced starting from $m = 48$ (see line constraint active). This implies, that unless demanding to restrict the offer set to contain products from at most 96% of the groups or even more, this requirement needs to be actively enforced.

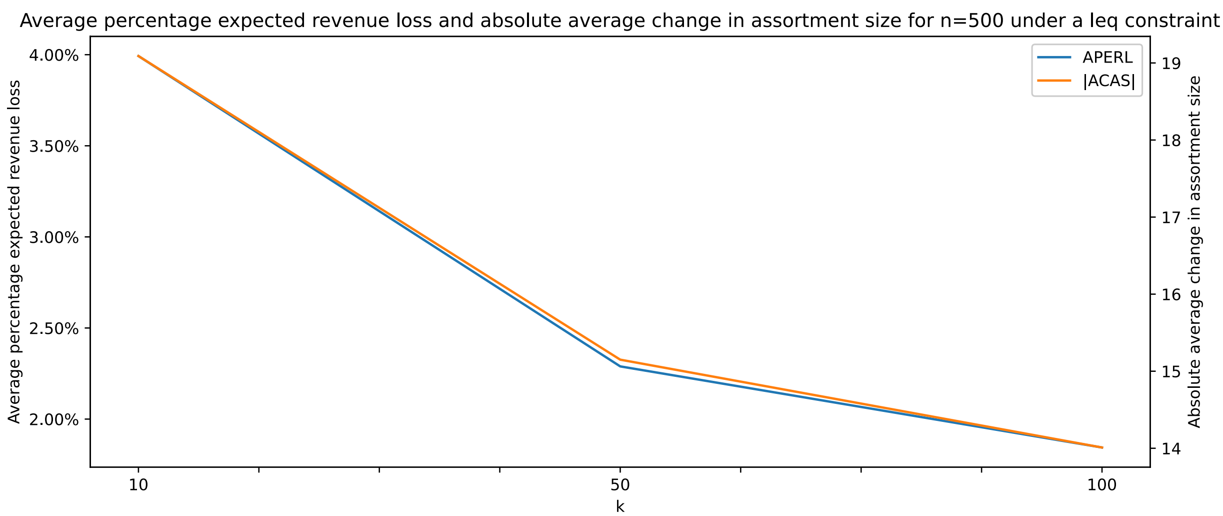
Overall, relaxing the constraint, i.e., allowing for a higher number of groups out of which products can at most be presented implies, that the average number of products that need to be removed from the assortment and thus also the APERL monotonously decrease with $m$ (see lines APERL and |ACAS| in part (b) of Fig. 8). Likewise, also the APERL per one unit reduction in assortment size decreases with $m$ (see line APERL/ACAS), indicating that removing one further product from the assortment becomes less costly in terms of expected revenue loss the looser the requirement regarding the number of groups from which products need to be offered is set. In line with this, the number of groups from which products are offered (line $\varnothing \sum_{i \in K} y_i$) departs further from its bound $m$ the less strong the constraint—i.e. the higher the value of $m$—is set.

Interestingly, requiring to offer products from at most one group—i.e., setting $m = 1$—results in an offer set

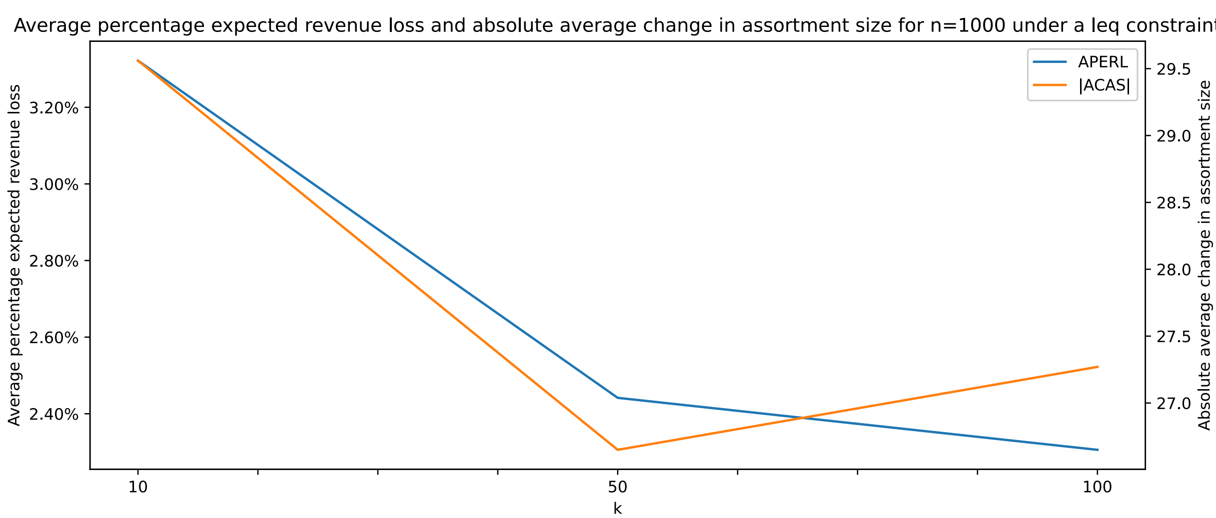
[4] Note that in Fig. 8, the average change in assortment size (line ACAS), the APERL per one unit increase in assortment size (line APERL/ACAS), the average number of groups from which products are offered (line $\varnothing \sum_{i \in K} y_i$), and the percentage of active constraints (line constraint active) are each scaled to values between 0 and 1 and plotted on the right y-axis.


Average percentage expected revenue loss and absolute average change in assortment size for n=250 under a leq constraint
APERL
|ACAS|
Average percentage expected revenue loss
Absolute average change in assortment size
4.00%
3.50%
3.00%
2.50%
2.00%
1.50%
1.00%
11
10
9
8
7
6
10
50
100
k

(a)


Average percentage expected revenue loss and absolute average change in assortment size for n=500 under a leq constraint
APERL
|ACAS|
Average percentage expected revenue loss
Absolute average change in assortment size
4.00%
3.50%
3.00%
2.50%
2.00%
19
18
17
16
15
14
10
50
100
k

(b)


Average percentage expected revenue loss and absolute average change in assortment size for n=1000 under a leq constraint
APERL
|ACAS|
Average percentage expected revenue loss
Absolute average change in assortment size
3.20%
3.00%
2.80%
2.60%
2.40%
29.5
29.0
28.5
28.0
27.5
27.0
10
50
100
k

(c)

◂**Fig. 7** Average percentage expected revenue loss and absolute average change in assortment size of (NOG) compared to (AOP) under the leq constraint for **a** $n = 250$, **b** $n = 500$, and **c** $n = 1000$

comprising products from one group instead of offering nothing, whereby the size of the assortment is reduced by 39.18 products on average compared to the optimal unconstrained offer set. This reduction in assortment size comes at an APERL of 22.06%. In comparison, the policy of offering exactly one product, i.e. (Offer One), reduces the offer set by 42 products at a cost of 31.57% in terms of average expected revenue loss. This shows that under NOG with $m = 1$, more than one product is offered from this single group—3.82 products on average to be precise—which results in a significantly lower APERL compared to (Offer One), implying that reducing the complexity of the offer set by imposing a leq constraint on the number of groups from which products can be offered instead of just limiting the size of the offer

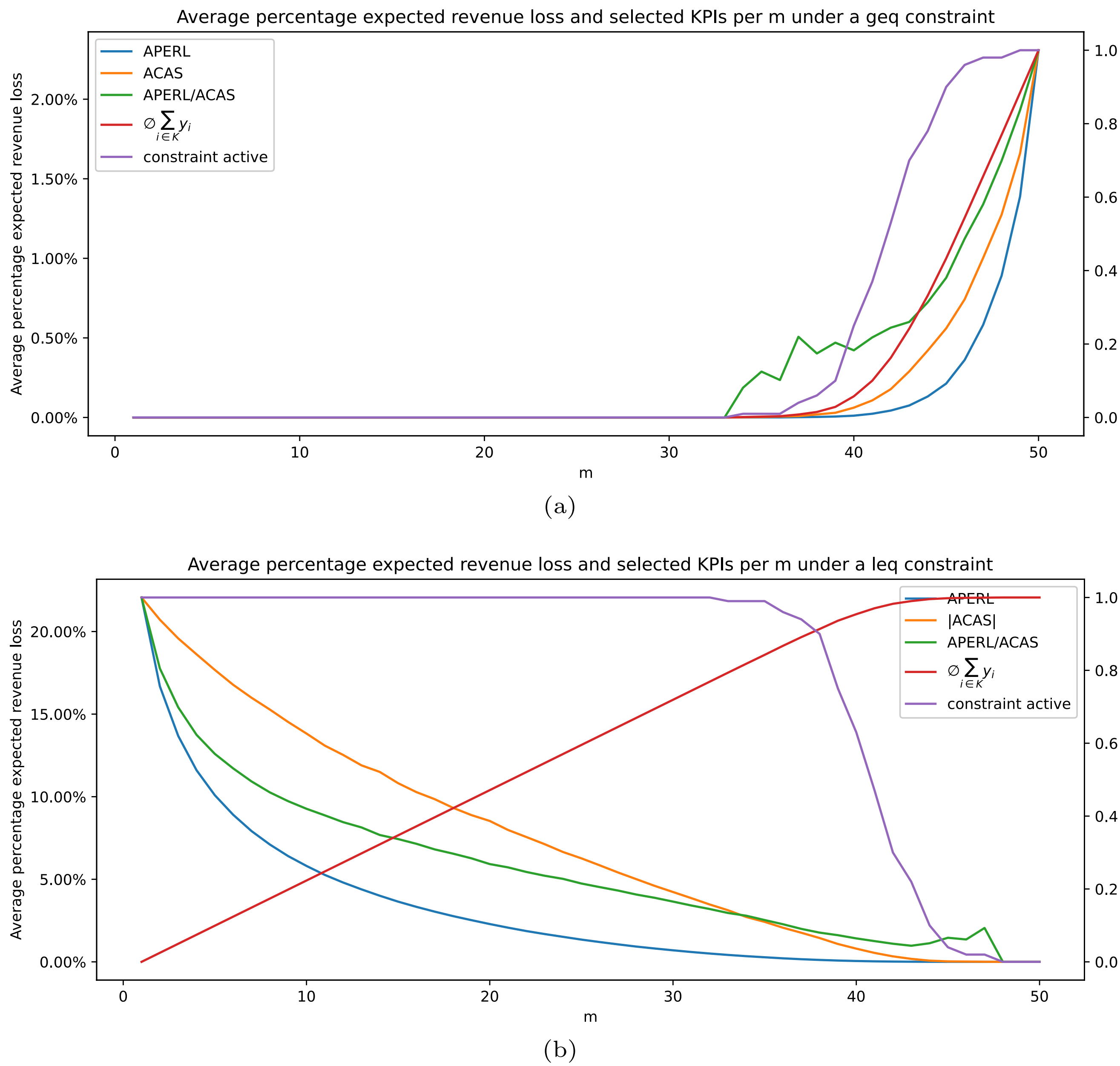



**Fig. 8** Average percentage expected revenue loss (left y-axis), (absolute) average change in assortment size, APERL per one unit change in assortment size, average number of groups from which products are offered, and percentage of active constraints (right y-axis) when varying $m$ between 1 and 50 under **a** the geq and **b** the leq constraint

set to be equal to one is way more meaningful in terms of expected revenue loss.

## Conclusion

Research on assortment optimization received a considerable boost in attention over the past decades. Particularly constrained settings where business requirements are incorporated into the optimization problem become increasingly popular. Our research is motivated by our joint project with BMW with the goal of determining an optimal assortment for the BMW new car locator platform while adhering to multiple business requirements.

To address all BMW requirements, we study the assortment problem under the MNL model while proposing two new types of group constraints. In these settings, the products are assigned to exactly one or to multiple groups and either the number of products per group or the number of groups from which a product is offered is restricted.

The constraint settings are each motivated by selected real-world examples from our project with BMW and formalized as binary fractional linear programs. These BFLPs are then reformulated to MILPs to enable the solvability via common standard solvers such as CPLEX or Gurobi. For the number-of-offered-groups constrained problem, we additionally propose a two-step solution approach.

Finally, we conduct a numerical study using synthetic data of realistic size and structure and find, that all instances of the products-per-group constrained problem can be solved efficiently within fractions of a second despite the usage of Jupyter Notebooks. For the number-of-offered-groups constrained problem, instances of small to medium size can be solved efficiently as well. The computation time of larger instance sizes depends on the respective instance; seemingly more complicated instances might take some seconds to be solved to optimality though near-optimal solutions are found very fast. In this case, if computation time is of interest, one could, e.g., limit the computation time and utilize the near-optimal solution or follow a two-stage approach for larger instances according to which the number of products and/or groups can be reduced in a rule-based way before applying the solution approach to the reduced set of products and/or groups.

Regarding the revenue performance, our numerical study provides evidence that offering all products or offering only one product is highly suboptimal. Moreover, we find that restricting the solution space by adding a products-per-group constraint or a number-of-offered-groups constraint to, e.g., incorporate diversity or assortment complexity requirements into the offer set results in low to moderate average percentage expected revenue losses compared to the unrestricted assortment. The magnitude of the losses depends on the way the bound on the products-per-group constraint and the number-of-offered-groups constraint is set.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1057/s41272-024-00486-6.

**Acknowledgements** The authors thank BMW for the funding of this research project.

**Funding** Open Access funding enabled and organized by Projekt DEAL.



## References

Agrawal, Shipra, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. 2019. MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research* 67 (5): 1453–1485. https://doi.org/10.1287/opre.2018.1832.

Aouad, Ali, Jacob Feldman, and Danny Segev. 2022. The exponomial choice model for assortment optimization: An alternative to the MNL model? *Management Science* 69 (5): 2814–2832. https://doi.org/10.1287/mnsc.2022.4492.

Aouad, Ali, Retsef Levi, and Danny Segev. 2018. Approximation algorithms for dynamic assortment optimization models. *Mathematics of Operations Research* 44 (2): 487–511. https://doi.org/10.1287/moor.2018.0933.

Bai, Yicheng, Jacob Feldman, Danny Segev, Huseyin Topaloglu, and Laura Wagner. 2023. Assortment optimization under the multipurchase multinomial logit choice model. *Operations Research Articles in Advance*. https://doi.org/10.1287/opre.2023.2463.

Bechler, Georg, Claudius Steinhardt, Jochen Mackert, and Robert Klein. 2021. Product line optimization in the presence of preferences for compromise alternatives. *European Journal of Operational Research* 288: 902–917. https://doi.org/10.1016/j.ejor.2020.06.029.

Bernstein, Fernando, A. Gürhan Kök, and Lei Xie. 2015. Dynamic assortment customization with limited inventories. *Manufacturing & Service Operations Management* 17 (4): 538–553. https://doi.org/10.1287/msom.2015.0544.

Bernstein, Fernando, Sajad Modaresi, and Denis Sauré. 2019. A dynamic clustering approach to data-driven assortment personalization. *Management Science* 65 (5): 2095–2115. https://doi.org/10.1287/mnsc.2018.3031.

Caro, Felipe, Victor Martínez-de Albéniz, and Paat Rusmevichientong. 2014. The assortment packing problem: Multiperiod assortment planning for short-lived products. *Management Science* 60 (11): 2701–2721. https://doi.org/10.1287/mnsc.2014.1991.

Charnes, A., and W.W. Cooper. 1962. Programming with linear fractional functionals. *Naval Research Logistics Quarterly* 9 (3–4): 181–186. https://doi.org/10.1002/nav.3800090303.

Chen, Kyle D., and Warren H. Hausman. 2000. Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science* 46 (2): 327–332. https://doi.org/10.1287/mnsc.46.2.327.11931.

Chen, Rui, and Hai Jiang. 2020. Assortment optimization with position effects under the nested logit model. *Naval Research Logistics* 67 (1): 21–33. https://doi.org/10.1002/nav.21879.

Davis, James, Guillermo Gallego, and Huseyin Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Working paper, Cornell University, Ithaca, NY.

Ettl, Markus, Pavithra Harsha, Anna Papush, and Georgia Perakis. 2019. A data-driven approach to personalized bundle pricing and recommendation. *Manufacturing & Service Operations Management* 22 (3): 461–480. https://doi.org/10.1287/msom.2018.0756.

Feldman, Jacob, and Danny Segev. 2022. Technical note—The multinomial logit model with sequential offerings: Algorithmic frameworks for product recommendation displays. *Operations Research* 70 (4): 2162–2184. https://doi.org/10.1287/opre.2021.2218.

Gao, Pin, Yuhang Ma, Ningyuan Chen, Guillermo Gallego, Anran Li, Paat Rusmevichientong, and Huseyin Topaloglu. 2021. Assortment optimization and pricing under the multinomial logit model with impatient customers: Sequential recommendation and selection. *Operations Research* 69 (5): 1509–1532. https://doi.org/10.1287/opre.2021.2127.

Ghuge, Rohan, Joseph Kwon, Viswanath Nagarajan, and Adetee Sharma. 2021. Constrained assortment optimization under the paired combinatorial logit model. *Operations Research* 70 (2): 786–804. https://doi.org/10.1287/opre.2021.2188.

Golrezaei, Negin, Hamid Nazerzadeh, and Paat Rusmevichientong. 2014. Real-time optimization of personalized assortments. *Management Science* 60 (6): 1532–1551. https://doi.org/10.1287/mnsc.2014.1939.

Heger, Julia and Robert Klein. 2024. Assortment optimization: A systematic literature review. *OR Spectrum*. https://doi.org/10.1007/s00291-024-00752-4.

Hense, Jonas, and Alexander Hübner. 2022. Assortment optimization in omni-channel retailing. *European Journal of Operational Research* 301 (1): 124–140. https://doi.org/10.1016/j.ejor.2021.09.045.

Hübner, Alexander, and Kai Schaal. 2017. An integrated assortment and shelf-space optimization model with demand substitution and space-elasticity effects. *European Journal of Operational Research* 261 (1): 302–316. https://doi.org/10.1016/j.ejor.2017.01.039.

Hübner, Alexander, Fabian Schäfer, and Kai Schaal. 2020. Maximizing profit via assortment and shelf-space optimization for two-dimensional shelves. *Production and Operations Management* 29 (3): 547–570. https://doi.org/10.1111/poms.13111.

Jena, Sanjay Dominik, Andrea Lodi, Hugo Palmer, and Claudio Sole. 2020. A partially ranked choice model for large-scale data-driven assortment optimization. *INFORMS Journal on Optimization* 2 (4): 297–319. https://doi.org/10.1287/ijoo.2019.0037.

Katsifou, Argyro, Ralf Seifert, and Jean-Sébastien. Tancrez. 2014. Joint product assortment, inventory and price optimization to attract loyal and non-loyal customers. *Omega* 46: 36–50. https://doi.org/10.1016/j.omega.2014.02.002.

Liu, Nan, Yuhang Ma, and Huseyin Topaloglu. 2020. Assortment optimization under the multinomial logit model with sequential offerings. *INFORMS Journal on Computing* 32 (3): 835–853. https://doi.org/10.1287/ijoc.2019.0910.

Lo, Venus, and Huseyin Topaloglu. 2021. Omnichannel assortment optimization under the multinomial logit model with a features tree. *Manufacturing & Service Operations Management* 24 (2): 1220–1240. https://doi.org/10.1287/msom.2021.1001.

Luce, R.D. 1959. *Individual choice behavior: A theoretical analysis*. New York: John Wiley.

McCormick, Garth P. 1976. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Mathematical Programming* 10: 147–175. https://doi.org/10.1007/BF01580665.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behaviour. In *Frontiers in econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.

Miao, Sentao, and Xiuli Chao. 2022. Online personalized assortment optimization with high-dimensional customer contextual data. *Manufacturing & Service Operations Management* 24 (5): 2741–2760. https://doi.org/10.1287/msom.2022.1128.

Miller, Christopher M., Stephen A. Smith, Shelby H. McIntyre, and Dale D. Achabal. 2010. Optimizing and evaluating retail assortments for infrequently purchased products. *Journal of Retailing* 86 (2): 159–171. https://doi.org/10.1016/j.jretai.2010.02.004.

Qiu, Jiaqing, Xiangyong Li, Yongrui Duan, Mengxi Chen, and Peng Tian. 2020. Dynamic assortment in the presence of brand heterogeneity. *Journal of Retailing and Consumer Services* 56: 102–152. https://doi.org/10.1016/j.jretconser.2020.102152.

Schön, Cornelia. 2010. On the optimal product line selection problem with price discrimination. *Management Science* 56 (5): 896–902. https://doi.org/10.1287/mnsc.1100.1160.

Strauss, Arne K., Robert Klein, and Claudius Steinhardt. 2018. A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research* 271 (2): 375–387. https://doi.org/10.1016/j.ejor.2018.01.011.

Sumida, Mika, Guillermo Gallego, Paat Rusmevichientong, Topaloglu Huseyin, and James Davis. 2020. Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. *Management Science* 67 (5): 2845–2869. https://doi.org/10.1287/mnsc.2020.3657.

Talluri, Kalyan, and Garrett van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50 (1): 15–33. https://doi.org/10.1287/mnsc.1030.0147.

Wang, Ruxian. 2013. Assortment management under the generalized attraction model with a capacity constraint. *Journal of Revenue and Pricing Management* 12: 254–270. https://doi.org/10.1057/rpm.2012.40.



**Julia Heger** studied mathematics at the Technical University of Munich, where she finalized her master's degrees in Mathematics as well as in Mathematical Finance and Actuarial Science in 2021. Since April 2022, Julia Heger is conducting doctoral research on the topic of assortment optimization with the Chair of Analytics & Optimization of the University of Augsburg in cooperation with the car manufacturer BMW. Julia Heger is particularly enthusiastic about solving real-world problems by transferring research results into practice.

**Robert Klein** studied business and information systems engineering at the Technical University of Darmstadt. Since April 2006, Prof. Dr. Robert Klein has held the Chair of Analytics & Optimization at the

University of Augsburg. His research focuses on applying operations research methods to pricing, revenue management, and last-mile logistics. He co-founded and led the pricing and revenue management working group of the German OR Society. Furthermore, he published the first German textbook on revenue management and conducted industry projects in this field with companies like Daimler, Lufthansa, Sixt, and TUI. He published in journals like the European Journal of Operational Research, Journal of Pricing & Revenue Management, and Transportation Science.