Flexible Services and Manufacturing Journal
https://doi.org/10.1007/s10696-023-09492-z

Decision support for managing assortments, shelf space, and replenishment in retail

Alexander Hübner[1] · **Heinrich Kuhn**[2]

Accepted: 3 March 2023

Abstract

Efficiently managing retail space is critical as the increase in product variety is in conflict with limited shelf space and instore replenishment constraints. This paper develops a general framework for retail space management and presents a decision support model with the related problems within the framework of optimizing assortment, shelf space assignment and replenishment. An integrative approach to these planning problems becomes particularly relevant for fast-moving consumer goods and groceries, where stores are regularly replenished from distribution centers. The planning problem at hand is a multi-product shelf space allocation problem where demand is a composite function of the shelf space allocated and assortment-related demand substitution, and actual replenishment practices from retail are incorporated. The model developed extends existing models of shelf space management by jointly considering space-elastic demand and assortment-based substitution and integrating restocking constraints. For the latter, we consider real-world replenishment processes of retailers that distinguish between period-based and ad-hoc replenishment from the backroom. We develop three solution approaches that are based on efficient pre-processing and a nonlinear binary integer programming formulation of the problem. The computation tests based on retail data show the efficiency of the solution approaches in terms of computation time and solution quality. We reveal the improvement in profit levels that can be achieved from integrating assortments, shelf space planning and replenishment where challenges arise in obtaining feasible solutions with limited shelf space and replenishment constraints. We also use sensitivity analyses to demonstrate the high impact of replenishment constraints on profits and solution structures.

Keywords Retailing · Store space management · Space elasticity demand · Substitution demand · Binary integer programming (BIP)

✉ Heinrich Kuhn
heinrich.kuhn@ku.de

Extended author information available on the last page of the article

Published online: 02 April 2023

1 Introduction

The objective and rationale of merchandise retailing is to offer customers a set of products and make them sufficiently available. This is the core of retail space management that provides decision support for defining the assortment (i.e., determining the variety of products), assigning shelf space (i.e., determining the space for each product offered) and determining the replenishment policy (i.e., determining how and how frequently to refill shelves) (Düsterhöft and Hübner 2022; Hübner et al. 2013; Ghoniem and Maddah 2015; Kök et al. 2015). These planning variables are dependent on each other since shelf space in an outlet is limited. Offering a broader assortment with more products limits the available space for each individual product. These factors influence the customer demand achievable and the resulting revenue of an outlet. Another aspect is that unlisted items indirectly impact total demand via substitution from unlisted to listed products. Additionally, the more space is allocated to a product, the higher the visibility for customers and the higher the demand will be. Moreover, the shelf space assigned determines the inventory. On the one hand, increasing the shelf space of a product means the shelves need refilling with it less often and shelf inventory can serve demand until the next regular replenishment period. On the other hand, if shelf space of a product is relatively low, shelf inventory may be insufficient to meet demand until the next replenishment period. This means that retailers need to replenish the shelves with stock from store's backroom, but refilling from backrooms is expensive (see e.g., Sternbeck 2015, 2022; Turgut et al. 2018). Empirical studies show that instore logistics amount to up to 50% of total retail logistics costs, which is also impacted by the shelf space assignment (Broekmeulen et al. 2006; Kotzab et al. 2005; Kuhn and Sternbeck 2013). The replenishment policy is especially relevant for fast-moving consumer goods and groceries, i.e., product categories that have limited shelf space, feature a noticeable shelf rotation rate, and where stores are replenished from distribution centers (DCs) on a frequent and regular basis, at least once a week. This is typically the case in all product categories of supermarkets and drugstores. In these cases an integrative optimization of assortment, shelf space assignment and replenishment becomes particularly relevant (see also Bianchi-Aguiar et al. 2021; Düsterhöft and Hübner 2022; Reiner et al. 2012; Zelst et al. 2009).

Several mutually reinforcing trends observable in retail practice and numerous empirical studies confirm the economic relevance of retail space management. Retailers and consumer goods producers rated "optimization of product portfolio and category management" the most important task for achieving performance goals (Breuer et al. 2009). This is not surprising as shelf space competition in retail stores is at an all-time high, driven by the competitive need to constantly introduce new products. The average number of items in overall store assortments has increased by 30% over the last decade (EHI Retail Institute 2014). However, there is empirical evidence that assortments have become so excessive that reducing variety may increase sales (Iyengar and Lepper 2000; Sloot and Verhoef 2008). Boatwright and Nunes (2001) found that significant item reductions (up to

54%) resulted in an average sales increase by 11% across 42 categories examined, and sales growth in more than two-thirds of these categories. Irrespective of their great relevance for retail practice, current literature largely addresses assortment, shelf space, and replenishment planning independently of each other (Bianchi-Aguiar et al. 2021). Only a few aspects have been integrated into the models, but not yet comprehensively. In particular, reflection on the actual replenishment systems applied in the stores constitute an underexplored area of research, and retail practice will benefit from integration of this issue into the available planning approaches (Düsterhöft and Hübner 2023; Hübner and Kuhn 2012; Kök et al. 2015). The aim of this contribution is to develop an integrated decision support model and corresponding solution approaches that solve practice-relevant problem instances. This paper comprises the following aspects. First, we develop a general modeling framework by considering both assortment, shelf space assignment, and replenishment in one integrated model. Within the model, we also consider an empirically motivated replenishment process that differentiates between period-based regular replenishment and ad-hoc replenishment from the backroom. Furthermore, we consider two types of exogenous demand effects, namely assortment-based substitution effects arising from unlisted items and demand effects resulting from shelf space assignment. Finally, we develop an exact and approximative solution approach as well as a practice-inspired heuristic.

The remainder is organized as follows. After setting the context and reviewing the literature in Sect. 2, the model is developed in Sect. 3. The solution procedure is explained in Sect. 4. Computational tests are presented in Sect. 5. Finally, Sect. 6 discusses the implications of the results on retail practice and provides an outlook on further areas of research.

2 Planning framework and literature review

This section details the conceptual background in retail space management and the decision problem with its main features in terms of scope and demand impacts. It is based on different collaborations with major European grocery retailers (see Düsterhöft et al. 2020; Hübner et al. 2013; Hübner and Schaal 2017a) and a review of related literature (see also Bianchi-Aguiar et al. 2021; Hübner and Kuhn 2012; Kotzab et al. 2005; Kök et al. 2015). We will first develop an overview of related planning questions for retail space management and then examine each planning problem individually. Each subsection also discusses the associated literature that presents related planning models and solution approaches.

2.1 Overview of related planning problems

Retailers are required to make multiple decisions related to tactical management of retail space that above all include which categories to offer at what volume and to what depth, which products to offer, how to allocate each product to the shelf, and how to refill the shelf with products. Retailers often solve the planning questions

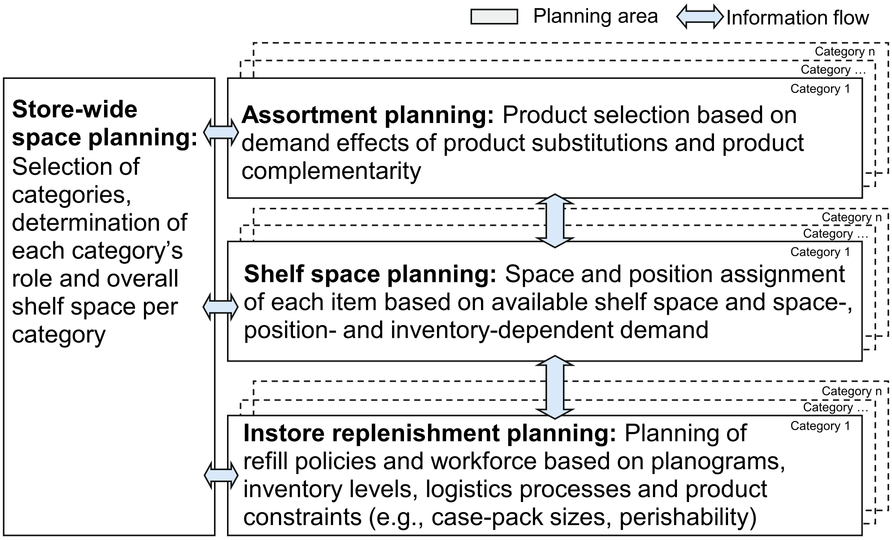

Fig. 1 Planning areas within the framework of retail space management

sequentially. First, they determine store-wide space allocation, followed by the composition of the assortment and allocation to shelves, and finally they manage instore replenishment (see also related frameworks in Hübner et al. (2013); Kök et al. (2015); Flamand et al. (2018) or Bianchi-Aguiar et al. (2021)). However, the planning problems are inevitably interdependent if the shelf space is scarce for the total store and each category. The space of one category can only be modified by adjusting one other category space as well at the very least. Within a category, a broader assortment may on the one hand increase the category demand, but it leads to less space for each product listed on the shelf and therefore increases the frequency of replenishment processes, and vice versa. A reduced assortment on the other hand may lead to substitutions and partial demand compensation. Additionally, fewer replenishments may become necessary with smaller assortments since the number of units stored for each product increases and may then cover the entire demand between two deliveries. This reduces the replenishment activities to fill up the shelves from the backroom when shelf inventories are insufficient to cover demand between two deliveries. For example, Broekmeulen et al. (2006) reveal consequences if shelf space allocation is not aligned with the replenishment regime. About 60% of the items are temporarily understocked in their empirical study, i.e., consumer demand is higher than shelf stock, thus requiring frequent backroom replenishment. Integrative planning approaches can help to overcome such sub-optimal situations. One prerequisite for developing efficient integrative planning approaches is the understanding of any interdependencies in the planning process. It is important to determine which decisions affect which part of which processes. Figure 1 illustrates such an integrated approach. The retail space management framework structures and connects the relevant decision problems of the planning process. An information flow among the planning issues is inevitable. Satisfactory planning output requires coordination among these areas at the very least. In a more comprehensive decision support, the planning steps are determined with sufficient regard to related overarching or subordinated domains.

The B2C bricks-and-mortar retail industry can be divided into food (e.g., grocery, drugstore, pharmacy) and non-food sectors (e.g., fashion, consumer electronics, household products, do-it-yourself, furniture). Assortment and shelf space issues are present and relevant in all these retail branches (Bianchi-Aguiar et al. 2021; Kök et al. 2015). However, integrated consideration with regard to replenishment questions is especially relevant for fast-moving consumer goods and groceries, i.e., product categories that feature a noticeable shelf rotation rate and where stores are replenished from distribution centers (DCs) on a frequent and regular basis, at least once a week. The planning approach we propose is therefore mainly applicable for food (ambient, dairy, chilled, frozen, and fresh products) and drugstore categories. Products in one product group should share a common sales area (shelf, presentation desk) and compete for the limited space on the sales floor. This is typically the case for many product segments of supermarkets and drugstores that are stored on regular shelves.

2.2 Store-wide space planning

Overarching store-wide space planning serves as input to the three subordinated planning problems: assortment, shelf space, and instore replenishment planning. It includes the selection of categories, the definition of each category role, and category shelf space (see e.g. Düsterhöft et al. 2020; Flamand et al. 2018; Irion et al. 2011; Ostermeier et al. 2021). Stores are designed using categories to group products of the same kind, define the sequence of categories within the store, i.e., which kinds of product are found at the entrance, which at the checkout (see e.g., Flamand et al. 2016), and to ensure a certain display size. The display size can be further determined by defining the dimensions of the shelves (e.g., depth and height of each level) (Düsterhöft et al. 2020). The number and size of categories depends on the value proposition, store type, location, and general product variety (Ostermeier et al. 2021).

2.3 Assortment planning

Assortment planning considers the question of which and how many different products to offer within a category (Fisher 2009; Hübner 2017; Kök et al. 2015). The main feature of assortment planning is the integration of a consumer's willingness to accept a substitute when his/her favorite product is not available. This becomes especially relevant if not all conceivable products of a category should or could be listed. It may be beneficial for the retailer to go without some (less profitable) products, thus forcing consumers to switch to substitutes that are more profitable. Assortment-related substitution hence becomes an integrated part of the planning process. For example, Gruen et al. (2002) report that an average of 45% of consumers substitute, i.e., buy one of the available items from that category. In related studies, ECR Europe (2003) concludes that 69% of product demand is substituted. Xin et al. (2009) note that this figure is as high as 75%, and Woensel et al. (2007) mention an even higher figure at 83%. The potential depends on product, situation, and consumer characteristics (Fitzsimons 2000).

2.3.1 Assortment planning models

The most popular approaches for integrating demand substitution in assortment planning are *multinomial logit models* (MNL) and *exogenous demand models* (ED). The MNL is a discrete consumer choice model that assumes that consumers are rational utility maximizers and derives anticipations of consumer behavior (see for example Mahajan and van Ryzin (2001), Gaur and Honhon (2006), Honhon et al. (2010) and Kunnumkal and Martinez-de Albeniz (2019)). In this paper, we focus on ED models since the MNL models usually neglect limited shelf capacity and ED models are mostly used when inventory levels become relevant and shelf space is limited. ED models directly specify the demand for each product. Consumers choose from a set of items. If the preferred item is not available, an individual consumer might accept another item as a substitute according to a defined substitution probability. The ED models generally allow one round of substitution. If the first alternative is also not available, sales are consequently lost. Only a very limited number of models (primary MNL models) allow for several rounds of substitutions (Honhon et al. 2012; Farahat and Lee 2018; Transchel 2017; Transchel et al. 2022). Kök and Fisher (2007) develop an ED model with a shelf space constraint to maximize total profit. They provide evidence of the impact of space limitations. They solve the model with substitution using an iterative heuristic to find the optimal number of facings for each product and a local search to derive the space required for a share of the entire assortment. Hübner et al. (2016) develop an advanced algorithm and show that their algorithm is applicable to large-scale problems. Numerical tests reveal that the heuristic procedure produces close-to-optimal solutions. However, none of these assortment models integrate facing-dependent demand.

2.4 Shelf space planning

The traditional shelf space planning tool for retailers is a planogram, representing an illustration of a shelf space plan of a specific category, showing exactly where each product should be physically displayed at the different shelf levels (*vertical allocation*), how it will be positioned horizontally (*horizontal location*), and how much space that product should have (*space assignment*) (Bianchi-Aguiar et al. 2021). The vertical allocation determines to which shelf level a product is assigned, i.e., the height level within the shelf. The horizontal location determines how products are arranged next to each other and how far a product is positioned from the aisle. Space assignment defines the number of facings, that is, the first row of units on the shelf, to selected products, taking into account the constraints of limited shelf space. If a product gets more facings, it is more likely to be seen by customers and purchased more frequently. This facing-dependent demand is also denoted as "space-dependent demand," i.e., item demand increases if more facings are allocated to that item. The term "space-elasticity" is the measure that expresses the responsiveness of the quantity demanded if the number of facings change (similar to price-elasticity for price changes). Shopper surveys and field experiments conclude that a significant relationship exists between the number of facings and the demand realized. The degree of significance depends on the type of item. Brown and Tucker (1961) recognized increasing space effects ranging from the group of unresponsive, inelastic products to general products for everyday purchases through to impulse purchases. Cox (1964) tests the impact of variations in facings on sales of staples and impulse-purchased items. Frank and Massy (1970) use an experiment to test the influence of facings on sales of grocery products. Curhan (1972) proved that fast-moving products have a higher facing-dependent demand effect than slow-moving items. Drèze et al. (1994) identify the impact on sales through reorganizing shelf configurations. Chandon et al. (2009) reveal that facing variation is the most significant instore factor, even stronger than positioning and pricing. Eisend (2014) conducted a meta-analysis comprising 1,268 estimates of space elasticity and concluded that the average space elasticity amounts to 17%, which implies that unit sales increase by 17% each time the number of facings is doubled. Although results relating to the magnitude of facing-dependent demand impacts differ, all studies conclude that item facings and item sales are positively correlated.

2.4.1 Shelf space planning models

One of the first shelf space planning models that considers facing-dependent demand goes back to the work of Hansen and Heinsbroek (1979), who formulate a non-linear model that considers various constraints, such as minimum and integer shelf quantities, and space elasticities. They apply a Lagrangian relaxation. Corstjens and Doyle (1981) propose a limited shelf space model that considers space and cross-space elasticities and maximizes retail profits while comprehensively taking into account space-elastic revenues. The model is solved by geometrical programming. However, the estimation and optimization procedures cannot be applied to large-scale problems and the model therefore works with product groups rather than SKUs. Zufryden (1986) formulates a deterministic model with space-elastic demand that is solved using dynamic programming. Yang and Chen (1999) assume a linear relation between space and demand within a constrained number of facings. They formulate a shelf space allocation problem with vertical and horizontal space allocation effects. Yang (2001) proposes a knapsack heuristic for the model. He found an optimal solution for simplified versions only. Lim et al. (2004) build on Yang's work by using meta-heuristics for optimization. A later approach of Hwang et al. (2005) provides a demand function that also incorporates neighborhood relationships of items in addition to several shelf segments. Hansen et al. (2010) create a model that integrates detailed location effects within their profit function. Irion et al. (2012) develop a non-linear model for space and cross-space elasticities that is then solved by piecewise linear approximation. They transform the model into an MIP with linear constraints. Their approach provides near-optimal solutions with a posteriori error bound. Gajjar and Adil (2010) build on Irion et al. (2012) and develop a local search heuristic. Bai et al. (2013) provide a model where several shelf segments are available that can each be defined with an individual height. Related to this one, Düsterhöft et al. (2020) and Hübner et al. (2021) provide the first models to address the problem with multiple shelf racks and multiple-sized shelf segments. Bianchi-Aguiar et al. (2018) formulate a model that considers product grouping and display-direction constraints, thus incorporating merchandizing rules. Hübner and Schaal (2017b) integrate assortment planning and model stochastic demand. Geismar et al. (2015) propose a two-dimensional shelf space optimization model that allows displays to extend across multiple shelf levels. The model is solved through a decomposition approach. Hübner et al. (2020) extend the model with a definition of assortment and item arrangement. Akkaş (2019) quantifies the level of shelf space that should be assigned to perishable products considering the fact that the assigned shelf space impacts product expiration. She models the problem as an infinite horizon Markov chain model assuming constant demand across periods and analyses the expiration rate of a product assuming different cycles of shelf rotation.

Some modeling and solution approaches also consider cross-space elasticity effects. However, the discussion of cross-space effects is ambiguous in the literature. Cross-space elasticity quantifies the effects of neighboring and listed items on the demand of another item. Zufryden (1986) argues that considering cross-space elasticity at a product level would be impossible in practice due to the overwhelming number of cross-elasticity terms that would need to be estimated. Eisend (2014) found only five studies over the past 40 years that have been able to identify cross-space elasticity at all. He computes an average cross-space elasticity of −1.6%, which means that sales fall by 1.6% when a neighboring product is presented with facings twice as large as before. Schaal and Hübner (2018) show that the impact of cross-space elasticity on product allocation and retail profit is limited. This also holds true if elasticities are significantly higher than the empirical values obtained so far. We therefore comply with these results and disregard cross-space elasticity within the decision model developed.

2.5 Instore replenishment planning

Instore replenishment planning determines the refilling process of the products on the shelf. It includes instore logistics processes, the determination of refilling quantities and cycles, reorder levels, and safety stocks. Its purpose is to achieve the required on-shelf inventory levels based on given shelf planograms (Donselaar et al. 2010; Zelst et al. 2009). Empirical studies highlight the need to reflect instore constraints and replenishment processes when determining the assortment offered (DeHoratius and Raman 2008; Raman et al. 2001). Zelst et al. (2009) and Kuhn and Sternbeck (2013) show that instore handling costs amount to between 38% and 48% of entire retail logistics costs, often also because shelf space allocation is not aligned with the replenishment regime.

The standard replenishment process happens as follows. Stores receive deliveries of the entire category from the distribution center (DC) on a regular basis. The frequency and exact day of deliveries are defined by so-called cyclic (weekly) delivery

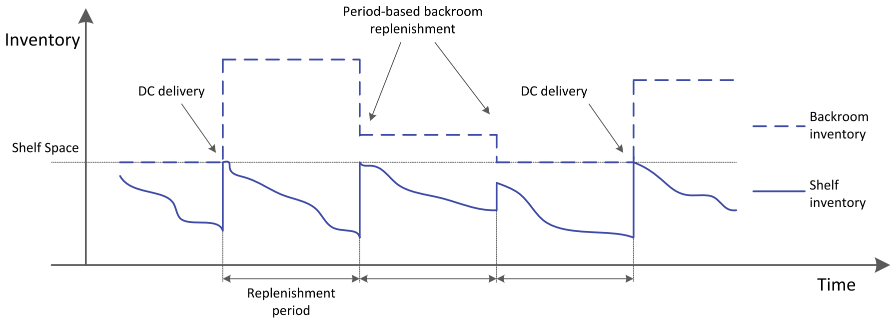

Fig. 2 Period-based replenishment process

patterns (Frank et al. 2021; Holzapfel et al. 2016; Mou et al. 2018; Sternbeck and Kuhn 2014). After a delivery from the DC, products are placed in front of the shelves and shelf-filling operations are then carried out by dedicated shelf refillers, mostly before opening hours (Hübner and Schaal 2017a; Kotzab and Teller 2005). If the shelf space for an item is insufficient for accommodating all the products delivered, they must be carried to the backroom, often on pallets or in-store roll cages, stored in the backroom and restocked later when free shelf space becomes available after consumer purchases (Eroglu et al. 2013; Pires et al. 2020; Reiner et al. 2012; Sternbeck 2015). In this process environment, the process of refilling the shelf from the backroom area has to be organized. Two main policies are applied in practice: (I) "period-based replenishment" and (II) "ad-hoc backroom replenishment."

(I) Period-based replenishment Shelves are replenished within fixed time intervals and order cycles. Generally this refilling process takes place at time instances when a regular store delivery arrives from the DC because then low-cost shelf-stackers can be deployed with processing the delivery (Kuhn and Sternbeck 2013). The backroom inventory and the new orders received are then placed together in front of the shelves of the associated product category and the refilling process starts. The individual products are stacked as far as possible on the shelves. The remaining products are again returned to the backroom and considered once more at the next shelf-stacking event. We denote this process a regular "period-based replenishment" that is scheduled to take place each period. Joint replenishment of backroom inventory and new incoming orders saves walking distances for the refillers and enables greater process efficiency (Hübner and Schaal 2017a; Reiner et al. 2012; Zelst et al. 2009). This refilling is usually carried out outside store hours, e.g., at the end of a sales day or in the morning before the shop opens. A retailer may prefer refilling shelves outside store hours in order to avoid the disruption for customers and regular staff (Berg et al. 1998). This "period-based replenishment" is visualized in Fig. 2 (see also Broekmeulen et al. (2017)).

A special case of "period-based replenishment" exists when all or almost all products of a DC delivery can be stacked on the shelves at each replenishment interval. This allows backroom inventories to be (almost) completely avoided. Retailers refer to this practice as "direct refilling process." This refill concept is becoming

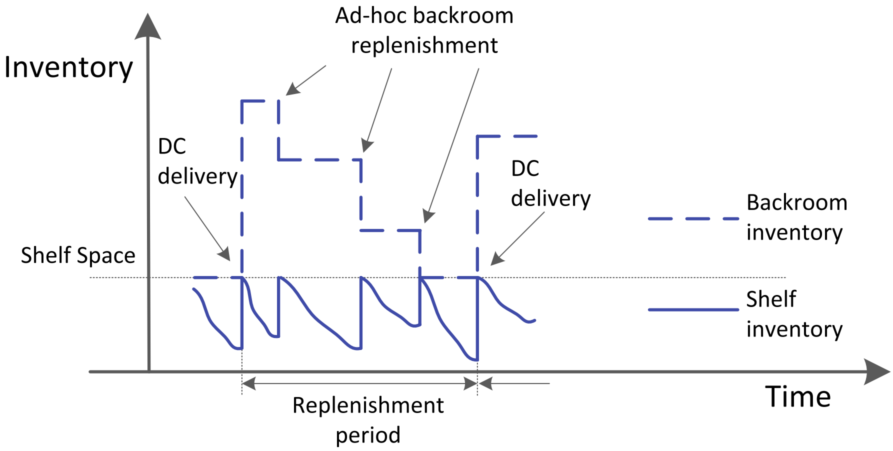

Fig. 3 Ad-hoc backroom replenishment process

increasingly popular in grocery retailing as it avoids inefficient backroom stocking and backroom refilling. However, this concept usually requires a high frequency of store deliveries from the DCs and short order lead times.

(II) Ad-hoc backroom replenishment Ad-hoc backroom replenishment has to take place if the shelf inventory of a product is lower than customer demand between two regular replenishment periods. Refilling from the backroom is then carried out during the opening hours of the store, i.e., at any time before the shelf stock gets depleted (Pires et al. 2015; Kotzab and Teller 2005). This refilling process is denoted "ad-hoc backroom replenishment," and is visualized in Fig. 3. In this case the staff will frequently transfer backroom stock to the shelf during opening hours, avoiding the situation where a product is in the store but not on the shelf. Ad-hoc backroom replenishment incurs a higher cost per unit replenished than either a period-based replenishment process or a direct-refilling process. This is due to several reasons (Sternbeck 2022). First, refilling usually takes place during opening hours, which causes longer processing times, as customers disrupt the process flow or customer purchases have to be considered. Second, refilling is usually performed by store personnel, who have a higher hourly rate than specialized shelf stackers. The latter also work more efficiently than regular staff. Third, the products on the roll cages are far less sorted and organized compared to roll cages and pallets that newly arrive from the DCs or are used in a period-based replenishment process. The roll cages mostly contain single consumer units instead of complete case packs. This makes searching and refilling less efficient. Restocking from the backroom at any time is therefore generally much more cost-intensive than better organized regular period-based replenishment, and should therefore be limited (Kuhn and Sternbeck 2013). "Ad-hoc backroom replenishment" is henceforth denoted "ad-hoc replenishment."

In this paper we assume that the retailer receives replenishment deliveries from the DC at fixed intervals, e.g., every other day. We also assume that the retailer replenishes the shelves between two consecutive delivery periods from the backroom inventory as soon as the shelf inventory is sold. The replenishment process applied therefore avoids temporary stockouts and partial depletion of safety stock on the shelves. This requires that showroom demand can be replaced immediately (ad hoc) with backroom stock (see for example Urban 1998). Here we follow the current literature on shelf space management (see Hübner and Kuhn (2012) and Bianchi-Aguiar et al. (2021)). In retail practice, for example REWE, EDEKA, dm-drogerie markt, two of the largest full-range retailers and one drugstore retailer in Germany, respectively, follow the general goal of achieving the highest possible product availability. REWE, for example, refers to a lack of product availability as "customer pain" and aims to minimize overall customer pain. Retailers therefore limit the number of products eligible for ad-hoc replenishment so that not too many replenishments occur between two regular replenishment periods. Store employees should not be overloaded with refill operations so that "phantom products" are avoided as much as possible, i.e., situations where products are available in backroom storage but not on the shelf and hence are not visible for customers (DeHoratius and Zeynep 2015; Pires et al. 2020). In addition, REWE (for example) marks certain products with black labels on the shelves, signaling to employees that these products should be continuously replenished from the backroom as needed.

2.5.1 Replenishment planning models in the context of shelf space management

Urban (1998) integrates inventory aspects into shelf space management by considering demand as a function of the shelf inventory displayed. The deterministic, continuous-review model considers inventory-elastic demand, since sales before replenishment reduce the number of items displayed. The problem is solved by a greedy heuristic and a genetic algorithm, but violates integer constraints for facings and order quantities. Hariga et al. (2007) propose a deterministic but non-linear model to determine assortment, replenishment, positioning and shelf space allocation considering shelf and storage constraints. The decision variables are the display locations, order quantities, and the number of facings in each display area. It was only possible to exactly solve the problem for a four-item case, although no integer facings values were assumed. Abbott and Palekar (2008) integrate inventory aspects into their shelf space problem to obtain optimal replenishment quantities and frequencies. They formulate an economic order quantity problem and assume a linear relationship between item shelf space and item sales. They determine – exactly for a single-product case and approximately for a multi-product case – the optimal replenishment cycles for products given the cost of restocking and the sales effects of inventory-elastic demand. These stylized models related to shelf space and replenishment planning have limitations in applicability to practical problem sizes as they are mainly developed to derive analytical insights, without focusing on efficient solution approaches for practice-relevant problems. Furthermore, some apply non-integer facings to obtain continuous decision variables. Also all models are restricted to individual product replenishment by an immediate refill by the retailer whenever an item stocks out, i.e., to ad-hoc restocking from backroom inventory. This represents a further limitation as retailers normally replenish products jointly because of joint delivery cycles from central warehouses. Bianchi-Aguiar et al. (2015) overcome this problem by a model that targets minimization of the differences among days' supply of items. However, they focus only on shelf space planning without assortment and space-elasticity effects. A new aspect was demonstrated by Hübner and Schaal (2017a), where replenishment costs for direct replenishment and from the backroom are specified. However, they omit assortment decisions.

2.6 Summary, research gap, and contribution

To summarize, the assortment models investigate substitution effects for unavailable items, but without facing-dependent demand and mostly neglecting limited shelf space. The shelf space models deal mainly with deterministic facing-dependent demand and a restriction in shelf space, but in most cases do not integrate substitution effects for unlisted items. Furthermore, they mostly neglect the effort of replenishing the stock on the shelf. In general it is assumed that a product on the shelf is immediately restocked from backroom inventory as soon as one unit of the product is sold. The shelf space models with replenishment considerations assume individual item refill and most solution approaches are limited to a very narrow item set. Furthermore, these papers simplify the replenishment system that is not differentiated between regular and ad-hoc replenishments as found in practice. We therefore develop a decision support model and solution approach that simultaneously integrates facing-dependent demand and substitution effects and considers the common replenishment processes applied by retailers. Our modeling and solution approach bounds the number of ad-hoc replenishments from backroom inventory for each item of the assortment. This avoids costly and possibly suboptimal replenishment processes since ad-hoc replenishments from the backroom incur much higher replenishment costs than the regular period-based replenishment process (Hübner and Schaal 2017a; Pires et al. 2020; Sternbeck 2022).

3 Model development

This section develops the decision model that is based on the assumptions of the demand model (see Sect. 3.1) and replenishment system (see Sect. 3.2). The profit-maximizing retailer needs to select items from a set of products, $\mathbb{N} = \{1, 2, \dots, i, \dots, |N|\}$. This is expressed by the binary variable $x_i, i \in \mathbb{N}$, which is set to 1 if item i is carried by the store and set to 0 if the item is not listed. Listed items are denoted by set $\mathbb{N}^+$ and unlisted items by $\mathbb{N}^-$. Thus, $\mathbb{N}^+, \mathbb{N}^- \in \mathbb{N}$, $\mathbb{N}^+ \cup \mathbb{N}^- = \mathbb{N}$ and $\mathbb{N}^+ \cap \mathbb{N}^- = \emptyset$. Furthermore, the retailer needs to decide on the integer number of facings k assigned to each item listed so that all items of the selected assortment fit onto the shelf with their respective space requirements. The shelf space for the category is limited to size S that limits the total number of facings of all items listed. Additionally, the limited capacity of the replenishment system has to be considered.

3.1 Modeling the demand function

The total period demand of a listed item i, $i \in \mathbb{N}^+$ is a composite function of (1) the base demand, (2) the facing-dependent demand, and (3) the substitution demand gained from unlisted items.

(1) *Base demand* The base demand rate α_i represents the retailer's forecast for an item that is independent from the facing and assortment decision (see also Hansen and Heinsbroek (1979), Irion et al. (2012) or Bianchi-Aguiar et al. (2015)). The forecast may be based on historical sales, but may also incorporate further demand effects such as shelf location in the store, pricing effects or other marketing efforts. We assume that the time unit of the demand rate corresponds to the length of the replenishment cycle of an item i. Parameter α_i thus equals the demand for the period between two regular replenishments.

(2) *Facing-dependent demand* Common denominators of shelf space models are item demand rates as a function of the number of facings allocated to an item (Bianchi-Aguiar et al. 2021; Hansen and Heinsbroek 1979). In accordance with prior research, the demand rate d_{ik} of the item i at facing level k is a deterministic function of base demand α_i, the number of facings k, with $k = 1, \dots, K$ and the space-elasticity β_i (with $0 \leq \beta_i \leq 1$). We denote this as facing-dependent demand as the demand is driven by the number of facings.

$$d_{ik} = \alpha_i \cdot k^{\beta_i} \qquad \forall\, i \in \mathbb{N},\ k = 1, \dots, K \tag{1}$$

The demand rate at one facing is equal to the base demand since $d_{i1} = \alpha_i \cdot 1^{\beta_i} = \alpha_i$.

(3) *Substitution demand* A shortcoming of space-allocation models using Eq. (1) for the demand computation is the assumption of "zero" demand when assigning "zero" facings for unlisted items (Hübner and Kuhn 2012), i.e., $d_{i0} = \alpha_i \cdot 0^{\beta_i} = 0$. These models therefore do not integrate latent consumer demand for products that are not available but are in the shopper's mind, and where the willingness to substitute is there. In this sense the classic "zero-facing zero-demand" property omits substitution. Estimating substitution demand requires two additional types of parameters: (3a) the potential latent demand for unlisted items $\mathbb{N}^-$, and (3b) the substitution rates of unlisted $\mathbb{N}^-$ to listed items $\mathbb{N}^+$.

(3a) The "potential demand" is denoted by d_{j0}, which is the demand rate at facing level $k = 0$ of an unlisted item $j, j \in \mathbb{N}^-$. This latent demand can be at a maximum as high as the demand at facing level one ($k = 1$) when the product would be listed. Without loss of generality, it is therefore assumed that the demand for unlisted items ranges from 0 to the demand of facing level one, i.e., $d_{j0} = [0; d_{j1}]$. We define parameter λ_j as $0 \leq \lambda_j \leq 1$, which quantifies the remaining demand for an unlisted item as a proportion of the demand at facing level one, d_{j1}.

$$d_{j0} = \lambda_j \cdot \alpha_j \qquad \forall\, j \in \mathbb{N} \tag{2}$$

(3b) The substitution rate of an unlisted item$j$, $j \in \mathbb{N}^-$ to a listed item $i, i \in \mathbb{N}^+$ is denoted μ_{ji}. The sum of the substitution rates of product j to all other products i is denoted $\hat{\mu}_j$ and must be less than or at most equal to 1, i.e., $\hat{\mu}_j = \sum_{i \in \mathbb{N}, i \neq j} \mu_{ji} \leq 1,\ j \in \mathbb{N}$. This intends to convey that every consumer chooses their favorite item j from set $\mathbb{N}$. If their favorite product j is not available for some reason ($j \in \mathbb{N}^-$), the substitution rate μ_{ji} predicts that a consumer will choose their second favorite item i from the set of listed items, $i \in \mathbb{N}^+$. We assume as in Smith and Agrawal (2000) and Kök and Fisher (2007) that one round of substitution in exchange for listed items is allowed ($i \in \mathbb{N}^+$). If consumers want to substitute their first choice by a product that is not listed ($j \in \mathbb{N}^-$), the sales are consequently lost. Kök (2003) shows that this is not too restrictive. The substitution demand $d_i^{\mathbb{N}^-}$ of a listed item $i, i \in \mathbb{N}^+$ gained from the set of unlisted items $\mathbb{N}^-$ can then be quantified by Eq. (3) and depends on the assortment variable x_i for all items $i, i \in \mathbb{N}$:

$$d_i^{\mathbb{N}^-} = \sum_{j \in \mathbb{N}^-, j \neq i} d_{j0} \cdot \mu_{ji} \qquad \forall\, i \in \mathbb{N}^+ \tag{3}$$

Adding up the facing-dependent demand d_{ik} and the substitution demand $d_i^{\mathbb{N}^-}$ quantifies the entire period demand $\hat{d}_{ik}$ of a listed item i:

$$\hat{d}_{ik}(\bar{x}) = d_{ik} + d_i^{\mathbb{N}^-} \qquad \forall\, i \in \mathbb{N}^+,\ k = 0, 1, \ldots, K \tag{4}$$

The total demand function (4) can be further extended if someone is also interested in considering cross-space-elastic (CSE) demand effects as quantified in Eq. (5) (see Schaal and Hübner 2018). The CSE factor between items i and j, δ_{ij} then quantifies the magnitude of the demand change for item i due to a facing change of another item j. Note that items can be linked to one another by complementary or substitution effects. This is reflected by negative or positive values of δ_{ij}, respectively.

$$\hat{d}_{ik}^{\mathrm{CSE}}(\bar{x}, \bar{k}) = \hat{d}_{ik}(\bar{x}) \cdot \prod_{j \in \mathbb{N}^+, j \neq i} (k_j)^{\delta_{ij}} \qquad \forall\, i \in \mathbb{N}^+,\ k = 0, 1, \ldots, K \tag{5}$$

As already noted and discussed in our description of the problem setting (see Sect. 2), we decided to disregard CSE effects within the decision model developed as their influences on the decisions to be made are very limited (see also Eisend 2014 and Schaal and Hübner 2018).

3.2 Modeling the replenishment system

The frequency of period-based replenishments is denoted by the order cycles and the associated patterns of deliveries from the DCs (van Donselaar and Broekmeulen 2013; Turgut et al. 2018). The modeling and solution approaches to determine the ordering cycles, i.e., the frequency of store replenishments, typically assume a given assortment and facing decision (Frank et al. 2021; Holzapfel et al. 2016; Sternbeck and Kuhn 2014). These approaches thus assume that the decision-relevant costs are independent of the assortment and facing decision. In the present context, however, we assume that the replenishment periods were determined in a previous planning step and that the quantity of refill operations from the backroom storage is limited to two replenishment periods. To ensure full shelves, sales staff refill any shelf gaps that arise between regular replenishments. Our direct information from various retailers indicates that shelf monitoring and replenishment by sales staff can be completed between other sales activities (Hübner et al. 2013; Kuhn and Sternbeck 2013) if the total volume of such immediate refills is limited. Sales staff fill shelves ad-hoc with backroom inventory during idle periods when items become depleted. We therefore do not explicitly model any replenishment costs that are difficult to obtain for each refill activity but introduce an operational constraint for each product to limit the refill requirements between two regular replenishments, otherwise sales staff could become overwhelmed with refilling activities (see also DeHoratius and Raman 2008; Hübner and Schaal 2017a; Mou et al. 2018). Defining a minimum share of the total demand that is fulfilled from the available shelf quantity allows operational flexibility at the shop floor level, but also limits intermediate ad-hoc refills. Disregarding this minimum share, as assumed in the current literature, could potentially result in all items having a minimum number of facings and store employees being permanently engaged in refilling shelves. Finally estimating the correct refill costs for each product is complex and demanding. It therefore seems appropriate to model this aspect using product-specific constraints. The effort of replenishing the stock on the shelf depends on two main issues: (1) the maximum stock level of a product on the shelf, and (2) the demand between two regular replenishment periods. Both quantities depend on the number of facings k a product is assigned on the shelf.

(1) The maximum inventory level of product i on the shelf results from multiplication of the number of facings k and the capacity of one facing, g_i. The parameter g_i indicates the integer number of units that can be stored behind a facing, including units that can possibly be stacked on top of each other, e.g., canned goods. This quantity q_{ik} is denoted as facing-dependent stock quantity:

$$q_{ik} = k \cdot g_i \qquad \forall\, i \in \mathbb{N},\ k = 0, 1, \ldots, K \tag{6}$$

The following example illustrates the facing-dependent stock quantity. Assume a product with $k = 3$ facings and four units per facing with $g_i = 4$. The maximum number of units of item i on the shelf then amounts to $q_{i,k=3} = 12$ units (see Fig. 4). Depending on the number of facings chosen, the stock quantity of item i amounts to the following values: $q_{ik} = \{4, 8, 12, \ldots\}$, $k = 1, 2, 3, \ldots$

(2) The demand between two regular replenishment periods is the total period demand $\hat{d}_{ik}$. In the following example, we ignore assortment-based substitution to simplify the illustration. Assuming a base demand of $\alpha_i = 10$ and a space elasticity of $\beta_i = 0.35$, the following facing-dependent demands can be obtained with Eq. (1): $d_{ik} = \{10.00, 13.01, 15.18, \ldots\}$, $k = 1, 2, 3, \ldots$. Figure 5 exemplifies the facing-dependent shelf stock quantity and the facing-dependent demand.

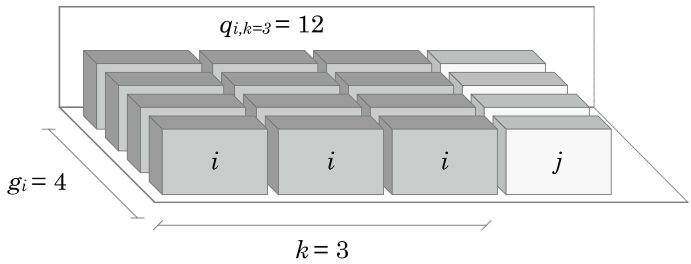

Fig. 4 Illustration of the facing-dependent stock quantity q_{ik}

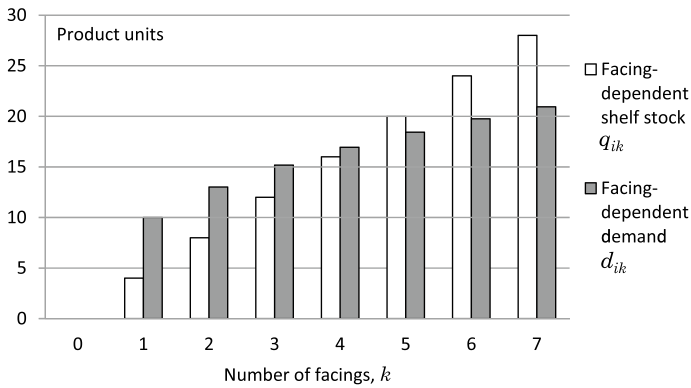

Fig. 5 Facing-dependent demand and shelf inventory, *without substitution*

The total demand within a replenishment cycle noticeably exceeds the maximum available stock on the shelf ($q_{i,k=1} < d_{i,k=1}$) when assigning only one facing ($k = 1$) to item i. The retailer has to additionally refill the shelf from backroom inventory multiple times between two regular replenishments as a result. Assigning five or more facings to item i, however, avoids additional ad-hoc backroom replenishments since the maximum stock level on the shelf covers the entire demand between two regular replenishment periods, $q_{ik} \geq d_{ik}$, $k \geq 5$.

We introduce a minimum share of the demand between two replenishment periods, $0 > f_i^{\min} \leq 1$, which has to be covered by the facing-dependent stock level q_{ik} on the shelf for each item. A minimum share of $f_i^{\min} = 1$ means that the demand during two regular replenishments is entirely satisfied by the available stock quantity. A minimum share less than one, i.e., $f_i^{\min} < 1$, potentially requires refilling processes in between those times. The number of ad-hoc replenishments from backroom inventory can therefore be limited for each item i. This avoids costly and possibly suboptimal refill processes. In the example above we assume a minimum share of 75% of the total demand that has to be fulfilled from the shelf, i.e., $f_i^{\min} = 0.75$. In that case the ratio $\min[\frac{q_{ik}}{d_{ik}}, 1] = \{0.40, 0.61, 0.79, 0.94, 1, 1, 1\}$, $k = 1, 2, 3, 4, 5, 6, 7$, defines the share of demand that can be fulfilled from the shelf with regular period-based replenishment. At least three facings are required in the example, $k \geq 3$, fulfilling the minimum share formulated.

In addition, retailers can also set $f_i^{\min} = 1$, $i \in \mathbb{N}$ for all products available for assortment selection. This completely avoids the need to replenish items from the backroom stock between two consecutive replenishment periods, and models a no-backroom-stocking strategy that also fulfills the "direct refilling process" within a period-based replenishment approach (see Subsect. 2.5). Demand between two regular replenishments is then fully covered by the available stock on shelf. This strategy also avoids ineffective allocation of safety stock in the store, so items may be "in store but not on the shelf" (Corsten and Gruen 2003; DeHoratius and Ton 2009). However, the number of products listed could then be significantly reduced, which would affect the profitability of the retailer. The tradeoff between replenishment from backroom stock and direct shelf replenishment can therefore be analyzed using the modeling approach proposed. Nevertheless, in several circumstances backroom stock cannot be avoided, e.g., in the case of scarce capacity at the DC that forces early transfer of inventory downstream to the stores, allowing large delivery intervals to save transportation costs, and intra-day replenishments for high-frequency product demand (Obermair et al. 2023; Pires et al. 2020; Tompkins 2014).

3.3 Modeling the capacitated assortment, shelf space, and replenishment problem

This section formulates the mathematical model for the **C**apacitated **A**ssortment, **S**helf-space and **R**eplenishment **P**roblem, abbreviated CASRP. The CASRP is based on the demand and replenishment concept developed above. We streamline the model and introduce the binary decision variable y_{ik}, $i \in \mathbb{N}$, $k = 0, 1, 2, \ldots, K$. It accounts for the number of facings k an item i is assigned to. If $y_{i,k=0}$, the item i is not listed, and if $y_{i,k\geq 1}$ then it is listed. The profit function (7) contains three terms. The first term quantifies the gross margin realized from listed items, which depends on the number of facings an item gets assigned to. The decision variable y_{ik} is multiplied for each item i by the related facing-dependent demand d_{ik} and the gross margin earned per unit, p_i. The second term summarizes the additional gain of gross margin realized by substitution demand from unlisted to listed items. If item i is listed (denoted by auxiliary variable $x_i = 1$), it gains substitution volume from the set of unlisted items, $\mathbb{N}^-$. This volume is quantified by $d_i^{\mathbb{N}^-}$. The substitution quantity however depends on the assortment decision for all products ($\bar{x}$), which results in a quadratic problem. The third term quantifies the listing costs arising. It is assumed that fixed costs c_i occur for each listed item i, representing costs for changing shelf layout or additional point-of-sales material.

$$\max \Pi(\bar{y}) \quad = \quad \sum_{i \in \mathbb{N}} \sum_{k=1}^{K} d_{ik} \cdot p_i \cdot y_{ik} + \sum_{i \in \mathbb{N}} d_i^{\mathbb{N}^-} \cdot p_i \cdot x_i - \sum_{i \in \mathbb{N}} c_i \cdot x_i \tag{7}$$

subject to

$$\sum_{i \in \mathbb{N}} \sum_{k=1}^{K} b_i \cdot k \cdot y_{ik} \leq S \tag{8}$$

$$f_i^{\min} \cdot \left[\sum_{k=1}^{K} \hat{d}_{ik} \cdot y_{ik} \right] - \sum_{k=1}^{K} q_{ik} \cdot y_{ik} \leq 0 \qquad i \in \mathbb{N} \tag{9}$$

$$\phi^{\min} \leq \sum_{k=0}^{K} k \cdot y_{ik} \leq \phi^{\max} \qquad i \in \mathbb{N} \tag{10}$$

$$\sum_{k=0}^{K} y_{ik} = 1 \qquad i \in \mathbb{N} \tag{11}$$

$$x_i - \sum_{k=1}^{K} y_{ik} = 0 \qquad i \in \mathbb{N} \tag{12}$$

$$y_{ik} \in \{0, 1\} \qquad i \in \mathbb{N},\ k = 0, 1, \ldots, K \tag{13}$$

$$x_i \in \{0, 1\} \qquad i \in \mathbb{N} \tag{14}$$

The shelf space constraint (8) considers the available space S in the front row on the shelf that can be distributed among the items listed. The shelf space utilized by item i depends on the number of facings assigned to this item and its breadth b_i. Here we only need to consider the shelf space required in the front row as this is the limiting factor. Replenishment constraints (9) ensure that the facing-dependent shelf stock level q_{ik} of each listed item i is large enough to cover the required minimum share f_i^{min} of its total period demand $\hat{d}_{ik}$. The constraints ensure that the ad-hoc replenishment from backroom inventory is limited. Constraints (8) and (9) are only relevant for listed items with $k \geq 1$. Constraints (10) set limits to the lower ($\phi^{\min}$) and upper ($\phi^{\max}$) bound of facings. This depicts business restrictions such as the presentation of certain item types, enforces minimum listings for new products, or sets an upper limit of shelf space assignable for item i. Constraints (11) ensure that only one number of facings is chosen for each item i. The binary decision variable $y_{ik} = 1$ expresses whether item i is placed on the shelf with k facings; $y_{i,k=0} = 1$ symbolizes that item i is not listed and has zero facings. Constraints (12) define the binary auxiliary variable x_i expressing the assortment decision. x_i is set to one if $k \geq 1$, i.e., if item i is placed on the shelf with at least one facing; otherwise, item i is not listed and x_i is set to zero. Constraints (13) and (14) define the binary variables.

A knapsack problem assuming a linear objective function and linear constraints is already known to be NP-hard (Kellerer et al. 2004). Model CASRP is a knapsack problem with a nonlinear and non-separable (quadratic) objective function and nonlinear and non-separable (quadratic) constraints (see Inequations (9)), whose coefficients need to be calculated for every combination of the decision variables. Assuming $b_i = 1, i \in \mathbb{N}$, the number of different solutions Y is yielded by the following binomial coefficient: $Y(N, S) = \binom{N+S-1}{S} = \frac{(N+S-1)!}{S!(N-1)!}$. The problem results in a combinatorial explosion of the number of solutions, e.g., for $N = |\mathbb{N}| = 10$ and $S = 100$ there are approximately $4.2 \cdot 10^{12}$ possible solutions. However, retailers have typical category sizes of 60–80 items on average (EHI Retail Institute 2014). This requires efficient solution approaches that will be presented in the following section.

4 Solution approaches

This section presents an exact (A), an approximate (B), and a heuristic (C) approach to solve model CASRP. All approaches rely on precalculations that are described first.

Precalculations To improve computational performance of the solution approaches, we apply precalculations to restrict the set of possible facings that are possible. These are as follows:

(1) The maximum number of integer facings, denoted by K, has to be defined. This bounds the number of parameters d_{ik} and decision variables y_{ik} required. The set K can be chosen such that realistic shelf configurations will result. Typically, retailers do not assign more than 10-15 facings to an individual item. Furthermore, the retailer may additionally fix an item-specific lower and upper limit of the number facings, $\phi^{\min}$ and $\phi^{\max}$ (see Constraints (10)).
(2) In addition to (1), the number of facings that are valid can be further reduced. We restrict these to facing levels that at least fulfill the Replenishment Constraints (9). We exclude such item-facing combinations that definitely do not fulfill the minimum facing-dependent stock level on the shelf, i.e., $q_{ik}/d_{ik} \leq f_i^{\min}$. This partitioning of the set of facings is also valid assuming substitution since it will increase the total demand of a listed item i.

Pre-processing of (1) and (2) reduces the set of all facings $\mathbb{K}$ to a set of possible facings $\overline{\mathbb{K}}_i$ for each item i, with $\overline{\mathbb{K}}_i \subset \mathbb{K}$. This also makes Constraints (10) redundant. As the number of facings k can only have integer values and is restricted to a set of facings, it is now possible to precalculate the facing-dependent demand d_{ik} for each item $i, i \in \mathbb{N}$ and each possible facing $k, k \in \overline{\mathbb{K}}_i$ using Eq. (1).

(A) Exact solution of model CASRP The pre-processed data described above are fed into a standard nonlinear BIP solver (in our case CPLEX) such that the CASRP model is solved as described in Sect. 3, see objective function (7) and Constraints (8) to (14). This approach is denoted as $\text{CASRP}_{\text{exact}}$. Note that this formulation results in a quadratic problem. However, we will show in the numerical examples that optimal solutions can be generated by a solver within reasonable time boundaries and an MIP-gap $\leq 1.0\%$. Here we apply the restricted set of facings $\overline{\mathbb{K}}_i$ obtained from precalculations, which reduces computation time significantly.

Step 1: Generate an initial solution without substitution

Set iteration index $\ell = 0$

Set $\mu_{ji} = 0 \quad \forall\ i, j \in \mathbb{N}$

Solve CASRP with Eq. (7) to (14)

Step 2: Update profit with substitution and generate sorted lists of the assortment

Set iteration index $\ell = \ell + 1$

Set $\mathbb{N}^{+(\ell)} = \left\{1, 2, \ldots, i, \ldots, |N^{+(\ell)}| \,\middle|\, x_i = 1, i \in \mathbb{N}\right\}$

Set $\mathbb{N}^{-(\ell)} = \left\{1, 2, \ldots, i, \ldots, |N^{-(\ell)}| \,\middle|\, x_i = 0, i \in \mathbb{N}\right\}$

Update total demand with $\hat{d}_{ik}^{(\ell)}\ \forall\ i \in \mathbb{N}^{+}$

Calculate profit with $\pi_{ik}^{(\ell)} = \hat{d}_{ik}^{(\ell)} \cdot p_i - c_i\ \forall\ i \in \mathbb{N}^{+}$

Set $\mathbb{N}_{[\text{asc}]}^{+(\ell)} = \left\{[1], [2], \ldots, [i], \ldots, [N^{+(\ell)}] \,\middle|\, \pi_{[i],k}^{(\ell)} \leq \pi_{[i+1],k}^{(\ell)}\right\}$

Step 3: Perform a feasibility check and adapt the number of facings

Step 3.1 For $[i] \in \mathbb{N}^{+}$:

If Eq. (9) is false, then set $k = k + 1$, else go to Step 4

end for

Step 3.2 If Eq. (8) is false, then start with first entry $a \in \mathbb{N}_{[\text{asc}]}^{+(\ell)}$ and set $k = k - 1$ and continue with next entries $a + 1, a + 2, \ldots$ until Eq. (8) is true, end if

Step 3.3 For $[i] \in \mathbb{N}^{+}$:

If $[i] = [a]$, then set $\hat{d}_{[i],k}^{(\ell)} = q_{[i],k} / f_{[i]}^{min}$, else go to Step 2

end for

Step 4: Calculate profit achieved

Return π_{ik} and calculate total profit according to Eq. (7)

Fig. 6 CASRP$_{\text{approx}}$ procedure

(B) Approximate solution of model CASRP Model CASRP is approximately solved by a BIP solver by initially neglecting the demand substitution and completing an ex-post update of demand and ensuring fulfillment of all constraints. This approach is denoted as CASRP$_{\text{approx}}$. It degenerates the original nonlinear problem into a bounded 0/1 multi-choice knapsack problem given a set of item-facing combinations, and each combination ($y_{ik} \in \{0, 1\}$) is associated with size $w_{ik} = b_i \cdot k$, facing-dependent profit π_{ik}, $i \in \mathbb{N}$ and a knapsack with capacity S.

The pseudo-code of the algorithm is displayed in Figure 6. In Step 1 an initial solution is obtained by disregarding substitutions ($d_i^{\mathbb{N}^-} = 0$) for all items and solving the CASRP with a BIP solver. A facing level of zero may result for some items. Hence we obtain a set with listed items $\mathbb{N}^+$ and unlisted items $\mathbb{N}^-$. In Step 2, the total demand and profit is updated with substitution demand. This means quantifying the substitution demand $d_i^{\mathbb{N}^-}$ for all items listed using Eq. (3) with the assortment obtained in the previous step. The additional demand is added to the facing-dependent demand and the new profit per item π_{ik} for all listed items is calculated.

Replenishment Constraints (9) are only considered without substitution demand at the initial solution of Step 1; adding the additional demand may result in non-feasible solutions since the new total demand (i.e., with substitution demand ex-post added) may be larger than the facing-dependent stock level and the minimum level f_i^{min}. To overcome these potential violations, we proceed as follows. We rank all items by their profit contribution π_{ik}. Items are then sorted in ascending ($\mathbb{N}^+_{[asc]}$) order. In Step 3.1, we first identify the items with a violation of Constraints (9) and then start increasing the number of facings of these items by one unit with the goal of obtaining feasible demand-stock ratios and fulfilling Constraints (9) again. To compensate for the additional space requirement, we decrease the number of facings by Step 3.2 for the items with low profitability from the ordered set $\mathbb{N}^+_{[asc]}$ until the Shelf Space Constraint (8) is also fulfilled again. We do this because usually these low profit items have a lower impact on total profit, lower total demand and usually a higher facing-dependent shelf stock than demand ($q_{ik} \geq \hat{d}_{ik}$). However, there may be situations where the number of facings for the same item goes up in Step 3.1 and then goes down again in Step 3.2. In these cases we cap total demand to the maximum shelf inventory available ($\hat{d}_{ik} = q_{ik}/f_i^{min}$) with Step 3.3. This is done as the Shelf Space Constraint is always binding and it overrules facing increases of Step 3.1. This also means that it may not be possible to fulfill the entire demand. These steps may also result in items being removed from the current assortment. Thus an additional iteration (Step 2) is required to update the item sets, demands and profits. We complete the iterations until all items fulfill the Replenishment Constraints (9) and total shelf space is met with Constraint (8).

(C) Heuristic solution of model CASRP In retail practice the limited shelf space is generally allocated proportionally to the market share of all items considered (see, e.g., Bianchi-Aguiar et al. 2021; Düsterhöft and Hübner 2022; Griswold 2007; Hübner and Kuhn 2012; Kök et al. 2015). High-volume items are therefore preferentially assigned to the assortment and achieve more facings than low-volume items (Hübner and Kuhn 2012). We base the heuristic on this intuitive approach found in practice as well. In a nutshell, the intuitive approach in retail practice is a ranking method where items and facings are determined based on their proportional sales contribution. To obtain feasible solutions, adjustments of facings are then required until space constraints are fulfilled. We further advance this basic approach by incorporating the idea of additional demand information in the heuristic by calculating the total demand with space-elastic effects and substitutions. The procedure is denoted as CASRP$_{\text{heu}}$. In Step 1 of procedure CASRP$_{\text{heu}}$ the number of facings per item i is approximated based on the relative profit r_i and is calculated as follows:

$$r_i = \frac{\alpha_i \cdot p_i}{\sum_{i \in \mathbb{N}} \alpha_i \cdot p_i} \;\forall\, i \in \mathbb{N} \tag{15}$$

These are then rounded to the nearest integer values to obtain the number of facings with:

$$k_i = \left\lfloor \frac{S \cdot r_i}{b_i} + 0.5 \right\rfloor \;\forall\, i \in \mathbb{N} \tag{16}$$

Note that a facing level of zero may result for some items. These items are excluded from the assortment in this part of the procedure. Afterwards we execute the same Steps 2 to 4 of CASRP$_{\text{approx}}$ to incorporate substitution demand as well as to maintain the space limit and replenishment constraints. During these steps, delisted items may again be added to the assortment.

5 Numerical analyses

This section analyzes the effectiveness of the solution approaches and develops managerial insights on some modeling aspects. Subsection 5.1 characterizes the test instances. Various numerical studies are then performed. Subsection 5.2 shows the effectiveness of the approaches suggested. In Subsect. 5.3 we analyze the impact of the demand parameters and the consequences of hard and soft replenishment constraints.

5.1 Test cases and data generation process

We apply several test cases proving the general applicability and capacity of the model and solution approaches suggested. The generation of test cases is based on a given data set from a real retail outlet. All items in the data set being considered belong to one category. Additional parameters not given in the real data set, i.e., potential demand of unlisted items and substitution rates, are generated based on empirical studies from literature.

Category problem sizes The numerical study should verify whether the solution approaches suggested can solve problem sizes faced by common retailers. The problem size depends on two parameters: the number of items, $|\mathbb{N}|$, and the number of facing levels, K. Problem cases are therefore chosen that include realistic values for those parameters. A typical hypermarket carries approximately 35,000 to 50,000 items within 600 categories, i.e., the average number of distinct items in a category is around 60 to 80 (EHI Retail Institute 2014). Direct information from the retailer suggests that it is rare for any category to exceed 250 items. Our test case from the retailer consists of multiple subcategories with a total of 300 items. In addition, retailers do not assign more than 10-15 facings to an individual item. This also holds true for our application. The number of facing levels is therefore limited to $K \leq 20$ for all items considered.

Specifying model parameters Demand data (i), financial data (ii), and shelf space data (iii) are quantified from the retail data set as follows. (i) The data set contains the annual sales volume for each item $i \in \mathbb{N}$. We use a proprietary data set of a grocery retailer and apply it to the beverage category. In this outlet, all beverages are stored on shelves and each product may have multiple facings. We use different subcategories to obtain different sizes of data sets. The sales volume across all items has an annual average of 9,868 units and a standard deviation of 9,534 units. The high standard deviation results since the data set contains both high- and low-volume items. The annual sales of each item are then divided by the number of replenishment periods per year, resulting in the sales between two replenishment periods. Sales and demand are assumed to be equal. This assumption is in line with the current literature (DeHoratius and Raman 2008). In cases where the out-of-stock rate shows relevant values, the daily demand distribution can be approximated by the demand values of days that have a positive stock level at the end of each day (Chuang et al. 2016; DeHoratius et al. 2023).

Given the number of facings realized, k, the period demand d_{ik} of each item i at facing level k is known. The base demand of each item i, α_i, is then calculated using the inverse of Eq. (1): $\alpha_i = d_{ik}/(k)^{\beta_i}$, $i \in \mathbb{N}$. The space elasticity is set to $\beta_i = 0.2$, $i \in \mathbb{N}$. This assumption is based on empirical studies found in literature (see Chandon et al. 2009; Drèze et al. 1994; Eisend 2014). An exogenous substitution estimate is applied representing the share substituted for the first favorite at an aggregated consumer level. The substitution intensity μ_{ij} vis-à-vis the first alternative item is 0.5, versus 0.2 for the second and 0.1 for the third. The products are ranked according to product attributes and first, second and third alternatives are assigned accordingly. The share of lost sales is consequently 0.2. This assumes that the base demand of an unlisted item j, d_{j0}, is added to the respective facing-dependent demand of the three alternative items if these items are listed. Thus, in sum, a maximum of $80\% = 50\% + 20\% + 10\%$ of the base demand could be transferred to the three possible alternative items. However, the associated demand shares are lost if the corresponding items are also not listed. Thus, we assume only one substitution round in each of the three possible substitution cases. Various empirical studies support these assumptions (e.g., Boatwright and Nunes 2001; Gruen et al. 2002; Xin et al. 2009). Furthermore, the proportion of demand at one facing that is still available at zero facings is assumed to be $\lambda_j = 0.8$, $j \in \mathbb{N}$.

(ii) Unit gross margins provided are uniformly distributed between $0.85 \leq p_i \leq 2.50$ currency units, $i \in \mathbb{N}$. Annual listing costs are set at $c_i = 1{,}000$ currency units for all items. (iii) The item breadth b_i is obtained from the given retail data set. Analyzing the correlation between profit and space results at $R^2 = 0.8055$. Since knapsack problems are generally hard to solve by a high profit-to-space correlation (see Pisinger 2005), the data applied will not only reflect retail practice but also lead to problems that are hard to solve. In each instance the available shelf space, S, is calculated based on the number of facings observed k and the breadth b_i of all items. The maximum stock volume per facing g_i is also obtained from the data set. We also apply bounds on the number of facings. The lower bound ($\phi^{\min}$) and upper bound ($\phi^{\max}$) are set at 25% and 400% of the observed number of facings in the given data set. This is aligned with the current practice of the retailer. Parameter $f_i^{\min}$, $i \in \mathbb{N}$ is set at 0.8. This ensures that the maximum stock level q_{ik} on the shelf covers at least 80% of the total demand between two replenishment periods. As a result, a maximum of 20% of the total demand needs to be refilled during the regular replenishment periods by ad-hoc shelf refilling processes from the backroom. Solution approaches are implemented in Java and CPLEX as the BIP solver is applied. All tests were run on an Intel Core i7-8550U CPU 1.80 GHz processor with 16 GB RAM.

Table 1 Profit impact and runtime of different solution approaches with varying items and facings, with $\beta_i = 0.2$, $\lambda_i = 0.8$, $\hat{\mu}_i = 0.8$ and $f_i^{min} = 0.8$, $i \in \mathbb{N}$

Case	$\|\mathbb{N}\|$	K	$\Delta\Pi_{approx}$ [%]	$\Delta\Pi_{heu}$ [%]	Runtime [sec.]		
					$CASRP_{exact}$	$CASRP_{approx}$	$CASRP_{heu}$
1	10	10	1.4	6.2	<1	<1	<1
2	30	10	0.7	5.5	<1	<1	<1
3	50	10	0.5	7.0	8	1	1
4	100	10	0.4	7.8	240	3	2
5	200	10	0.5	8.4	602	5	3
6	250	10	0.3	8.3	652	9	4
7	300	10	0.3	8.9	866	10	4
8	80	5	0.4	7.4	185	2	2
9	80	10	0.4	7.5	265	8	2
10	80	20	0.6	8.4	268	12	3
11	80	25	0.5	7.9	298	12	3
12	80	30	0.5	8.2	295	12	3
13	200	20	0.4	8.2	613	6	4
14	250	20	0.5	8.3	706	15	4
15	300	20	0.6	8.7	951	21	4
16	300	30	0.8	7.8	1,016	22	4

5.2 Effectiveness of modeling and solution approaches

We evaluate the efficiency of each solution approach by assessing the impact on the solution structure by identifying the changes in the number of facings, and by assessing the impact on the objective values by comparing total profitability. For the latter, we evaluate the profit $\Pi_{[...]}$ for each approach and then quantify the relative impact compared to the exact approach:

$$\Delta\Pi_{approx} = \frac{\Pi_{exact} - \Pi_{approx}}{\Pi_{exact}} \quad \text{and} \quad \Delta\Pi_{heu} = \frac{\Pi_{exact} - \Pi_{heu}}{\Pi_{exact}} \tag{17}$$

Table 1 summarizes the results for varying problem sizes. First, it can be seen that the intuitive application of a space assignment based on the relative profit contribution (expressed with the $CASRP_{heu}$) leads to about 7.5% lower profitability compared to the exact approach. The heuristics have the advantage of being fast and intuitive, but at the cost of solution quality. The direct integration of substitution effects (as applied in $CASRP_{exact}$) increases the profit by up to 1.4% compared to the approximative approach with ex-post integration of substitutions (with $CASRP_{approx}$). The profit disadvantage remains very limited for small and large problem sets. However, the approximative approach has clear advantages with respect to the computational effort. Even large-scale realistic problem sizes can be solved within seconds. The solution approach with precalculation of the facing-dependent demand and efficiently restricting the number of facings makes it possible to calculate solutions

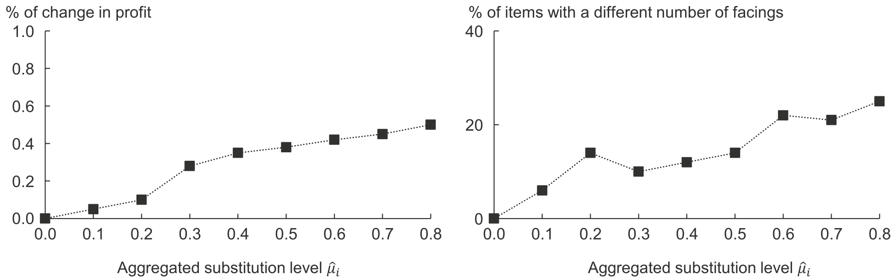

Fig. 7 Impact of substitution rates on solution quality and structure, $\mathrm{CASRP}_{\mathrm{exact}}$ vs. $\mathrm{CASRP}_{\mathrm{approx}}$, example with $|\mathbb{N}| = 50$, $f_i^{\min} = 0.8$, $\lambda_i = 0.8$ and $\beta_i = 0.2$, $i \in \mathbb{N}$

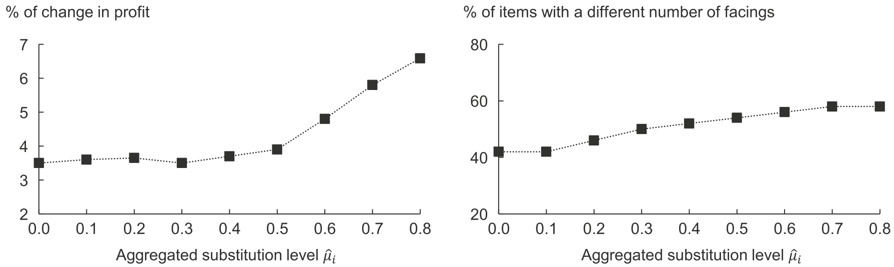

Fig. 8 Impact of substitution rates on solution quality and structure, $\mathrm{CASRP}_{\mathrm{exact}}$ vs. $\mathrm{CASRP}_{\mathrm{heu}}$, example with $|\mathbb{N}| = 50$, $f_i^{\min} = 0.8$, $\lambda_i = 0.8$ and $\beta_i = 0.2$, $i \in \mathbb{N}$

time-efficiently for the problem being considered. All solution approaches can handle all relevant problem sizes. Retailers usually optimize shelves for each category individually. The average category size is about 60-80 items, and generally there are not more than 15 facings per item. The large cases with 200 and more items represent the most complex problems. Looking at the runtime we can confirm that the computation times for all test problems with $\mathrm{CASRP}_{\mathrm{exact}}$ and $\mathrm{CASRP}_{\mathrm{approx}}$ are still within reasonable boundaries for the mid-term planning problem considered.

Exact versus approximative solution We additionally analyze the impact of substitution rates on solution quality and structure between the exact ($\mathrm{CASRP}_{\mathrm{exact}}$) and the approximate ($\mathrm{CASRP}_{\mathrm{approx}}$) solution approach for the case $|\mathbb{N}| = 50$. Figure 7 shows the results. The magnitude of substitution rates has an impact on the efficiency of $\mathrm{CASRP}_{\mathrm{approx}}$. As the approximative approach only integrates substitutions in a post-optimization run, it is obvious that the difference vs. the exact approach grows with increasing substitution levels. However, the difference is still at a low level given that the substitution rates will usually be between 0.5 and 0.7. The change in the solution structures is about 15-20% and also depends on the substitution rate. This has a limited impact as the change of one facing always implies the change of another facing. In the case of 50 items, it means that changes to 5 items most likely result in total changes to 10 items, which is a ratio of 20%.

Table 2 Impact of substitution and replenishment level, case $|\mathbb{N}| = 50$ and $\beta_i = 0.2, i \in \mathbb{N}$

Substitution rate $\widehat{\mu}_i$	0%			50%			80%		
Min. replenish. $f_i^{\min}$	0%	50%	100%	0%	50%	100%	0%	50%	100%
Profit potential[1]	0.0%	0.4%	0.5%	12.7%	12.4%	12.6%	13.8%	12.9%	12.4%
Assortment size[2]	85%	78%	72%	64%	56%	48%	58%	54%	45%
Facing changes[3]	0%	18%	22%	74%	82%	86%	75%	78%	80%

[1] Calculation: ($\text{CASRP}_{\text{exact}}$ profit / CASRP* profit)-1

[2] Optimized assortment size as a share of total possible assortment

[3] Share of items with different facings when comparing solutions of $\text{CASRP}_{\text{exact}}$ and CASRP*

Exact versus heuristic solution A further study compares the exact solution of the CASRP model with the results of the $\text{CASRP}_{\text{heu}}$ procedure, which represents common retail practice (see Sect. 4, Subsection (C) as well as Hübner and Kuhn 2012). We consider a test case with 50 items and up to 10 facings taken from the data set described. The left-hand side of Figure 8 shows the changes in total profit deriving from increasing substitution intensity. $\text{CASRP}_{\text{exact}}$ significantly outperforms $\text{CASRP}_{\text{heu}}$ with profits that are up to 7% higher in the test cases considered. The substitution works here as a further penalty for non-optimal decisions. The exact solution not only results in higher profits but also significantly changes the structure of the assortment and the number of assigned facings. The exact solution allocates significantly different facing levels to the items. The right-hand side of Figure 8 shows that up to two-thirds of the items receive non-optimally facing levels if $\text{CASRP}_{\text{heu}}$ is used.

5.3 Sensitivity analyses

In this subsection we investigate the value of integration and how the results are affected by the replenishment constraints and the various demand parameters considered.

Value of integration First we analyze the value of integrating substitution demands and replenishment constraints together in a single modeling and solution approach. This illustrates the relevance of the model extensions proposed. To do so we define the following benchmark approach that we denote as CASRP*. In the first step we solve the $\text{CASRP}_{\text{exact}}$ with $\widehat{\mu}_i = 0$ for all $i \in \mathbb{N}$, i.e., we ignore the substitution effects in the optimization approach. We subsequently consider the additional substitution demands and adjust (if necessary) the number of facings so that the replenishment and shelf space constraints are satisfied. This post-optimization process proceeds similarly to step 3 of $\text{CASRP}_{\text{exact}}$. The step is relevant to achieve comparable results to the approach $\text{CASRP}_{\text{exact}}$ in which substitution demands and replenishment constraints are directly integrated. Please note that CASRP* equals $\text{CASRP}_{\text{exact}}$ when $\widehat{\mu}_i = 0$ and $f_i^{\min} = 0$ for all $i \in \mathbb{N}$.

Table 2 reveals the profit potential and impact on solution structure of the integrated approach $\text{CASRP}_{\text{exact}}$. Depending on the magnitude of substitution rates and

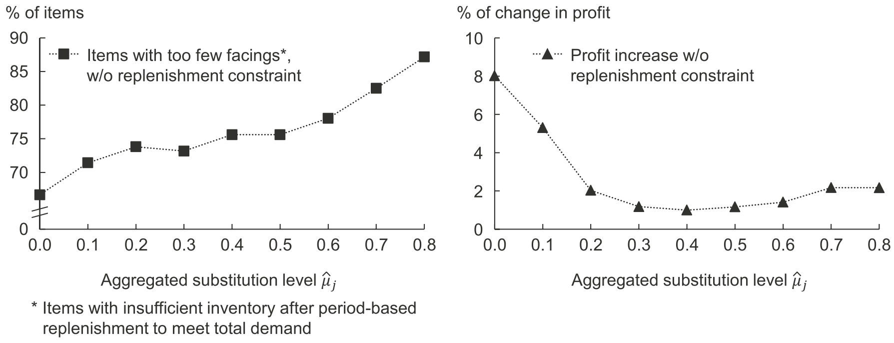

Fig. 9 Impact of replenishment constraints on solution structures (left) and total profits (right), case $|\mathbb{N}| = 50$, $f_i^{\min} = 0.8$, $\lambda_i = 0.8$ and $\beta_i = 0.2$, $i \in \mathbb{N}$

replenishment constraints, the retailer can achieve a profit increase of up to 14% when substitution is directly integrated. A significant reduction in the assortment is recommended across all scenarios independent of the magnitude of substitution rates and replenishment levels. The unlisted items all have low demand and/or a small (absolute) profit margin. Increasing willingness of customers to substitute (expressed by $\hat{\mu}_i$, $i \in \mathbb{N}$) and increasing minimum replenishment levels (expressed by $f_i^{\min}$, $i \in \mathbb{N}$) usually result in smaller assortments. Higher substitution rates make it possible to no longer list less profitable products as the respective demand is transferred to listed products with possibly higher profitability. The higher the minimum replenishment level, the more products need higher inventories on the shelf, which results in a higher number of facings and in unlisting further products. This is also represented in the steadily increasing facing changes with increasing replenishment levels f_i^{min}.

Impact of replenishment constraints The following analysis investigates the dependencies and impact of the replenishment constraints. The replenishment constraints ensure that part of the total demand can be fulfilled with the stock on the shelf available after a regular period-based replenishment. Not reflecting these constraints may result in frequent ad-hoc replenishments from the backroom. Figure 9 shows the trade-off decision for applying the constraints with $f_i^{\min} = 100\%$, $i \in \mathbb{N}$ versus neglecting the replenishment constraints completely, i.e., $f_i^{\min} = 0\%$, $i \in \mathbb{N}$ depending on the aggregated substitution level $\hat{\mu}_i$, $i \in \mathbb{N}$. Disregarding the constraints theoretically increases profits between 1 and 8%. With higher substitution levels, less profitable products are no longer listed, as they could be replaced by products with higher profit margins. Profit losses due to the replenishment constraints therefore decrease at higher substitution levels. In summary, replenishment constraints affect the achievable profit only moderately at a relevant substitution level ($\hat{\mu}_i \geq 0.2$, $i \in \mathbb{N}$), by 2% at most. However, the additional effort of replenishing shelves from the backroom has to be considered when neglecting the replenishment constraints. Between 67% and 87% of the items receive insufficient facings when replenishment constraints are neglected, a fact that becomes particularly relevant at higher substitution levels. This means that without the replenishment constraints,

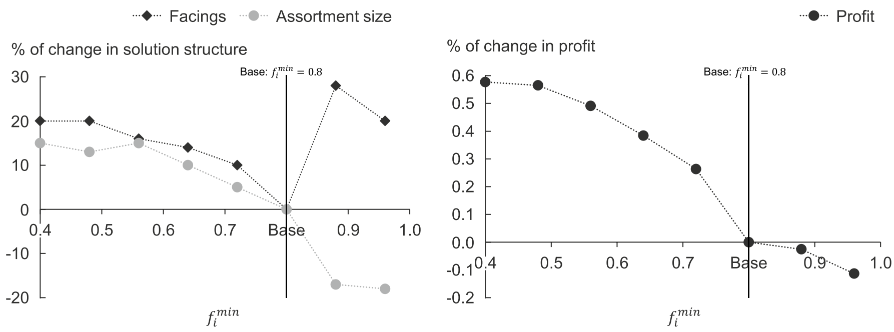

Fig. 10 Sensitivity analysis with replenishment constraints, case $|\mathbb{N}| = 50$, $\beta_i = 0.2$, $\lambda_i = 0.8$ and $\hat{\mu}_i = 0.8, i \in \mathbb{N}$

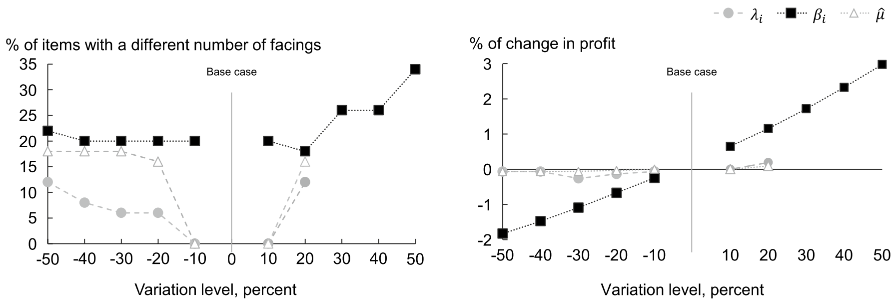

Fig. 11 Impact of variations of demand parameters on solution structures (left) total profits (right), CASRP$_{\text{approx}}$, example with $|\mathbb{N}| = 50$, $f_i^{\min} = 0.8$, $i \in \mathbb{N}$ and for the respective base cases (0%): $\lambda_i = \hat{\mu}_i = 0.8, \beta_i = 0.2, i \in \mathbb{N}$

the entire demand for more than two-thirds of the items cannot be fulfilled with the shelf inventory available, and therefore much more frequent restocking activities are necessary than allowed. This ultimately leads to much lower profits due to high operational costs. It is therefore essential that a retailer considers potential instore replenishment options when making assortment and facing decisions.

Figure 10 highlights the impact of varying minimum replenishment levels $f_i^{\min}$, $i \in \mathbb{N}$ on solution structure and total profits. Compared to the base value of $f_i^{\min} = 80\%$, applying a smaller level for $f_i^{\min}$ offers more opportunities to list additional items. Vice versa, the assortment sizes decrease with higher values for $f_i^{\min}$. The change in facings is mostly driven by the change in the assortment. The profit impact of the variation of $f_i^{\min}$ is in line with results from Figure 9. The higher the $f_i^{\min}$, the lower the total profits will be. However, the additional replenishment costs are not considered in this calculation. The lower the replenishment constraints, the more costly ad-hoc replenishments become necessary that by far outweighs the slight profit increase.

Impact of demand parameters A multitude of demand parameters need to be estimated. Errors and deviations cannot easily be excluded, so this section evaluates the impact of these parameters on the results of the model using sensitivity analyses. The parameters β_i and λ_i, $i \in \mathbb{N}$ and μ_{ij}, $i,j \in \mathbb{N}, j \neq i$ need to be estimated using consumer experiments and thus by their nature may deviate from real consumer behavior. In the following we vary each set of parameters separately by +/− 50% for β_i, and −50% and +20% for λ_i and $\widehat{\mu}_i$, $i \in \mathbb{N}$ compared to the base values. Note that the base values of λ_i and $\widehat{\mu}_i$ are set to 0.8 and cannot be greater than 1.0. The left side of Figure 11 shows a remarkable impact of varying space elasticity β_i, $i \in \mathbb{N}$ on total profit. Profit sensitivities are between −1.8% and +3.0% compared to the base value. The impact on the decision structure is also moderate (see right side of Figure 11). Up to 35% of the items have different facings. The impact on profit is very low and on solution structure only moderate for each of the demand shares λ_i and substitution levels $\widehat{\mu}_i$, $i \in \mathbb{N}$ analyzed. As this also represents the substitution volume in general, the results are in line with the substitution analysis above. The higher accuracy in estimating consumer behavior requires expensive consumer research. These sensitivity results may be used by business to justify additional investments for a better understanding of consumer behavior.

6 Conclusions and future areas for research

6.1 Summary

The increasingly competitive retail business requires ever greater customer orientation and operational efficiencies. Despite significant investments in infrastructure and IT, retailers are still losing potential revenue due to their inability to get the right goods to the right places at the right time. Analytics and optimization approaches will support companies on this journey to obtain effective structures and planning tools for retail space management. Efficiently managing retail shelf space is critical as the increase in product variety is in conflict with limited shelf space and instore replenishment constraints. This paper develops a general framework for retail space management and presents a decision support model for the related problems within the framework of optimizing assortment, shelf space assignment, and replenishment.

The model and solution approach described contributes to retail space management and extends shelf space models by replenishment constraints. It is based on consumer decisions affected by space effects and substitution. The operational constraints introduced ensure that shelf space decisions are aligned with the retailer's replenishment processes. Standard retail data have mainly been used, such as observed sales, which are available at a store level and have been supported by experimental data. We develop a two-step solution approach that precalculates the facing-dependent demand for all possible integer values of the facings and then solve the integrated assortment and shelf space problem with an exact, approximative and heuristic approach. Additionally, our approach provides results for retail-specific problem sizes within reasonable computation times, even for extreme cases with correlated profit-to-space data for hard knapsack problems.

Managerial implications extend across multiple dimensions. First, the integrated model achieves a profit increase of up to 14% in our case study, which demonstrates the need to integrate this into category planning processes and software applications. Second, integrating substitution and replenishment into the shelf space models does not just improve accuracy, it also allows better objective values and more realistic planograms to be achieved. Third, space effects and replenishment matter in terms of profit impact, assortment size and space assignment. This means that retailers gain from better understanding of the categories underlying space and substitution effects and replenishment systems. Finally, a constraint for the stock level achieved by regular period-based replenishment is required to obtain efficient instore positions and to avoid costly ad-hoc replenishment processes.

6.2 Future areas of research

The modeling and solution approach proposed can be extended in several directions, opening up new areas for future research: (1) supply assumptions, (2) demand-generating effects and (3) certainty of demand data. (1) The model assumes fixed restocking cycles, unlimited transportation, warehouse, and backroom capacity, as only showroom effects are considered. Additionally, delivery aspects from DCs to stores and the associated delivery patterns as well as enlarging the scope upwards of the supply chain will provide additional insights. (2) Further marketing activities and demand-generating effects should also be investigated in this context. These particularly include positioning effects to account for different shelf layers and "eye-level" demand, price effects with price and cross-price elasticity, and other marketing variables that generate instore demand. (3) Out-of-stock substitution does not occur as long as shelf space is sufficient in our model as efficient instore logistics are assumed and safety stocks are applied. Consumer demand is assumed to be deterministically known but is in reality subject to a certain volatility depending on external factors such as seasonality, temperature or the day of the week. Incorporating seasonality, stochastic demand and out-of-stock substitution would add an important and realistic modeling feature. Focusing on demand volatility would imply the development of a stochastic model for our decision problem. Out-of-stock substitutions resulting from potentially insufficient shelf and backroom quantities for specific items would need to be taken into consideration in such cases. A stochastic model would need to balance the tradeoffs between under- and overstock situations, which are specifically relevant in the case of perishable items.

Acknowledgements The authors would like to thank the anonymous reviewers and the editor for their valuable recommendations, which have significantly improved the paper.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Abbott H, Palekar US (2008) Retail replenishment models with display-space elastic demand. Eur J Oper Res 186(2):586–607

Akkaş A (2019) Shelf space selection to control product expiration. Prod Oper Manag 28(9):2184–2201

Bai R, van Woensel T, Kendall G, Burke EK (2013) A new model and a hyper-heuristic approach for two-dimensional shelf space allocation. 4OR 11(1):31–55

Berg JPvd, Sharp GP, Gademann A, Pochet Y (1998) Forward-reserve allocation in a warehouse with unit-load replenishments. Eur J Oper Res 111(1):98–113

Bianchi-Aguiar T, Silva E, Guimaraes L, Carravilla MA, Oliveira JF (2018) Allocating products on shelves under merchandising rules: multi-level product families with display directions. Omega 76:47–62

Bianchi-Aguiar T, Hübner A, Carravilla MA, Oliveira JF (2021) Retail shelf space planning problems: a comprehensive review and classification framework. Eur J Oper Res 289(1):1–16

Bianchi-Aguiar T, Carravilla MA, Oliveira JF (2015) Replicating shelf space allocation solutions across retail stores. In: Working paper University Porto, pp 1–24

Boatwright P, Nunes JC (2001) Reducing assortment: an attribute-based approach. J Mark 65:50–63

Breuer P, Fritsch Av, Prauschke C, Steegmann J (2009) Looking for the right offer: assortment and understanding customer behavior are the hot topics in the year of crisis. Akzente 4(3):8–13

Broekmeulen R, van Donselaar K, Fransoo J, Woensel Tv (2006) The opportunity of excess shelf space in grocery retail store. Oper Res 49:710–719

Broekmeulen RA, Sternbeck MG, van Donselaar KH, Kuhn H (2017) Decision support for selecting the optimal product unpacking location in a retail supply chain. Eur J Oper Res 259(1):84–99

Brown WM, Tucker WT (1961) The marketing center: vanishing shelf space. Atlanta Econ Rev 9:9–13

Chandon P, Hutchinson WJ, Bradlow ET, Young SH (2009) Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. J Mark 73(November):1–17

Chuang HH-C, Oliva R, Liu S (2016) On-shelf availability, retail performance, and external audits: a field experiment. Prod Oper Manag 25(5):935–951

Corsten DS, Gruen TW (2003) Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. Int J Retail Distrib Manag 31(12):605–617

Corstjens M, Doyle P (1981) A model for optimizing retail space allocations. Manage Sci 27(7):822–833

Cox K (1964) The responsiveness of food sales to shelf space changes in supermarkets. J Mark Res 1(2):63–67

Curhan RC (1972) The relationship between shelf space and unit sales in supermarkets. J Mark Res 9(4):406–412

DeHoratius N, Raman A (2008) Inventory record inaccuracy: an empirical analysis. Manage Sci 54(4):627–641

DeHoratius N, Holzapfel A, Kuhn H, Mersereau AJ, Sternbeck M (2023) Evaluating count prioritization procedures for improving inventory accuracy in retail stores. Manuf Serv Oper Manag 25(1):288–306

DeHoratius N, Ton Z (2015) The role of execution in managing product availability. In: Agrawal N, Smith SA (eds) Retail supply chain management. International Series in Operations Research & Management Science, 2nd edn. Springer, Boston, MA, pp 53–77

Donselaar KHv, Gaur V, Woensel Tv, Broekmeulen RA, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. Manage Sci 56(5):766–784

Drèze X, Hoch SJ, Purk ME (1994) Shelf management and space elasticity. J Retail 70(4):301–326

Düsterhöft T, Hübner A (2023) Differences, problems and opportunities of applied optimization models in retail space analytics. In: Ghoniem A, Maddah B (eds) Retail space analytics. Springer, Boston, MA

Düsterhöft T, Hübner A, Schaal K (2020) A practical approach to the shelf-space allocation and replenishment problem with heterogeneously sized shelves. Eur J Oper Res 282(1):252–266

EHI Retail Institute (2014) Retail data 2014: Structure, key figures and profiles of international retailing

Eisend M (2014) Shelf space elasticity: a meta-analysis. J Retail 90:168–181

Eroglu C, Williams BD, Waller MA (2013) The backroom effect in retail operations. Prod Oper Manag 22(4):915–923

Europe ECR (2003) Handbook ECR-Demand Side. EHI, Cologne

Farahat A, Lee J (2018) The multiproduct newsvendor problem with customer choice. Oper Res 66(1):123–136

Fisher M (2009) Or forum-rocket science retailing: the 2006 philip mccord morse lecture. Oper Res 57(3):527–540

Fitzsimons GJ (2000) Consumer response to stockouts. J Consum Res 27:249–266

Flamand T, Ghoniem A, Maddah B (2016) Promoting impulse buying by allocating retail shelf space to grouped product categories. J Oper Res Soc 67(7):953–969

Flamand T, Ghoniem A, Haouari M, Maddah B (2018) Integrated assortment planning and store-wide shelf space allocation: an optimization-based approach. Omega 81:134–149

Frank RE, Massy WF (1970) Shelf position and space effects on sales. J Mark Res 7(1):59–66

Frank M, Ostermeier M, Holzapfel A, Hübner A, Kuhn H (2021) Optimizing routing and delivery patterns with multi-compartment vehicles. Eur J Oper Res 293(2):495–510

Gajjar HK, Adil GK (2010) A piecewise linearization for retail shelf space allocation problem and a local search heuristic. Ann Oper Res 179(1):149–167

Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. Manage Sci 52:1528–1543

Geismar HN, Dawande M, Murthi BP, Sriskandarajah C (2015) Maximizing revenue through two-dimensional shelf-space allocation. Prod Oper Manag 24:1148–1163

Ghoniem A, Maddah B (2015) Integrated retail decisions with multiple selling periods and customer segments: optimization and insights. Omega 55:38–52

Griswold M (2007) Space management: align business challenges and it vendors. AMR Res, pp 1–17

Gruen WT, Corsten S, Bharadwaj S (2002) Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses: report. Grocery Manufacturers of America, Washingtion and D.C

Hansen P, Heinsbroek H (1979) Product selection and space allocation in supermarkets. Eur J Oper Res 3(6):474–484

Hansen JM, Raut S, Swami S (2010) Retail shelf allocation: a comparative analysis of heuristic and meta-heuristic approaches. J Retail 86(1):94–105

Hariga MA, Al-Ahmari A, Mohamed A-RA (2007) A joint optimisation model for inventory replenishment, product assortment, shelf space and display area allocation decisions. Eur J Oper Res 181(1):239–251

Holzapfel A, Hübner A, Kuhn H, Sternbeck M (2016) Delivery pattern and transportation planning in grocery retailing. Eur J Oper Res 252(1):54–68

Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. Oper Res 58(5):1364–1379

Honhon D, Jonnalagedda S, Pan XA (2012) Optimal algorithms for assortment selection under ranking-based consumer choice models. Manuf Serv Oper Manag 14(2):279–289

Hübner A (2017) A decision support system for retail assortment planning. Int J Retail Distrib Manag 45(7/8):808–825

Hübner AH, Kuhn H (2012) Retail category management: a state-of-the-art review of quantitative research and software applications in assortment and shelf space management. Omega 40(2):199–209

Hübner A, Schaal K (2017) Effect of replenishment and backroom on retail shelf-space planning. Bus Res 10(1):123–156

Hübner A, Schaal K (2017) An integrated assortment and shelf-space optimization model with demand substitution and space-elasticity effects. Eur J Oper Res 261(1):302–316

Hübner AH, Kuhn H, Sternbeck MG (2013) Demand and supply chain planning in grocery retail: an operations planning framework. Int J Retail Distrib Manag 41(7):512–530

Hübner A, Kuhn H, Kühn S (2016) An efficient algorithm for capacitated assortment planning with stochastic demand and substitution. Eur J Oper Res 250(2):505–520

Hübner A, Schäfer F, Schaal K (2020) Maximizing profit via assortment and shelf-space optimization for two-dimensional shelves. Prod Oper Manag 29(3):547–570

Hübner A, Düsterhöft T, Ostermeier M (2021) Shelf space dimensioning and product allocation in retail stores. Eur J Oper Res 292(1):155–171

Hwang H, Choi B, Lee M-J (2005) A model for shelf space allocation and inventory control considering location and inventory level effects on demand. Int J Prod Econ 97(2):185–195

Irion J, Lu J-C, Al-Khayyal FA, Tsao Y-C (2011) A hierarchical decomposition approach to retail shelf space management and assortment decisions. J Oper Res Soc 62(10):1861–1870

Irion J, Lu J-C, Al-Khayyal F, Tsao Y-C (2012) A piecewise linearization framework for retail shelf space management models. Eur J Oper Res 222(1):122–136

Iyengar P, Lepper M (2000) When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol 79(6):995–1006

Kellerer H, Pferschy U, Pisinger D (2004) Knapsack problems. Springer, Berlin

Kök G (2003) Management of product variety in retail operations: Ph.D. Disseration. The Wharton School, University of Pennsylvania, Pennsylvania

Kök GA, Fisher ML (2007) Demand estimation and assortment optimization under substitution: methodology and application. Oper Res 55(6):1001–1021

Kök G, Fisher ML, Vaidyanathan R (2015) Assortment planning: review of literature and industry practice. In: Agrawal N, Smith SA (eds) Retail Supply Chain Management, vol 223. International Series in Operations Research & Management Science. Springer, US, pp 175–236

Kotzab H, Teller C (2005) Development and empirical test of a grocery retail instore logistics model. Br Food J 107(8):594–605

Kotzab H, Reiner G, Teller C (2005) Instore-logistik als neuralgischer brückenkopf einer supply chain: Modellbildung, empirische bestandsaufnahme und simulation. In: Lasch R (ed) Logistik-Management. Gabler, Wiesbaden, pp 281–293

Kuhn H, Sternbeck MG (2013) Integrative retail logistics: an exploratory study. Oper Manag Res 6(1–2):2–18

Kunnumkal S, Martinez-de Albeniz V (2019) Tractable approximations for assortment planning with product costs. Oper Res 67(2):436–452

Lim A, Rodrigues B, Zhang X (2004) Metaheuristics with local search techniques for retail shelf-space optimization. Manage Sci 50(1):117–131

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. Oper Res 49:334–351

Mou S, Robb DJ, DeHoratius N (2018) Retail store operations: literature review and research directions. Eur J Oper Res 265(2):399–422

Obermair E, Holzapfel A, Kuhn H (2023) Operational planning for public holidays in grocery retailing - Managing the grocery retail rush. Oper Manag Res. https://doi.org/10.1007/s12063-022-00342-z

Ostermeier M, Hübner A, Düsterhöft T (2021) A model and solution approach for store-wide shelf space allocation. Omega 102:102425

Pires M, Camanho A, Amorim P (2020) Solving the grocery backroom sizing problem. Int J Prod Res 58(18):5707–5720

Pires M, Amorim P, Liz J, Pratas J (2015) Design of retail backroom storage: a research opportunity? In: Studies in Big Data. Springer International Publishing, pp 167–174

Pisinger D (2005) Where are the hard knapsack problems? Comput Oper Res 32(9):2271–2284

Raman A, DeHoratius N, Ton Z (2001) Execution: the missing link in retail operations. Calif Manage Rev 43(3):136–152

Reiner G, Teller C, Kotzab H (2012) Analyzing the efficient execution on in-store logistics processes in grocery retailing - the case of dairy products. Prod Oper Manag 22(4):924–939
Schaal K, Hübner A (2018) When does cross-space elasticity matter in shelf-space planning? A decision analytics approach. Omega 80:135–152
Sloot LM, Verhoef PC (2008) The impact of brand delisting on store switching and brand switching intentions. J Retail 84(3):281–296
Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. Oper Res 48(1):50–64
Sternbeck MG (2015) A store-oriented approach to determine order packaging quantities in grocery retailing. J Bus Econ 85(1):569–596
Sternbeck M (2022) Robot-assisted instore operations in grocery retailing. In: Villani L, Natale C, Beetz M, Siciliano B (eds) Robotics for intralogistics in supermarkets and retail stores. Springer International Publishing, Cham, pp 1–40
Sternbeck MG, Kuhn H (2014) An integrative approach to determine store delivery patterns in grocery retailing. Transp Res Part E Logist Transp Rev 70:205–224
Tompkins (2014) Retail backrooms. Tompkins International. Technical Report
Transchel S (2017) Inventory management under price-based and stockout-based substitution. Eur J Oper Res 262(3):996–1008
Transchel S, Buisman ME, Haijema R (2022) Joint assortment and inventory optimization for vertically differentiated products under consumer-driven substitution. Eur J Oper Res 301(1):163–179
Turgut Ö, Taube F, Minner S (2018) Data-driven retail inventory management with backroom effect. OR Spectr 40(4):945–968
Urban TL (1998) An inventory-theoretic approach to product assortment and shelf-space allocation. J Retail 74(1):15–35
van Donselaar KH, Broekmeulen RA (2013) Determination of safety stocks in a lost sales inventory system with periodic review, positive lead-time, lot-sizing and a target fill rate. Int J Prod Econ 143(2):440–448
Woensel Tv, Donselaar KHv, Broekmeulen RACM, Fransoo JC (2007) Consumer responses to shelf out-of-stocks of perishable products. Int J Phys Distrib Logist Manag 37(9):704–718
Xin G, Messinger PR, Li J (2009) Influence of soldout products on consumer choice. J Retail 85(3):274–287
Yang M-H (2001) An efficient algorithm to allocate shelf space. Eur J Oper Res 131(1):107–118
Yang M-H, Chen W-C (1999) A study on shelf space allocation and management. Int J Prod Econ 60–61(4):309–317
Zelst Sv, Donselaar Kv, Woensel Tv, Broekmeulen R, Fransoo J (2009) Logistics drivers for shelf stacking in grocery retail stores: Potential for efficiency improvement. Int J Prod Econ 121(2):620–632
Zufryden FS (1986) A dynamic programming approach for product selection and supermarket shelf-space allocation. J Oper Res Soc 37(4):413–422

Alexander Hübner is a Full Professor for Supply and Value Chain Management at Technical University of Munich, Germany. He received his Ph.D. from Catholic University of Eichstätt-Ingolstadt. His research focus on decision support systems in service operations. His interests include combinatorial optimization, heuristics and retail operations.

Heinrich Kuhn is a Full Professor for Supply Chain Management and Operations at the Catholic University of Eichstätt-Ingolstadt, Germany. He received his Ph.D. in Industrial Engineering from Darmstadt University of Technology, Germany, and completed his formal qualifications as a professor in Business Administration at the University of Cologne, Germany. His research interests focus on performance analysis of service and production systems, resource scheduling, last-mile delivery, retail operations and inventory management.

Authors and Affiliations

Alexander Hübner[1] · **Heinrich Kuhn**[2]

Alexander Hübner
alexander.huebner@tum.de

1 Technical University of Munich, Supply and Value Chain Management, Am Essigberg 3, 94315 Straubing, Germany

2 Catholic University Eichstätt-Ingolstadt, Supply Chain Management & Operations, Auf der Schanz 49, 85049 Ingolstadt, Germany