ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Bsc2 Econometrics/Economics

Enhanced integration of assortment planning and store-wide shelf space allocation: an exploration of demand effects

Name student: Jingli Gao

Student ID number: 450076

Supervisor: Kuryatnikova, O.

Second assessor: Maasakkers, F.J.L. van

Date final version: 4th of July, 2021

ABSTRACT

This thesis paper considers the incorporation of multiple practically relevant considerations into an integrated model for assortment selection and store-wide shelf space allocation. The integration is fortified by incorporating demand effects and reformulating the model of Flamand et al. (2018) to adhere to the Newsvendor model setting. Moreover, an state-of-the-art model is formulated and the practical feasibility of implementing it is assessed.

Contents

1 Introduction

Deciding which products to put where in the store has always been a challenge for retailers. One difficulty of this shelf space allocation problem is its interdependency with the retailer's assortment and in-store stock replenishment decisions (Campo and Gijsbrechts, 2005). The academic literature on these category management decision areas is progressively gaining popularity (Bianchi-Aguiar et al., 2020), as well as the practical relevance of good solutions. For instance, the competitive landscape for retailers has intensified because of the continuous introduction of new products (Hübner and Kuhn, 2012). Additionally, in a post COVID-19 environment, the demand for good solutions is possibly even larger, as retailers may decide to reduce their store footprint (Gecili, 2020) due to the switch to online shopping. It is therefore essential for retailers to efficiently utilize their scarce shelf space. However, while on the one hand the academic literature becomes more vast, and on the other hand the demand for suitable solutions increases, there still exists a large discrepancy between the academic models and the tools used in practice. Hübner and Kuhn note that commercial software applications used by practitioners are too simplistic and lack the integration of mathematical optimization and relevant consumer demand effects, whereas state-of-the-art models from the academic literature are practically infeasible for large problem instances. This thesis paper aims to contribute to reducing this discrepancy between academics and practice by assessing the computational burden and practicality of the inclusion of several key demand effects into the integrated modelling approach of Flamand et al. (2018). Another contribution of this paper is the incorporation of several findings from multiple papers into one integrated model. The results give insights into which practically relevant demand effects are feasible and interesting to incorporate into a model that has shown promising results in terms of the computational burden of large problem instances.

As a starting and reference point, the computational study of Flamand et al. (2018) on their proposed Mixed Integer Programming (MIP) model formulation and heuristic approach is used. Their model does not consider any demand effects at all, despite the significance of those effects shown in for example Hübner et al. (2020). The sensitivity of the computational performance is therefore assessed when enhancing the model with several key demand effects, such as space elasticity and substitution effects, and switching from deterministic demand to stochastic demand. Simultaneously, the (economic) relevance of all the considered enhancements is elaborated on. In combination with the results of the computational study on the performance of the heuristic approach of Flamand et al., a discussion is then given on the trade-off between computational burden and practical relevance. The results can be used by practitioners to select an optimal set of considerations to include for their integrated assortment planning and store-wide shelf space planning problem. In summary, this thesis paper aims to provide an answer to the research question formulated below:

- *Which practically relevant considerations are feasible to incorporate when modelling an integrated retail assortment planning and store-wide shelf space allocation problem?*

To answer the above research question, the following sub-questions are considered:

- *What practically relevant considerations are addressed in the existing literature on assortment planning and shelf space allocation?*
- *Which of these considerations can be incorporated as an extension to the model of Flamand et al. (2018)?*
- *What is the effect of including these considerations on the optimality gap and computational time?*

The remainder of this thesis paper is structured as follows. First, in Section 2 the existing literature is reviewed and consolidated into a list of relevant considerations retailers could potentially make in practice. Then, the methodology for incorporating these considerations into a model and the corresponding computational study is introduced in Section 3. The results of the computational study are given in Section 4 and a conclusion and discussion is made in Section 5.

2 Literature review

This section addresses the first sub-question by providing an overview of considerations a retailer possibly needs to make for their assortment and shelf space planning problem. Campo and Gijsbrechts (2005) note that these decisions are interdependent because of supply- and demand-based interdependencies. On the supply-side, the retailer has to bear in mind that a larger assortment results in a relatively smaller availability of shelf space per product and therefore potentially lower sales due to for example stockouts. On the demand-side, the retailer needs to estimate the customers' reaction to a change in assortment and/or shelf space allocation. Thus, there exists a clear need for an integrated approach for these two category management decisions, towards which the paper of Flamand et al. (2018) constitutes a step forward. They formulate the integrated problem as a Mixed Integer Program (MIP) with as objective the maximization of a measure of overall store profitability. In their profitability calculation, they account for: *(a) impulse purchasing potential* of a product category, differing attractiveness of shelf segments, space elastic demand effects, and profit margins. Their interpretation of (a) impulse purchasing potential is the likelihood of an unplanned (impulsive) purchase, given that the product is seen by the customer. The authors note that their definition is similar to the impulse purchase rate in for example Inman et al. (2009), which show that the baseline probability of a unplanned purchase is about 46% and can go up to 93% depending on so-called contextual factors such as the hedonic nature of a product. In this thesis paper a more generalized interpretation is considered and given in Section 3. Another practically relevant consideration the authors make is multiplying the expected demand with a term they refer to as the visibility of a product category on the shelves. This term can be interpreted as an aggregated form of *(d) vertical position-dependent demand*, *(e) horizontal position-dependent demand*, and *(b) space elastic demand effects*, which is elaborated on in Section 3. The economic intuition is that certain positions on the shelf, and the location of the shelf as a whole, are more likely to be visited than others, thereby increasing the likelihood of noticing the products on those shelves. For example, products have higher sales when displayed on eye level instead of on the bottom shelf (Dreze et al., 1994), and when displayed towards the end of an aisle instead of in the middle (Van Nierop et al., 2008). In addition, the authors' formulation implies that they assume this effect on sales to be additive with respect to the number of shelf segments a product is assigned to. This can be interpreted as a form of (b) space elastic demand effects, as the demand increases with the amount of shelf space allocated to a product. Finally, for the profit margins in the profitability calculation, the authors simply consider revenue minus the unit costs on product category level, but then aggregated into one overall profit margin term. In this thesis paper the profit margin is disaggregated into an extended profit function by also considering salvage values and shortage costs, which is elaborated on in Section 3.

While Flamand et al. (2018) do account for the aforementioned practically relevant considerations, their model is still too simplistic in the sense that it only considers an aggregated form of a limited set of demand effects, does not consider any supply limitations, and does not consider the width, length, and height components of product sizes and shelf capacities. Therefore, the focus of the remainder of this literature review is to explore how these limitations have been addressed in other studies. Special attention is given to considerations on the demand and the supply side in the field of assortment planning and shelf space allocation, in order to strengthen the integration of assortment selection and shelf space planning. To do this, each decision area is assessed separately. Afterwards, in Section 3 an attempt is made to incorporate several of these considerations into the model formulation of Flamand et al. (2018).

2.1 Assortment planning

Retailers usually first determine which products to include in their product offering (Bianchi-Aguiar et al., 2020). This step is known as *assortment planning*, for which an important parameter to consider is a customer's willingness to substitute within a particular product category (Kök et al., 2015). Kök et al. identify three sources of consumer driven substitution: *(f) out-of-assortment substitution* (OOA), *(g) out-of-stock substitution* (OOS), and an utility-based buying decision. For the first two sources, the customer buys a substitute because the original product is stocked out or because the store does not offer it, respectively. The customer is thus aware that it bought a substitute. However, for the third source, the customer might not be aware, as the substitute was simply bought because it would give a higher utility than not buying anything at all. It could thus well be that there exists another product that would have given an even higher utility, but the customer was simply not aware of its existence. While Hübner (2011) note that about 69% up to 84% of quantity demanded can be substituted, depending on the perishability of the product (Van Woensel et al., 2007), it is also important to note that stock-out losses disproportionately grow with the frequency and duration of the unavailability of a product (Campo et al., 2004). With an eye on the long-term strategy of a retailer, it is reasonable expected that the incorporation of this undesirability of stock-outs is important.

Kök et al. (2015) also outline three demand models that typically serve as the foundation for assortment planning. The most stylized model is the Multinomial Logit (MNL) model, which assumes that every individual gets a certain (expected) utility for each possible choice and will choose the option that maximizes the expected utility. The major advantage is that it is relatively easy to add, for example, marketing mix variables to the model. The major limitation is that the MNL model is too restrictive in terms of substitution patterns due to its Independence of Irrelevant Alternatives (IIA) property. For instance, suppose there exists a store that sells one non-alcoholic beverage and one alcoholic beverage. Then if the retailer were to add another alcoholic beverage, one would expect this to cannibalize the sales of the other alcoholic beverage more than it would cannibalize the non-alcoholic beverage. However, the IIA property implies that the cannibalization effect would be the same for both products. Another utility-based model commonly used to reflect consumer behavior in assortment planning, is the Locational Choice model, as described in Kök et al. (2015). The major difference is that the Locational Choice model allows to have different substitution effects for different products, whereas the MNL model assigns an equal effect to all products regardless of their similarities with the original product. The third demand model used in assortment planning is the Exogenous Demand model, which is the most commonly used option in the literature to reflect substitution effects in inventory management (Kök et al., 2015). The biggest difference with the previous two models is that for the Exogenous Demand (ED) model, there is no underlying consumer behavior model. The ED model can therefore only capture aggregated consumer demand. Instead of looking at individual consumer demand, it directly specifies both the initial and substitution demand for each product category. The advantage is that it allows for more flexibility in terms of incorporating demand effects such as within-subgroup substitution, resulting in a more realistic representation of expected demand. Other examples include the ability to distinguish between OOS and OOA substitution, and to differentiate the demand for value packs versus single products, as is also noted in Kök et al. (2015).

2.2 Shelf space allocation

After retailers have decided on their assortment, usually the next step is to decide where and how much of the selected product categories to put on the shelf. Bianchi-Aguiar et al. (2020) performed a comprehensive review on the literature related to the shelf space allocation problem. They note that there are essentially three decisions a retail shelf space planner needs to make for each product category: (I) the total shelf quantity, (II) the vertical shelf level, and (III) its horizontal distance from the aisle. Furthermore, Bianchi-Aguiar et al. (2020) identified two different demand types for which multiple empirical studies have shown that they are positively influenced by shelf space allocation decisions: (1) space-dependent demand, and (2) position-dependent demand.

Space-dependent demand effects originate from the amount of space assigned to a product category. One of them is *(b) space elastic demand*, which essentially reflects that the demand for a product grows with the amount of space it is assigned, as the visibility will then be higher. And naturally, if the visibility is higher, the likelihood of customers noticing and buying the product also increases. Another space-dependent demand effect is *(c) cross-space elastic demand*, which reflects the interdependency of a pair of products. If for example two substitutes were to be put adjacent to each other on the shelf, one would expect the sales to be lower than if they were put on another shelf. Similarly, if two complementary products were to be put next to each other, it would probably stimulate overall sales. Bianchi-Aguiar et al. (2020) also note that the common practice for both of these space-dependent demand effects is to incorporate them as a power function, implying that the (implicit) implementation of space elastic demand by Flamand et al. (2018) is uncommon.

Position-dependent demand effects originate from the vertical and horizontal location a product is allocated to on a shelf rack. These include (d) *vertical position-dependent demand* and (e) *horizontal position-dependent demand*, which reflect that certain parts of a shelf are more likely to be noticed by consumers than others, as noted earlier in this section. Bianchi-Aguiar et al. (2020) note that these two demand effects are commonly studied jointly in the existing literature, implying that the aggregated formulation of Flamand et al. (2018) is consistent with the literature in this case. Another demand effect is arrangement-dependent demand, which reflects that well-organized shelves are more likely to attract the attention of a customer than shelves with no structure. To incorporate this effect, a model needs to incorporate merchandising rules, which are constraints for product arrangements to reflect consumer buying behavior and/or the strategy of a company (Bianchi-Aguiar et al., 2020). This also greatly adds to the practical feasibility of the final solution, as it limits the likelihood of constructing a product category arrangement that feels unnatural to the customer. Flamand et al. (2018) already incorporated some merchandising rules by defining product category interdependence, taking into account that some product categories should be put adjacent to each other and others absolutely not. For example, putting detergents and vegetables next to each other would most likely feel unnatural to the customer. Another well-known example of merchandising rules in the literature is the preference for product categories to be presented in block formations, given that the product category is allocated to multiple shelf segments (Geismar et al., 2015). Although the incorporation of additional merchandising rules is definitely of practical relevance, it is unfortunately considered to be out of the scope of this thesis paper.

Next to only considering demand effects which mainly affect the revenue side of the profit maximization goal, it is also of interest for retailers to consider the cost side. Most literature on shelf space management focuses primarily on the demand side (Hübner and Kuhn, 2012). This thesis paper generalizes the objective function of Flamand et al. to allow for distinguishing between revenues, salvage values, unit costs, and shortage costs, based on the objective function of the model in Hübner et al. (2016). Salvage value represents the revenue from leftover products and shortage costs can be interpreted as penalty costs for not having enough quantity of a product in stock. The cost side can be incorporated even more by for example integrating inventory problems as well, but this integration is beyond the scope of this study.

Finally, allowing for distinguishing between the width, length, and height components of product sizes and shelf spaces might significantly add to the practicality of the model solution, as is demonstrated in for example Geismar et al. (2015) and Rabbani et al. (2018). This would allow for a more realistic representation of the amount of stock to put on the shelves as the dimensions of a product's packaging, and sometimes also the shelf itself, differ significantly in practice. Take for example some product A with dimensions

$(width, length, height) = (5, 5, 5)$, some product B with dimensions $(5, 10, 5)$, and a shelf with dimensions $(5, 10, 5)$. If a model only considers one dimension, say width, then the model can assign at most one unit of product A or one unit of product B, whereas in reality the retailer could have placed two units of product A, instead of only one. The model of Flamand et al. only considers one dimension of both products and shelves, and could therefore be enhanced by adjusting their model formulation to allow for 2D or 3D considerations. However, this is also out of the scope of this thesis paper.

2.3 Integrated assortment planning and shelf space allocation

Assortment planning models tend to have a stochastic nature, whereas shelf space decisions are usually modeled with deterministic demand (Hübner and Kuhn, 2012). Although most assortment planning and shelf space allocation decisions have been investigated separately in the literature, there also exists a variety of integrated model approaches. However, Hübner et al. (2016) note that an efficient and effective integrated solution approach has not yet been developed. They note that common limitations of existing models are their unsuitability for large problem sizes, disregarding substitution effects, disregarding costs, or limited applicability when demand is stochastic. Based on the extensive literature review of Bianchi-Aguiar et al. (2020), which have classified related planning problems of 55 different papers, only about 7% of the publications in the shelf space planning literature consider stochastic demand. Interesting to note is that for all the papers Bianchi-Aguiar et al. have identified to consider stochastic demand, the objective profit function is based on the newsvendor model (Edgeworth, 1888). Hübner et al. (2016) note that accounting for stochastic demand is especially important when considering perishable products and in addition it also reflects consumer behavior more comprehensively, allowing the retailer to derive more consumer insights. Other benefits of using a newsvendor model setting are its considerations for salvage value and shortage costs and its concave formulation (Khouja, 1999). The objective function in Flamand et al. (2018) is therefore reformulated to adhere to the newsvendor model setting. This is done based on the methodology of Hübner et al. (2016), as stated before.

3 Methodology

Generally, a retailer's objective is to have their assortment allocated on the shelves in such a way that overall store profits are maximized, given some constraints. In the previous section, it has been made clear that the decision making for assortment planning and shelf space allocation are intertwined. Furthermore, it has been noted that currently there does not exist a paper in the literature that effectively and efficiently jointly models the assortment and shelf space allocation problems, while also taking into account the demand and cost effects elaborated on in Section 2. This paper constitutes a step towards a more practical tool for jointly deciding on assortment and shelf space allocation by incorporating demand and cost effects into a model that has shown promising computational results for large problem instances. In the remainder of this paper, a simplistic deterministic demand function is used as a starting point and enhanced with the practically relevant considerations (a) — (g) described in Section 2, thereby addressing the second sub-question. Before this, first the original model formulation of Flamand et al. is given and interpreted in Section 3.1. Then, the objective function is reformulated and the incorporation of model enhancements (a) — (g) is formulated in Section 3.2. Afterwards, the feasibility of implementing these formulations into the MIP of Flamand et al. (2018) is assessed in Section 3.3. Finally, the methodology for the computational study used to assess the computational burden and practical relevance of implementing the feasible set of model enhancements is given in Section 3.4.

3.1 Original model formulation

The notation of Flamand et al. (2018) for their original model formulation of the Mixed Integer Programming (MIP) model APSA is given in Table 3.1. Given the notation, Flamand et al. formulate the MIP as given in Equations (3.1a) — (3.1s). To enhance this model formulation, most changes are made to the objective value function (3.1a). Considering one product j only and splitting the function into two terms, $\Phi_j \cdot \sum_{k \in \mathcal{K}} \frac{f_k \cdot s_{k,j}}{c_k}$, helps to understand the formulation better and makes it more clear where adjustments might be interesting. The first term, Φ_j, represents the (aggregated) largest possible profit of product category j and is calculated as the product of expected demand volume v_j, profit margin ρ_j, and impulse purchase potential γ_j. Flamand et al. do not consider any substitution demand effects, and assume v_j to be deterministic. The model can thus be enhanced by replacing v_j with a (stochastic) demand function d_j, which is elaborated on in the next subsection. The profit margin ρ_j is also only considered in its aggregated form in Flamand et al., disregarding possible fluctuations in its revenue and cost components. This thesis paper generalizes their formulation and considers the profit on a more granular level based on the item-specific profit function π_j given in Hübner et al. (2016) with revenue rev_j, unit costs cos_j, salvage costs sal_j, shortage costs sho_j, and their dependency on (stochastic) demand. The formulation is in accordance with the newsvendor model setting, which is a single-period model and is outlined in for example Khouja (1999), as stated in Hübner and Schaal (2017). More elaboration is given in the next subsection. For γ_j, Flamand et al. consider two disjoint sets of products: (1) the set of fast-moving products $\mathcal{F}$, defined as the 20% most selling products that jointly account for 80% of overall sales in terms of quantities, typically having low profit margins, and (2) the set of slow-moving products, defined as $\mathcal{N} \setminus \mathcal{F}$, typically having relatively high profit margins. The authors assume that for fast-moving products $\gamma_j = 1$, and for slow-moving product $\gamma_j \in (0, 1]$. Based on this definition, γ_j can be interpreted as a penalty cost term for products that are less likely to be bought impulsively. This would also imply that the expected demand v_j for a product j should include all impulse purchases as well. However, in this thesis paper the original demand for a product is considered to be the base demand α_j as in Kök and Fisher (2007). In this setting it would be more intuitive and logical to define $\gamma_j \in [1, \infty)$ as a reward term, representing the additional demand for a product j as a result of unplanned impulse purchases. Another advantage is that this alteration allows a retailer to adjust γ_j to represent temporary increases in demand,

Table 3.1

Notation used for the Mixed Integer Program (MIP) model APSA, as given in Flamand et al. (2018)

Notation	Description
Indices	
i	Index for all shelves $\in \mathcal{B}$
j	Index for all product categories $\in \mathcal{N}$
k	Index for all shelf segments $\in \mathcal{K}$
Sets	
$\mathcal{N}$	Set with all product categories
$\mathcal{F} \subset \mathcal{N}$	Subset with approximately 20% of product categories in $\mathcal{N}$ that jointly contribute nearly 80% of the expected sales
$\mathcal{I} \equiv \mathcal{N} \setminus \mathcal{F}$	Subset with remaining 80% of product categories in $\mathcal{N}$
$\mathcal{L}$	Set with product pairs $(j, j')\in \mathcal{N}^2$ exhibiting allocation disaffinity
$\mathcal{H}_1$	Set with product pairs $(j, j')\in \mathcal{N}^2$ exhibiting symmetric assortment affinity
$\mathcal{H}_2$	Set with product pairs $(j, j')\in \mathcal{N}^2$ exhibiting asymmetric assortment affinity
$\mathcal{H}_3$	Set with product pairs $(j, j')\in \mathcal{N}^2$ exhibiting allocation affinity
$\mathcal{B}$	Set with all shelves
$\mathcal{K}_i$	Set with consecutive shelf segments along shelf $i \in \mathcal{B}$
$\mathcal{K} \equiv \cup_{i\in\mathcal{B}}\mathcal{K}_i$	Set with all shelf segments in the store
Parameters	
ρ_j	Profit margin of product category $j \in \mathcal{N}$
v_j	Expected demand volume of product category $j \in \mathcal{N}$
$\gamma_j \in (0,1]$	Likelihood of impulsive purchase of product $j \in \mathcal{N}$ (impulse purchase potential)
$f_k \in (0,1]$	Likelihood of a customer visiting shelf segment $k \in \mathcal{K}$ (shelf attractiveness)
$\Phi_j \equiv \gamma_j \times \rho_j \times v_j$	Largest possible profit of product category $j \in \mathcal{N}$
l_j	Lower limit of shelf space allocated to product category $j \in \mathcal{N}$
u_j	Upper limit of shelf space allocated to product category $j \in \mathcal{N}$
ϕ_j	Minimum space to be allocated to product category $j \in \mathcal{N}$ along any segment $k \in \mathcal{K}$ it is assigned to
α_i	Smallest index of a segment $k \in \mathcal{K}$ that belongs to shelf $i \in \mathcal{B}$
β_i	Largest index of a segment $k \in \mathcal{K}$ that belongs to shelf $i \in \mathcal{B}$
c_k	Capacity of segment $k \in \mathcal{K}$
$c^{max} = \max_{k\in\mathcal{K}} c_k$	Maximum shelf segment capacity of all shelf segments $k \in \mathcal{K}$
$C_i = \sum_{k\in\mathcal{K}} c_k$	Capacity of shelf $i \in \mathcal{B}$
Decision variables	
$x_{i,j} \in \{0,1\}$	$x_{i,j} = 1$ if and only if product category $j \in \mathcal{N}$ is assigned to shelf $i \in \mathcal{B}$
$y_{k,j} \in \{0,1\}$	$y_{k,j} = 1$ if and only if product category $j \in \mathcal{N}$ is assigned to shelf segment $k \in \mathcal{K}$
$z_{j,j'} \in \{0,1\}$	$z_{j,j'} = 1$ if and only if product categories j and $j' \in \mathcal{N}$ are selected in the assortment simultaneously
$q_{k,j} \in \{0,1\}$	$q_{k,j} = 1$ if and only if product category $j \in \mathcal{N}$ is assigned to both shelf segments k and $k+1 \in \mathcal{K} \setminus \{\beta_i : i \in \mathcal{B}\}$
$s_{k,j} \in [0,\infty)$	Amount of space allocated to product category $j \in \mathcal{N}$ along shelf segment $k \in \mathcal{K}$

for example if a product is on promotion. In the remainder of this paper γ_j is therefore interpreted as a reward term $\gamma_j^{'}$ defined on $[1, \infty)$.

The second term, $\sum_{k \in \mathcal{K}} \frac{f_k \cdot s_{k,j}}{c_k}$, is described by the authors as the visibility of a product j on the shelves. This is calculated as the sum of the attractiveness f_k of all shelf segments, proportionate to the amount of shelf space $s_{k,j}$ product j is assigned relative to the total capacity c_k of that shelf segment k. As an illustration, consider a product j with expected demand volume $v_j = 10$, profit margin $\rho_j = 1$, impulse purchase potential $\gamma_j = 1$, and that occupies 10 cm of shelf segment k and 5 cm of shelf segment $k+1$, which have a shelf attractiveness of 0.6 and 0.7, respectively, and a shelf capacity of 10 cm. Then the model formulation of Flamand et al. (2018) calculates for this product j a profit of $v_j \cdot \rho_j \cdot \gamma_j \cdot (\frac{f_k \cdot s_{k,j}}{c_k} + \frac{f_{k+1} \cdot s_{k+1,j}}{c_{k+1}}) = 10 \cdot 1 \cdot 1 \cdot (\frac{0.6 \cdot 10}{10} + \frac{0.8 \cdot 5}{10}) = 10$. Given the non-negative domains of all these parameters and variables, it becomes clear that the last additive term always increases the profit for product j if more shelf space $s_{k,j}$ is assigned to this product, which can be interpreted as a form of (b) space elastic demand effects. The differing values for $0 < f_k \leq 1$ represent the likelihood of a shelf segment to be visited by a customer, or shelf attractiveness in short, and can be seen as the aggregation of (d) vertical position-dependent demand and (e) horizontal position-dependent demand effects, as explained in Section 2. Given its domain, one can also interpret f_k as a penalty cost term for not assigning a product category to the most attractive shelf segment. To gain better insights of the demand for products, this paper consolidates all demand effects identified into a demand function d_j, which is elaborated on in the next subsection.

3.2 Enhancing the model

This subsection formulates possible enhancements of model APSA by incorporating the practically relevant considerations (a) — (g) into a demand function d_j^* and reformulating the objective function. Note that this is done solely on what is recommended methodologically in the existing literature. The feasibility of implementing these formulated enhancements into model APSA is assessed in Section 3.3. An overview of all the newly introduced notation is given in Table 3.2. Enhanced models are referred to with a suffix. In addition, both deterministic *det* and stochastic *sto* demand formulations are considered. For example, APSA-sto-(a)(b) is model APSA that considers stochastic demand and also incorporates (a) impulse purchasing potential and (b) space elastic demand effects.

$$\max \sum_{k\in\mathcal{K}} \sum_{j\in\mathcal{N}} \Phi_j \frac{f_k s_{k,j}}{c_k}, \text{ subject to} \tag{3.1a}$$

$$\sum_{i\in\mathcal{B}} x_{i,j} \leq 1, \quad \forall j \in \mathcal{N} \tag{3.1b}$$

$$\sum_{j\in\mathcal{N}} s_{k,j} \leq c_k, \quad \forall k \in \mathcal{K} \tag{3.1c}$$

$$l_j \sum_{i\in\mathcal{B}} x_{i,j} \leq \sum_{k\in\mathcal{K}} s_{k,j} \leq u_j \sum_{i\in\mathcal{B}} x_{i,j}, \quad \forall j \in \mathcal{N} \tag{3.1d}$$

$$\phi_j y_{k,j} \leq \min\{c_k, u_j\} y_{k,j}, \quad \forall j \in \mathcal{N}, k \in \mathcal{K} \tag{3.1e}$$

$$s_{k_2,j} \geq c_{k_2}(y_{k_1,j} + y_{k_3,j} - 1), \quad \forall j \in \mathcal{N}, i \in \mathcal{B}, k_1, k_2, k_3 \in \mathcal{K}_i \mid k_1 < k_2 < k_3 \tag{3.1f}$$

$$y_{k,j} \leq x_{i,j}, \quad \forall i \in \mathcal{B}, j \in \mathcal{N}, k \in \mathcal{K}_i \tag{3.1g}$$

$$x_{i,j} \leq \sum_{k\in\mathcal{K}_i} y_{k,j}, \quad \forall i \in \mathcal{B}, j \in \mathcal{N} \tag{3.1h}$$

$$q_{k,j} \geq y_{k,j} + y_{k+1,j} - 1, \quad \forall i \in \mathcal{B}, j \in \mathcal{N}, k \in \mathcal{K}_i \setminus \{\beta_i\} \tag{3.1i}$$

$$\sum_{j\in\mathcal{N}} q_{k,j} \leq 1, \quad \forall i \in \mathcal{B}, k \in \mathcal{K}_i \setminus \{\beta_i\} \tag{3.1j}$$

$$x_{i,j} + x_{i,j'} \leq 1, \quad \forall (j, j') \in \mathcal{L}, i \in \mathcal{B} \tag{3.1k}$$

$$x_{i,j} - x_{i,j'} = 0, \quad \forall (j, j') \in \mathcal{H}_1, i \in \mathcal{B} \tag{3.1l}$$

$$x_{i,j} \leq x_{i,j'}, \quad \forall (j, j') \in \mathcal{H}_2, i \in \mathcal{B} \tag{3.1m}$$

$$x_{i,j} - x_{i,j'} \leq 1 - z_{j,j'}, \quad \forall (j, j') \in \mathcal{H}_3, i \in \mathcal{B} \tag{3.1n}$$

$$x_{i,j} - x_{i,j'} \geq 1 - z_{j,j'}, \quad \forall (j, j') \in \mathcal{H}_3, i \in \mathcal{B} \tag{3.1o}$$

$$z_{j,j'} \leq \sum_{i\in\mathcal{B}} x_{i,j}, \quad \forall j,\ j' \in \mathcal{H}_3 \tag{3.1p}$$

$$z_{j,j'} \leq \sum_{i\in\mathcal{B}} x_{i,j'}, \quad \forall j,\ j' \in \mathcal{H}_3 \tag{3.1q}$$

$$z_{j,j'} \geq \sum_{i\in\mathcal{B}} x_{i,j} + \sum_{i\in\mathcal{B}} x_{i,j'} - 1, \quad \forall j,\ j' \in \mathcal{H}_3 \tag{3.1r}$$

$$\mathbf{x},\ \mathbf{y},\ \mathbf{z} \text{ binary}, \mathbf{s},\ \mathbf{q} \geq 0. \tag{3.1s}$$

To generalize the formulation and allow for demand effects, the objective function of model APSA (3.1a) is reformulated based on Hübner et al. (2016), which is in accordance with the Newsvendor model (Edgeworth, 1888). The Newsvendor model is a single-period profit optimization model and is outlined in for example Khouja (1999). The generalized new objective function of model APSA is given in Equation 3.2, where π_j represents the product-specific profit and is formulated with deterministic demand in Equation 3.3 and with stochastic demand in Equation 3.4. Here the indicator function is denoted by $\mathbb{1}_A$, which equals one if constraint A is satisfied and zero otherwise. When considering stochastic demand, $f_{d_j^*}$ is used to refer to the probability density function of the demand d_j^* for product j, where the superscript * denotes the same

Table 3.2

Additional notation used for the enhanced Mixed Integer Program (MIP) model variations of APSA.

Notation	Description
Sets	
$\mathcal{N}^+ \subseteq \mathcal{N}$	Set with all products selected in the assortment
$\mathcal{N}^- \equiv \mathcal{N} \setminus \mathcal{N}^+$	Set of all products not selected in the assortment
Parameters	
α_j	Base (latent) demand for product $j \in \mathcal{N}$
β_j	Space demand effect (elasticity)
$\beta_{j,j'}$	Cross space elasticity relationships
$\gamma_j \in [1, \infty)$	Reward term for additional sales due to unplanned purchases
$\theta^{OOA}_{j',j}$	Probability of buying product j in case product j' is not in the assortment
$\theta^{OOS}_{j',j}$	Probability of buying product j in case product j' is out of stock
cos_j	Unit cost of product j
rev_j	Revenue of product j
sal_j	Salvage value of product j
sho_j	Shortage cost of product j
Auxiliary variables	
$x_j = \sum_i^{\mathcal{B}} x_{i,j}$	Total number of shelves product j is assigned to
$y_j = \sum_k^{\mathcal{K}} y_{k,j}$	Total number of shelf segments product j is assigned to (number of facings)
$s_j = \sum_k^{\mathcal{K}} s_{k,j}$	Total shelf capacity assigned to product j
$\theta^{OOA}_{j'} = \sum_{j \neq j'}^{\mathcal{N}} \theta^{OOA}_{j',j}$	Fraction of demand $d_{j'}$ for delisted product j that can be substituted
$\theta^{OOS}_{j'} = \sum_{j \neq j'}^{\mathcal{N}} \theta^{OOS}_{j',j}$	Fraction of unsatisfied demand $d_{j'} - qty_{j'}$ for product j that can be substituted
$qty_{j'} = \left\lfloor \frac{s_{j'}}{\phi_{j'}} \right\rfloor$	Total available quantity in units of product j on all the shelves

sequence of demand effects suffixes. For example, for APSA-sto-(a)(b), the corresponding demand function is denoted as $d_j^{(a)(b)}$. For simplicity and without loss of generality, this thesis paper only considers normally distributed demand for all APSA-(sto) models, similar to Hübner and Schaal (2017).

$$\max \sum_j^{\mathcal{N}} \pi_j \tag{3.2}$$

$$\pi_j = -cos_j \cdot qty_j + rev_j \cdot \min\{qty_j, d_j^*\} + \mathbb{1}_{qty_j > d_j^*} \cdot sal_j \cdot (qty_j - d_j^*) - \mathbb{1}_{qty_j < d_j^*} \cdot sho_j \cdot (d_j^* - qty_j) \tag{3.3}$$

$$\begin{aligned} \pi_j = -cos_j \cdot qty_j + rev_j \cdot \int_0^{qty_j} h \cdot f_{d_j^*}\, dh + rev_j \cdot \int_{qty_j}^{\infty} qty_j \cdot f_{d_j^*}\, dh \\ +sal_j \cdot \int_0^{qty_j} (qty_j - h) f_{d_j^*}\, dh - sho_j \cdot \int_{qty_j}^{\infty} (h - qty_j) f_{d_j^*}\, dh \end{aligned} \tag{3.4}$$

In the remainder of this section, the demand function d_j^* is derived by including the practically relevant considerations (a) — (g) step-by-step. This is mainly done based on the methodology described in Hwang et al. (2005) and Hübner et al. (2020). As starting point, d_j is considered to be equal to the base demand α_j for product j, representing the demand of a product if it would be assigned to exactly one shelf segment k, excluding unplanned (impulse) purchases:

$$d_j = \alpha_j, \tag{3.5}$$

(a) impulse purchasing potential. To account for an increase in expected demand due to unplanned purchases, the terminology of Flamand et al. (2018) is slightly adjusted to be interpreted as a reward term instead of a penalty term, as mentioned before. This allows to reflect (temporary) increases in expected demand due to promotional activities, such as in-store banners, discount coupons, or external advertisement by the supplier itself. More specifically, incorporating the impulse purchase potential $\gamma_j^{'} \geq 1$ into the demand function gives

$$d_j^{(a)} = \alpha_j \cdot \gamma_j^{'}. \tag{3.6}$$

(b) space elastic demand. To consider space elastic demand effects, the common practice in the shelf space planning literature is to model the elasticity as a power function (Bianchi-Aguiar et al., 2020). This general relationship was first found by Hansen and Heinsbroek (1979). The main idea is that an increase in assigned shelf space for a product j, increases the expected demand for that product. Incorporating this into Equation 3.6 gives

$$d_j^{(a)(b)} = \alpha_j \cdot \gamma_j^{'} \cdot y_j^{\beta_j}, \tag{3.7}$$

where $y_j = \sum_{k \in \mathcal{K}} y_{k,j}$ represents the number of shelf segments product category j is assigned to equals, which is equivalent to the number of facings. The space elastic demand relationship is represented by β_j, and to calibrate observed demand with the observed space effect, one can calculate $\alpha_j = \frac{d_j}{\gamma_j^{'} \cdot y_j^{\beta_j}}$ as outlined in Hübner (2011). Note that the number of facings is the determining factor for space elastic demand, since only the product on the front row is visible to the customer and not the products stored behind, as described in Hübner et al. (2021). Also note that if a product j is not selected in the assortment, Equation 3.6 would mathematically result into $d_j^{(a)} = \alpha_j \cdot \gamma_j^{'} \cdot y_j^{\beta_j} = \alpha_j \cdot \gamma_j^{'} \cdot 0^{\beta_j}$, which is either zero or undefined, depending on the value of β_j. However, this is not realistic as it is likely that there is still demand for a product, even though it is not included in the assortment. Therefore if $y_j = 0$, the demand for product j is considered to be $d_j^{(a)} = \alpha_j$, as is similarly defined in Hübner et al. (2020). This holds for all demand functions d_j^* defined in this paper. Considering (b) space-elastic demand effects thus requires the estimation of an additional $|\mathcal{N}| - 1$ parameters.

(c) cross-space elastic demand. The general space elastic relationship formulated by Hansen and Heinsbroek is extended upon by Corstjens and Doyle (1981) to include cross-space effects. Their extension has been very influential and forms the foundation for many models that consider cross-space elastic demand (Bianchi-Aguiar et al., 2020). The demand function considering both (a) space elastic demand and (b) cross-space elastic demand is formulated as

$$d_j^{(a)(b)} = \alpha_j \cdot \gamma_j \cdot y_j^{\beta_j} \cdot \prod_{j'}^{\mathcal{N}^+} y_{j'}^{\beta_{j,j'}}, \tag{3.8}$$

where $\beta_{j,j'}$ represents the cross space elasticity relationship, which is positive if product category j' is complementary to j, negative if j' is a substitute to j (Hwang et al., 2005), and by definition zero if $j' = j$. The set $\mathcal{N}^+ \subseteq \mathcal{N}$ is defined as the set of product categories j' with $y_{j'} \geq 1,\ \forall j' \in \mathcal{N}$. This is necessary as mathematically $d_j^{(a)(b)} = 0$ if $y_{j'} = 0,\ \forall j' \in \mathcal{N}$. Enhancing the demand function with (c) cross-space elastic demand effects requires the estimation of an additional $|\mathcal{N}| \cdot |\mathcal{N}|$ parameters. An advantage of choosing for this general formulation of Corstjens and Doyle is its forward compatibility with extensions, such as the ones described in Irion et al. (2012). Another important note to make is that cross space elasticity is often not considered in practice (Kök et al., 2015). The main reasons being that it is practically infeasible to consider all cross-elasticity terms if $|\mathcal{N}|$ is large and even if the effect is present, it is often negligibly small (Bianchi-Aguiar et al., 2020).

(d) vertical position-dependent demand and *(e) horizontal position-dependent demand.* Hwang et al. (2005) consider location effects by multiplying the model of Corstjens and Doyle with an additional term $attractiveness_k \geq 1$, representing the difference in attractiveness of shelf segments given their position. Albeit the model formulation of Flamand et al. (2018) does not explicitly mention these demand effects, they are implicitly aggregated into the shelf attractiveness $0 < f_k \leq 1$, as stated before. However, an important difference is the defined range of both parameters. The former is defined on $[1, \infty)$, implying that the base demand α_j represents the demand for a product category j if it would be assigned to the least attractive shelf segment k. In this case, the location effects can then be interpreted as an additional increment on the base demand if a product category is assigned to a relatively more attractive shelf segment. Flamand et al. (2018) on the other hand, define their shelf attractiveness f_k on $(0, 1]$, implying that the base demand α_j should represent the demand for a product category j if it would be assigned to the most attractive shelf segment k. In this case, the location effects can thus be interpreted as a penalty term for assigning a product category to a less than most attractive shelf segment. In this paper the definition of Flamand et al. (2018) for f_k is used, thus regarding it as a penalty term. However, an adjustment is required, since Flamand et al. (2018) formulate their shelf attractiveness f_k in an aggregated form $\sum_{k \in \mathcal{K}} \frac{f_k \cdot s_{k,j}}{c_k}$ with space elastic demand effects, as elaborated on before, and this paper wishes to isolate these effects in the demand function. Therefore, the location effect for a product category j is assumed to be the weighted average value of f_k of the shelf segments k it is assigned to, as is similarly done in Hwang et al. (2005). To illustrate this, consider a product j assigned to shelf segments k and $k+1$ with shelf space $s_{k,j} = 3$ and $s_{k+1,j} = 6$, and shelf attractiveness $f_k = 0.5$ and $f_{k+1} = 0.8$. Then the weighted average value of f_k for product j is assumed to be $\frac{s_{k,j} \cdot f_k + s_{k+1,j} \cdot f_{k+1}}{s_{k,j} + s_{k+1,j}} = \frac{3 \cdot 0.5 + 6 \cdot 0.8}{3+6} = 0.7$, where the denominator is denoted by $s_j = \sum_{k \in \mathcal{K}} s_{k,j}$ representing the total shelf space allocated to product j. Incorporating this in the demand function in Equation 3.8 gives

$$d_j^{(a)(b)(c)(d)(e)} = \alpha_j \cdot \gamma_j^{'} \cdot y_j^{\beta_j} \cdot \prod_{j'}^{\mathcal{N}^+} y_{j'}^{\beta_{j,j'}} \cdot \frac{\sum_{k \in \mathcal{K}} s_{k,j} f_k}{s_j}. \tag{3.9}$$

(f) out-of-assortment substitution. Similarly to Hübner et al. (2016), this paper considers the additional demand for a product $j \in \mathcal{N}^+$ following from the substitution of delisted products $j' \in \mathcal{N}^-$, where $\mathcal{N}^+$ denotes the set of product categories selected in the assortment, $\mathcal{N}^-$ denotes the set of product categories

not selected in the assortment and $\mathcal{N}^+ \cup \mathcal{N}^- = \mathcal{N}$. For simplicity, it is assumed there is one round of substitution and aggregated consumer demand is modelled with an exogenous demand model, which is common across most assortment literature as is also described in Kök et al. (2015). This implies that when a customer wants to buy a product $j' \in \mathcal{N}^-$ that is out of assortment, there is a probability of $\theta^{OOA}_{j',j}$ that it will buy product $j \in \mathcal{N}^+$ instead, where obviously $\theta^{OOA}_{j',j} = 0$ for $j' = j$. Considering all products $j \in \mathcal{N}$ gives $\theta^{OOA}_{j'} = \sum_j^{\mathcal{N}} \theta^{OOA}_{j',j}$, representing the fraction of demand $d_{j'}$ for a delisted product j' that is expected to be substituted by some other product $j \in \mathcal{N}$, assuming that product j is in the assortment and not out of stock either. Note that the demand $d_{j'}$ for a delisted product $j' \in \mathcal{N}^-$ is equal to the base demand $\alpha_{j'}$ as $y_{j'} = 0$. This substitution process is modeled similar to Smith and Agrawal (2000), as described in Hübner et al. (2016). The additional demand resulting from out-of-assortment substitution is formulated as Equation 3.10 and adding it to Equation 3.9 results in Equation 3.11. Incorporation of (f) out-of-assortment substitution effects into the demand function thus requires the estimation of an additional $|\mathcal{N}| \cdot |\mathcal{N}|$ parameters.

$$d_j^{(f)} = \sum_{j'}^{\mathcal{N}^-} \alpha_{j'} \cdot \theta^{OOA}_{j',j} \tag{3.10}$$

$$d_j^{(a)(b)(c)(d)(e)(f)} = \alpha_j \cdot \gamma_j \cdot y_j^{\beta_j} \cdot \prod_{j'}^{\mathcal{N}^+} y_{j'}^{\beta_{j,j'}} \cdot \frac{\sum_{k \in \mathcal{K}} s_{k,j} f_k}{s_j} + \sum_{j'}^{\mathcal{N}^-} \alpha_{j'} \cdot \theta^{OOA}_{j',j}, \tag{3.11}$$

(g) out-of-stock substitution. Hübner et al. (2016) also provide a model formulation for incorporating demand effects resulting from the substitution of products that are temporarily out of stock. An empirical study of Campo et al. (2004) indicates that retailer losses are significantly smaller in case of a temporary product unavailability than those in the case of a permanent unavailability. This may thus be an indication of out-of-assortment substitution to be significantly smaller than out-of-stock substitution, implying the importance of distinguishing between the two when modelling expected demand. In the case of out-of-stock substitution, there is an additional demand $d_j^{(g)}$ for a product $j \in \mathcal{N}^+$ following from the substitution of another product $j' \in \mathcal{N}^+$ that is temporarily not available on the shelves. It is assumed that this type of substitution only takes place when the expected demand $d^*_{j'}$ exceeds the available quantity $qty_{j'} = \left\lfloor \frac{s_{j'}}{\phi_{j'}} \right\rfloor$ on the shelves, where $s_{j'} = \sum_k^{\mathcal{K}} s_{k,j'}$. If product j is in the assortment and not out-of-stock, $\theta^{OOS}_{j'} = \sum_{j \neq j'}^{\mathcal{N}} \theta^{OOS}_{j',j}$ is the fraction of the unsatisfied demand $d^*_{j'} - qty_{j'}$ that is expected to be substituted, where $\theta^{OOS}_{j',j}$ is the probability of a customer substituting product j' with product $j \in \mathcal{N}$, for which again obviously $\theta^{OOS}_{j',j} = 0$ if $j' = j$. The additional demand resulting from out-of-stock substitution is formulated as Equation 3.12 and adding it to Equation 3.11 results in Equation 3.13. Incorporation of (g) out-of-stock substitution effects into the demand function thus requires the estimation of an additional $|\mathcal{N}| \cdot |\mathcal{N}|$ parameters. Again, note that for all demand functions in this section $d^*_j = \alpha_j$ if $y_j = 0$, as motivated earlier.

$$d_j^{(g)} = \sum_{j'}^{\mathcal{N}^+} [(d_{j'} - qty_{j'}) | d_{j'} > qty_{j'}] \cdot \theta^{OOS}_{j',j} \tag{3.12}$$

$$d_j^{(a)(b)(c)(d)(e)(f)(g)} = \alpha_j \cdot \gamma_j^{'} \cdot y_j^{\beta_j} \cdot \prod_{j'}^{\mathcal{N}^+} y_{j'}^{\beta_{j,j'}} \cdot \frac{\sum_{k \in \mathcal{K}} s_{k,j} f_k}{s_j} + \sum_{j'}^{\mathcal{N}^-} \alpha_{j'} \cdot \theta_{j',j}^{OOA} + \sum_{j'}^{\mathcal{N}^+} [(d_{j'} - qty_{j'}) | d_{j'} > qty_{j'}] \cdot \theta_{j',j}^{OOS} \tag{3.13}$$

3.3 Feasibility of MIP

This subsection assesses which of the considered practically relevant model enhancements are feasible to be implemented in the Mixed Integer Programming (MIP) model of Flamand et al. (2018). First, the new objective function with deterministic demand is evaluated (Equation 3.3), after which the one with stochastic demand (Equation 3.4), and finally the feasibility of implementing all the variants of the demand function d_j^* itself is considered.

The deterministic version of the profit function (Equation 3.3) is feasible to implement into APSA. For this, the logical operators $\min\{qty_j, d_j^*\}$, $\mathbb{1}_{qty_j > d_j^*}$, and $\mathbb{1}_{qty_j < d_j^*}$ need to be written more explicitly. Two of the most common ways to achieve this are using the so-called bigM formulation and disjunctive programming formulation (Bonami et al., 2015). In this thesis paper the bigM formulation is used, mainly because of its straight-forwardness and taking into account the time constraints for writing this paper. However, it is important to note that using disjunctive programming techniques instead would most probable result in (computationally) more promising results, as is demonstrated in the numerical study of Bonami et al. (2015). Also note that the feasibility of the implementation depends on the choice for demand function d_j^*, but this is assessed in isolation later in this subsection. Using bigM formulation, the logical operators in the objective function and their corresponding constraints are given in Table 3.3.

The MIP implementation of the stochastic version of the profit function (Equation 3.4) is considered to be out of the scope for this paper. The presence of the integrals would namely result in a non-linear objective function, thereby classifying model APSA as a Mixed Integer Nonlinear Program (MINLP) and significantly increasing the complexity of the model. In addition, the profit function can be shown to be non-convex by following the same approach as Khouja (1999) to use Leibniz's rule to obtain first and second order derivatives. Burer and Letchford (2012) note that although both the convex and non-convex MINLPs are NP-hard in general, the non-convex case is much harder to solve in both theory and practice. An additional problem for the non-convex case is that the continuous relaxation of the problem would not provide a valid bound anymore, as the solution might get stuck in a local optimum. A suggestion for further research is to rewrite the integrals using Taylor expansion, such that the problem becomes a MINLP with polynomials. Burer and Letchford (2012) note that these expressions can then be reformulated as quadratic MINLPs by using additional constraints and variables. To deal with the non-convexity, convex relaxations can be applied for which popular approaches are based on semidefinite programming (SDP) and the Reformulation-Linearization Technique (RLT), as outlined in Burer and Letchford (2012). They also note that for quadratic MINLPs, a combination of SDP and RLT is feasible and would result in an even better convex relaxation.

To determine which versions of the demand function d_j^* to consider in the computational study, a similar motivation as above applies. So in essence, this thesis paper only evaluates the incorporation of demand functions d_j^* for which the resulting objective function would not contain any polynomials. This effectively means that next to the base demand function d_j given in Equation 3.5, only the enhancements (a) impulse purchasing potential, (f) out-of-assortment substitution, and a combination of thereof are considered. Moreover, similar to the logical operators in the deterministic version of the profit function, the reformulation of the model enhancements is also given in Table 3.3.

Table 3.3

Overview of reformulated logical operators to implement into the MIPs. The constraints take into account the maximization of the objective function.

Original	Reformulated	Constraints ($\forall j, j' \in \mathcal{N}$)
$qty_j = \left\lfloor \frac{s_j}{\phi_j} \right\rfloor$	qty_j	$qty_j \leq \frac{\sum_k^{\mathcal{K}} s_{k,j}}{\phi_j}$ $qty_j \in \mathbb{N}$
$\min\{qty_j, d_j^*\}$	$I_{1,j}$	$qty_j \geq I_{1,j}$ $d_j^* \geq I_{1,j}$
$\mathbb{1}_{qty_j > d_j^*} \cdot (qty_j - d_j^*)$	$I_{2,j}$	$0 \leq I_{2,j} \leq M(1 - b_{I2,j})$ $qty_j - d_j^* \leq I_{2,j} \leq qty_j - d_j^* + M \cdot b_{I2,j}$ $b_{I2,j} \in \{0, 1\}$
$\mathbb{1}_{qty_j < d_j^*} \cdot (d_j^* - qty_j)$	$I_{3,j}$	$0 \leq I_{3,j} \leq M(1 - b_{I3,j})$ $d_j^* - qty_j \leq I_{3,j} + I_{4,j} \leq d_j^* - qty_j + M \cdot b_{I3,j}$ $0 \leq I_{4,j} \leq M(1 - b_{I4,j})$ $b_{I4,j} \leq qty_j \leq M \cdot b_{I4,j}$ $b_{I3,j},\ b_{I4,j} \in \{0, 1\}$
$\sum_{j'}^{\mathcal{N}^-} \alpha_{j'} \cdot \theta_{j',j}^{OOA}$	$\sum_{j'}^{\mathcal{N}} (1 - I_{5,j'}) \alpha_{j'} \cdot \theta_{j',j}^{OOA}$	$\sum_i^{\mathcal{B}} x_{i,j'} \leq I_{5,j'}$ $I_{5,j'} \in \{0, 1\}$

Note. The value of M is defined such that $M \geq \max(|qty_j - d_j^*|, d_j^*),\ \forall j \in \mathcal{N}$. Variables $I_{4,j}$ and $b_{I4,j}$ make sure that there are no shortage costs when there is no quantity put on the shelves.

3.4 Computational study

For the computational study, this thesis paper mainly adheres to the structure of Flamand et al. (2018). To gain a better insight into the computational burden of the considered model enhancements, first the main results of Flamand et al. (2018) are replicated and interpreted. This allows for a fair comparison between the original APSA and the ones suggested and considered in this paper, in the sense that all the results are analyzed in the same modeling and simulation environment. Next to the differences in hardware used in this study, the differences in software are likely to be the most contributing factor in explaining the dissimilarity between some of the results of Flamand et al. (2018) and the replication done as part of this paper. For instance, it is noteworthy that they use CPLEX 12.6, whereas the present study uses CPLEX 20.1.0, which contains improvements resulting in a better performance of solving MIPs. Another difference is that the replication is done in Python for both the CPLEX solver and the heuristic approach, whereas Flamand et al. use C++, which is generally known to be more efficient in terms of computations.

After the replication part, the models in Table 3.5 are analyzed, but then with less sets and instances due to time constraints. The testbed for these models consists of two sets with only two random instances each, respectively including 30 and 50 shelves, and 240 and 400 products. A time limit of 3,600 CPU seconds to solve a problem instance is imposed for both CPLEX and the heuristic approach. The data generation scheme for all the test instances is based on the existing literature, and is given in Table 3.4. The % optimality gap for the heuristic is as defined in Flamand et al. (2018) as $\frac{BUB-Incumbent}{BUB} \cdot 100$, where BUB is the best upper bound found by CPLEX at termination and $Incumbent$ is the best objective value found at termination for the heuristic approach. In Flamand et al. (2018), the heuristic terminates when the time limit is reached, the relative gap between the LP-based upper bound and the incumbent is less than or equal to $\epsilon = 0.5\%$, or when the algorithm repeats itself several times without any improvement in the incumbent. This thesis paper uses the same termination rules for the replication part. However, for the enhanced APSA models the LP-based upper bound needs to be replaced, as in all the considered cases this would give an upper bound that is several orders of magnitudes larger than the best upper bound found by CPLEX at termination. It is therefore deemed impossible for the heuristic to reach an optimality gap of $\epsilon = 0.5\%$ when using the LP-based upper bound. As an alternative upper bound, this paper therefore uses the best bound found by the CPLEX solver when it has found five feasible MIP solutions. Due to time limitations, this number of five is simply a guesstimate based on a limited number of problem instances. Consequently, it is important to note that the performance of the heuristic could potentially be improved by fine-tuning this number by for example terminating CPLEX when the five solutions have been found and/or when no significant improvement in the best bound has been made after a certain number of nodes visited. For completeness, the algorithms for the heuristic approach of Flamand et al. (2018), including the adjusted step for determining the upper bound, are given in Appendix B.

As mentioned in the previous subsection, the computational study evaluates the deterministic APSA models with base demand function d_j and model enhancements (a) and (f). For comparability purposes, also a new model APSA-(Flam) is defined as the analogue to the original APSA model given in Equations (3.1a) — (3.1s), but then with the profit of each product category Φ_j disaggregated to adhere to the deterministic newsvendor model profit function (Equation 3.3), without taking into account the terms with salvage values sal_j and shortage costs sho_j, resulting in Equation 3.14. Furthermore, all components that influence demand are consolidated into the demand function $d_j^{(Flam)}$ (Equation 3.15), which accounts for (a) impulse purchasing potential and an aggregated form of (b) space elastic, (d) vertical position-dependent, and (e) horizontal position-dependent demand effects, as described in Section 2. Note again that this aggregated form sums the shelf attractiveness f_k proportionally to the amount of shelf space $s_{k,j}$ product j occupies, with respect to the the overall capacity c_k of that shelf. The space elastic demand effect relationship is therefore modeled differently than the common practice mentioned in Section 3.2. Finally, note that when product j is not

Table 3.4

Data generation scheme for all the parameters for model APSA and its proposed enhanced versions.

Parameter	Data generation description	Based on
α_j	For deterministic demand equals μ_j, which is randomly generated from the uniform distribution $U(0.2, 13)$. For stochastic demand, follows normal distribution $N(\mu_j, \sigma_j)$, where $\sigma_j = CV_j \cdot \mu_j$, and CV_j is the coefficient of variation, randomly generated from uniform distribution $U(0.01, 0.40)$. The generated value is rounded to two decimal places	Hübner et al. (2020), Hübner et al. (2021)
β_j	Equals 0.17 for all products j for simplicity	Eisend (2014)
$\beta_{j,j'}$	Equals -0.016 for all $j \neq j'$ and zero for $j = j'$ for simplicity	Eisend (2014)
γ_j	Equals 1.46 for all products j for simplicity	Inman et al. (2009)
$\theta^{OOA}_{j',j}$	Equals $\frac{\alpha_j}{\sum_{n \neq j'}^{N} \alpha_n}$ (market share-based substitutions), rounded to four decimal places, and zero for $j' = j$	Çömez-Dolgan et al. (2021)
$\theta^{OOS}_{j',j}$	Equals $\theta^{OOA}_{j',j}$ for simplicity	Hübner et al. (2020)
rev_j	Randomly generated from $U(20, 25)$ with two decimal places	Hübner et al. (2020)
cos_j	Randomly generated from $U(4, 9)$ with two decimal places	Hübner et al. (2020)
sal_j	Randomly generated from $U(4, cos_j)$ with two decimal places	Hübner et al. (2020)
sho_j	Randomly generated from $U(1, 3)$ with two decimal places	Hübner et al. (2020)
C_i	Equals 18 and each shelf has three segments	Flamand et al. (2018)
c_k	Equals 6 for all shelf segments k	Flamand et al. (2018)
l_j	Randomly generated from $U(1, \frac{C_i}{6})$, and is rounded to the nearest integer	Flamand et al. (2018)
u_j	Randomly generated from $U(l_j, \frac{C_i}{3})$, and is rounded to the nearest integer	Flamand et al. (2018)
f_k	Randomly generated such that all shelves $i \in \mathcal{B}$ are evenly distributed to have the same $t \in \{5\%, 25\%, 45\%, 65\%, 85\%\}$, which is used to generate for each of those shelves the attractiveness of the middle shelf segment randomly from $U(t, t + 0.05)$ and for the end segments randomly from $U(t + 0.06, t + 0.1)$, rounded to two decimal places	Flamand et al. (2018)
ϕ_j	Equals 0.1 units for all j	Flamand et al. (2018)

selected in the assortment its demand $d^*_j = \alpha_j$, as motivated in Section 3.2 as well.

$$\pi_j^{(Flam)} = -cos_j \cdot qty_j + rev_j \cdot \min\{qty_j, d^*_j\} \tag{3.14}$$

$$d_j^{(Flam)} = \alpha_j \cdot \gamma_j^{'} \cdot \sum_{k \in \mathcal{K}} \frac{s_{k,j} f_k}{c_k} \tag{3.15}$$

$$d_j^{(Flam)(f)} = \alpha_j \cdot \gamma_j^{'} \cdot \sum_{k \in \mathcal{K}} \frac{s_{k,j} f_k}{c_k} + \sum_{j'}^{\mathcal{N}^-} \alpha_{j'} \cdot \theta^{OOA}_{j',j} \tag{3.16}$$

To assess the practical relevance and importance of model enhancements (a), (f), and the aggregated enhancement of (a), (d) and (e), this study looks at the differences in computational time, objective value, and changes in selected assortment. This joint comparison addresses the third sub-question defined in the introduction of this paper. To define the differences in objective values, model APSA-det-(Flam)(f) is used as a benchmark, which refers to model APSA with the objective function in Equation 3.2, deterministic profit function in Equation 3.3, and demand function given in Equation 3.16. To compare with the benchmark model, this thesis paper considers the calculation of $\Delta\text{Profit}(\%) = \frac{Incumbent_{adjusted} - Incumbent}{Incumbent} \cdot 100$, where $Incumbent_{adjusted}$ denotes the objective value resulting from plugging the solution at termination into the objective function of the benchmark model. Since the objective function is the profit function, one can interpret ΔProfit as the expected surprise in profit by assuming that the objective function represents the expected profit by the retailer, and the objective function of the benchmark model represents the true realized (underlying) profit. For example, $Incumbent_{adjusted}$ for model APSA-det-(Flam) is calculated by first solving model APSA-det-(Flam) and then plugging the solution found at termination into the objective function of APSA-det-(Flam)(f). The difference between $Incumbent_{adjusted}$ and $Incumbent$ then reflects the overestimation of expected profits as a result of not considering (f) out-of-assortment substitution effects when determining the assortment and shelf space allocation. This comparative approach is similar to Yücel et al. (2009). The change in selected assortment is defined as the percentage of products whose assigned shelf space s_j changed compared to the best solution found at termination for the benchmark model, which is similar to the analysis approach of demand effects in Hübner and Schaal (2017). The results are given in Section 4 and are produced using the Python API of IBM ILOG CPLEX Optimization Studio 20.1.0 on a Windows 10 64 Bit machine with 16 gigabyte RAM and an Intel(R) Core(TM) i7-10510U CPU with 1.80 gigahertz.

Table 3.5

Definition of considered models and overview of the enhancements they include.

Model	Demand Type	Salvage Value	Shortage Cost	Model Enhancements (a)	(b)(d)(e)	(f)
APSA-(Flam)	det			✓	✓	
APSA-det-(Flam)	det	✓	✓	✓	✓	
APSA-det-(Flam)(f)	det	✓	✓	✓	✓	✓
APSA-det-(a)(f)	det	✓	✓	✓		✓

Note. The model enhancements are (a) impulse purchasing potential, (b) space elastic demand, (d) vertical position-dependent demand, (e) horizontal position-dependent demand, and (f) out-of-assortment substitution.

4 Results

This section outlines the results from the numerical study used to address the third sub-question, focusing on the computational burden and the effect on profits when model formulation APSA is enhanced. First, the main results of the computational study of Flamand et al. (2018) are reproduced and compared. Afterwards, the numerical results for the enhanced APSA models are considered. An important annotation to make is that all the results from the computational study are tracked in real time elapsed, while for an unbiased comparison the use of CPU seconds would have been better. However, due to time limitations it was not feasible to reproduce all the results using CPU seconds as a measure of time. For simplicity, time and CPU time are used interchangeably in the remainder of this paper. Only when the interpretation of the results is expected to be significantly biased, it is explicitly mentioned that a possible explanation for the non-intuitiveness is the use of wall-clock time instead of CPU time.

4.1 Replication of computational study in Flamand et al. (2018)

Overall, the main conclusions in Flamand et al. (2018) also hold for the replication done as part of this paper. The complete tables for all the replications are given in Appendix A.1 — A.4. An extract of the results for the CPLEX solver and the heuristic with neighborhood size $\tau \in \{2, 3, 4\}$ and optimality gap target $\epsilon = 0.5\%$ is given in Table 4.1, which shows that CPLEX never terminates within the preset time limit of 3,600 seconds, whereas the heuristic approach with $\tau = 3$ and $\tau = 4$ successfully manages to find a feasible solution with an % optimality gap satisfying the preset upper bound of $\epsilon = 0.5\%$ for all test instances. Furthermore, it does so while on average only using about 3.3% of the preset time limit for Set 1, up to about 28.4% for Set 5. Also interesting to note is that for Flamand et al., these numbers were around 1.8% and 55.4%, respectively. This improvement in time savings for larger problem sizes is most likely explained by the use of CPLEX version 20.1.0 instead of CPLEX 12.6. From Table 4.1 it can also be seen that CPLEX sometimes constructs a solution with lower optimality gap than the heuristic for the smallest data set, albeit requiring significantly more time, which is similar to Flamand et al. (2018). The practical relevance of the heuristic approach becomes especially apparent for larger problem instances, as then the heuristic consistently constructs high quality solutions in terms of the optimality gap, while the solutions found by CPLEX deteriorate. Zooming into the results of the heuristic with $\tau = 4$, an interesting dissimilarity with Flamand et al. is the lower relative increase in computational time required for the heuristic when increasing the problem size. For example, when comparing the results for test set 3 with test set 5 in Flamand et al. (2018), the average computational time required for the heuristic to terminate increases with about 750%. In contrast, this increase in CPU seconds is only about 260% for the replicated results. Also noteworthy is the difference in number of explored branch&bound/cuts nodes by the CPLEX solver. For problem set 1, the number of nodes explored is about 81% lower compared Flamand et al., and this difference seems to change signs as the problem size grows. For instance, for problem set 5 the average number of nodes visited is about 466% higher than in Flamand et al. for the same problem size. This difference is most likely explained by the use of different CPLEX versions. Another dissimilarity can be found when comparing the optimality gaps corresponding to the CPLEX solver. While this paper uses a newer version of CPLEX with improvements in the performance of solving MIPs, the optimality gaps in the replication are consistently lower than the ones found in the original results for almost all problem instances. An explanation for this inferior performance might be the implementation in Python instead of C++, as mentioned before. This would effectively imply that the difference in relative increase in computational time needed for the heuristic approach when the problem size grows, is even more beneficial than the replication in this paper shows. It is namely expected that this relative increase in CPU seconds is even lower when implementing the heuristic in C++ and running it on a machine with similar specs as the authors. This also fortifies the demonstration of the increasing practical relevance and importance of heuristic solution approaches when solving MIPs, as finding solutions for large problem instances and/or complex mathematical programs becomes more feasible.

Table 4.1

Extract of Table A.2 in Appendix A.2, showing the performance of the CPLEX solver and the effect of the neighborhood size τ on the performance of the heuristic.

Set	($\|\mathcal{B}\|$,$\|\mathcal{N}\|$)	Inst.	CPLEX			Heuristic $\tau = 2$, $\epsilon = 0.5\%$		Heuristic $\tau = 3$, $\epsilon = 0.5\%$		Heuristic $\tau = 4$, $\epsilon = 0.5\%$	
			CPU(s)	*B&B/C*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*
Set 1	(30,240)	1	3,601	3,818	0.95	1,003	0.62	75	0.45	124	0.43
		2	3,601	4,703	0.43	1,003	0.56	51	0.48	53	0.47
		8	3,601	2,567	0.96	497	0.49	98	0.49	96	0.40
Set 2	(40,320)	1	3,602	2,969	0.99	1,014	0.54	131	0.48	155	0.47
		2	3,602	2,964	0.87	1,010	0.53	134	0.48	112	0.40
		8	3,601	2,166	2.28	1,005	0.54	108	0.48	278	0.46
Set 3	(50,400)	1	3,601	1,947	2.40	1,019	0.56	138	0.46	329	0.43
		2	3,602	2,299	1.69	1,000	0.58	189	0.48	478	0.45
		10	3,601	1,877	1.73	740	0.49	141	0.45	147	0.37
Set 4	(60,480)	1	3,601	3,534	2.77	1,002	0.62	275	0.50	315	0.50
		2	3,601	1,986	2.28	1,014	0.58	336	0.44	274	0.48
		3	3,602	1,806	1.80	1,010	0.76	315	0.47	641	0.50
Set 5	(100,800)	1	3,608	465	5.07	1,092	0.96	1,165	0.49	1,324	0.43
		2	3,606	1,384	3.04	1,060	0.90	1,096	0.49	852	0.47
		3	3,603	462	6.14	1,096	0.94	1,004	0.50	946	0.45

Note. Column *B&B/C* displays the number of explored branch&bound/cut nodes. The time limit for the heuristic with $\tau = 2$ is set to 1,000 CPU seconds.

Next, the replication of the sensitivity analysis with respect to the neighborhood size τ, and target optimality gap ϵ is compared. Similar to Flamand et al. (2018), the model with $\tau = 2$ performs badly, as can be seen in the extract in Table 4.1. The heuristic only manages to find a solution satisfying the optimality gap target of $\epsilon = 0.5\%$ for 6 out of all the 50 considered problem instances. Furthermore, for set 4 and set 5 the heuristic did not manage to find a single satisfactory solution, even when the time constraint of 1,000 CPU seconds was disregarded. Interestingly enough though, in the replicated study the heuristic did terminate successfully for one instance in set 3, whereas in the original study of Flamand et al. it did for none. For both $\tau = 3$ and $\tau = 4$, the heuristic successfully terminated with a solution that satisfies the optimality gap target for all instances, consistent with the results of Flamand et al.. The relation between the computational times of $\tau = 3$ and $\tau = 4$ on the other hand, are inconsistent with the original results. In the original study, the CPU seconds for $\tau = 4$ are lower than the ones for $\tau = 3$ for almost all problem instances, whereas in the replicated results the relation is reversed for problem sets 1 to 4. For problem set 5 the computational time for $\tau = 4$ is still on average approximately 10% lower than for $\tau = 3$. A possible explanation for this inconsistency is the non-optimal environment the code was run in for the heuristic with $\tau = 4$, as at the time of running part of the machine's CPU was occupied by another application, whereas for $\tau = 3$ this was not the case. As mentioned earlier, due to time limitations these test instances were not recomputed using CPU seconds instead of real time elapsed, but this is not expected to significantly affect the main conclusions of this thesis paper. Given that for the largest problem set the heuristic still shows the best results for $\tau = 4$, and the possible upward bias for the computational time for $\tau = 4$, only the heuristic with $\tau = 4$ is considered in the remainder of this computational study.

The results for the sensitivity analysis of differing values for $\epsilon \in \{0.5\%, 1.0\%, 1.5\%\}$ are largely similar to Flamand et al. (2018) as well. The only odd observation is that the CPU seconds for $\epsilon = 1.0\%$ seem to be significantly larger than the ones for $\epsilon = 1.5\%$, while both provide similar optimality gaps, as can be seen in Table 4.2. The most plausible explanation is again that the codes for those instances were executed in differing environments, as the machine was also used for other tasks while the program was running. Comparing the computational results for $\epsilon = 0.5\%$ with $\epsilon = 1.5\%$, an average increase of about 40% in CPU seconds is required in return for an average reduction of 0.42 percentage points in the optimality gap, which is similar to original study.

Finally, the replication of the results for the sensitivity analysis of the inclusion of five product pairs in each of the affinity sets ($\mathcal{L}$, $\mathcal{H}$, $\mathcal{H}_2$ and $\mathcal{H}_3$) is assessed. An extract of the results is given in Table 4.3. Similar to Flamand et al. (2018), it can be concluded that including five product category pairs in affinity sets $\mathcal{L}$, $\mathcal{H}_2$ and $\mathcal{H}_3$ does not seem to significantly affect the computational burden on average, while both the optimality gap and CPU seconds used do considerably increase when including affinity set $\mathcal{H}$ and all four sets simultaneously. More specifically, the optimality gap increased on average from 0.45% to 0.56% when including five product pairs in affinity set $\mathcal{H}$, and this increase was about 0.25 percentage points when including five product pairs in all of the affinity sets at the same time. Note that the differences are relatively larger for the smaller problem sizes, as the inclusion of five product pairs affects a relatively larger part of the assortment than for larger problem sizes. A possible outlier is the result for Set 1, instance 5, as it is the only result for which the inclusion of affinity set $\mathcal{L}$ does not construct a solution within the optimality gap target within the time limit. The most plausible explanation is again that sometimes the machine used for executing the code had to share part of its CPU.

4.2 Results for the enhanced APSA models

In this subsection the computational results for the enhanced APSA models denoted in Table 3.5 are outlined. Recall that model APSA-(Flam) represents the reformulation of the original APSA model to adhere to the Newsvendor model setting, excluding the terms with salvage values sal_j and shortage costs sho_j (Equation 3.14), and also includes the consolidation of all demand effects into the aggregate demand function $d_j^{(Flam)}$ (Equation 3.15). Model APSA-det-(Flam) is similar, but then with salvage values and shortage costs included in the profit function (Equation 3.3). The results in Table 4.4 show that both CPLEX and the heuristic approach did not manage to construct a feasible solution that satisfies the target optimality gap of $\epsilon =$ 0.5%. However, it can be concluded that the heuristic approach does a better job in general, as its average optimality gap is about 1.26 and 1.25 percentage points lower for models APSA-(Flam) and APSA-det-

Table 4.2

Extract of Table A.3 in Appendix A.3, showing the effect of target optimality gap ϵ on the performance of the heuristic.

Set	($\vert\mathcal{B}\vert$,$\vert\mathcal{N}\vert$)	Inst.	CPLEX			Heuristic $\tau = 4$, $\epsilon = 1.5\%$		Heuristic $\tau = 4$, $\epsilon = 1\%$		Heuristic $\tau = 4$, $\epsilon = 0.5\%$	
			CPU(s)	*B&B/C*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*
Set 1	(30,240)	1	3,601	3,818	0.95	43	0.93	111	0.93	124	0.43
		2	3,601	4,703	0.43	30	0.82	77	0.82	53	0.47
		3	3,601	2,619	1.07	57	0.84	114	0.84	105	0.47
Set 2	(40,320)	1	3,602	2,969	0.99	78	0.90	177	0.90	155	0.47
		2	3,602	2,964	0.87	73	0.84	149	0.84	112	0.40
		9	3,601	3,277	0.96	113	1.04	274	0.79	219	0.48
Set 3	(50,400)	1	3,601	1,947	2.40	84	0.79	198	0.79	329	0.43
		9	3,601	2,140	1.73	160	1.16	369	0.66	241	0.40
		10	3,601	1,877	1.73	236	0.69	204	0.69	147	0.37
Set 4	(60,480)	1	3,601	3,534	2.77	479	0.82	445	0.82	315	0.50
		2	3,601	1,986	2.28	226	0.86	362	0.86	274	0.48
		3	3,602	1,806	1.80	238	0.85	261	0.85	641	0.50
Set 5	(100,800)	1	3,608	465	5.07	829	0.88	743	0.88	1,324	0.43
		9	3,605	501	5.93	793	0.83	953	0.83	1,068	0.48
		10	3,608	1,121	6.42	618	0.86	821	0.86	970	0.50

(Flam), respectively. Also remarkable is that for problem instance 1 of Set 1, the heuristic did terminate before the preset time limit of 3,600 CPU seconds. Nevertheless, the constructed solution is still significantly larger than $\epsilon = 0.5\%$. The reason for this premature termination is that the heuristic stopped because its incumbent was within 0.5% of the upper bound found in the heuristic. For this it is important to note that the upper bound for the heuristic is constructed by taking the best bound found by the CPLEX solver that terminates after it has found five MIP solutions, whereas in Table 4.4 the *Gap*(%) is relative to the best upper bound found by the CPLEX solver that is only restricted by the time limit. This result is therefore an indication that further research is recommended for improving the methodology to determine the upper bound for the heuristic. Another notable observation is that the quality of the solutions in terms of the optimality gap is better for model APSA-det-(Flam) than for APSA-(Flam), even though the former includes more practically relevant considerations (salvage values and shortage costs). More specifically, on average the optimality gap is about 31% lower when solved with CPLEX and 40% lower when using the heuristic approach. This result is not intuitive as the former model has $6 \cdot |\mathcal{N}|$ additional variables ($I_{2,j}$, $I_{3,j}$, $I_{4,j}$, $b_{I2,j}$, $b_{I3,j}$, $b_{I4,j}$) and $8 \cdot |\mathcal{N}|$ additional constraints (see Table 3.3). A more plausible explanation is again that the results for APSA-(Flam) are upward biased as the corresponding code was executed in less optimal circumstances than the code for APSA-det-(Flam). Finally, when looking at the change in profits when plugging the solution of the model into the profit function of APSA-det-(Flam)(f), the results show that all profits become extremely negative. When considering APSA-det-(Flam), this would mean that the solution found by both CPLEX and the heuristic at termination, lead to an profit that is more than 200% lower than the retailer initially expected, just because it did not account for (f) out-of-assortment substitution effects. However, judging from the data generation scheme (Table 3.4), it is unlikely that the difference can

Table 4.3

Extract of Table A.4 in Appendix A.4, showing the effect of affinity relationships for the heuristic. For each relationship set $\in \{\mathcal{L}, \mathcal{H}, \mathcal{H}_2, \mathcal{H}_3\}$*, five product pairs are considered.*

Set	($\vert\mathcal{B}\vert$,$\vert\mathcal{N}\vert$)	Inst.	No Affinity		$\mathcal{L}$		$\mathcal{H}$		$\mathcal{H}_2$		$\mathcal{H}_3$		$(\mathcal{L}, \mathcal{H}, \mathcal{H}_2, \mathcal{H}_3)$	
			CPU(s)	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*
Set 1	(30,240)	3	105	0.47	94	0.47	122	0.85	95	0.48	131	0.75	122	1.50
		4	86	0.35	73	0.35	84	0.42	64	0.41	72	0.42	123	0.42
		5	314	0.48	219	1.18	132	0.75	122	0.60	165	1.36	131	1.32
Set 2	(40,320)	1	155	0.47	144	0.49	200	0.70	133	0.46	195	0.44	195	0.88
		9	219	0.48	185	0.63	195	1.02	198	0.66	185	0.60	199	0.82
		10	252	0.44	97	0.46	186	0.99	94	0.49	155	0.47	198	0.97
Set 3	(50,400)	1	329	0.43	159	0.43	207	0.49	175	0.45	183	0.48	261	0.45
		9	241	0.40	244	0.40	911	1.28	249	0.40	257	0.44	952	1.44
		10	147	0.37	154	0.37	302	0.57	120	0.47	157	0.50	302	0.70
Set 4	(60,480)	1	315	0.50	326	0.50	388	0.44	310	0.47	308	0.45	493	0.50
		2	274	0.48	274	0.50	337	0.46	304	0.47	274	0.44	360	0.45
		3	641	0.50	303	0.48	527	0.65	359	0.41	485	0.49	512	1.07
Set 5	(100,800)	1	1,324	0.43	1,109	0.42	1,292	0.46	1,096	0.46	1,117	0.50	1,388	0.47
		2	852	0.47	831	0.49	1,068	0.50	1,030	0.48	915	0.47	1,142	0.46
		3	946	0.45	1,126	0.44	1,223	0.48	1,033	0.42	1,506	0.47	2,454	0.49

Note. The time limits are 120, 180, 300, 480, and 2400 CPU seconds for test sets 1-5 respectively, as motivated in Flamand et al. (2018). However, this time limit is only checked after all shelves have been visited in the algorithm.

be this high. Further research into the determination of ΔProfit is therefore definitely recommended and the interpretation of this is therefore disregarded in the remainder of this paper.

Next, the effects of incorporating (f) out-of-assortment demand and the aggregated form of (b) space elastic, (d) vertical position-dependent, and (e) horizontal-position-dependent demand effects is assessed. Recall that the consideration of this aggregated form is referred to by adding (Flam) as an suffix to the model. Due to time constraints, only the first problem instance of Set 1 is considered, for which the results are given in Table 4.5. The most interesting observation is that for the full model APSA-det-(Flam)(f), both the CPLEX solver and the heuristic approach perform badly. To understand which of the practically relevant considerations is the bottleneck, the results of the different models and their components are compared. First, it is remarkable to note that the heuristic spent about 3,800 seconds for determining an upper bound. This effectively means that the CPLEX solver took more than one hour to find five feasible MIP solutions, as this is the stopping criterion for finding the upper bound. The heuristic finally terminated after about 134 minutes when it completed the Single Shelf Space (SSP) problem, which corresponds to Algorithm 1 in Flamand et al. (2018) and is also given in Appendix B. This algorithm essentially constructs a feasible initial solution by visiting all the shelves one by one, while determining for each shelf the optimal set of products out of the set of products that are not (yet) in the assortment. The fact that the heuristic still took relatively long for all the single shelf problems, indicates that the bottleneck is likely to be related to the constraints for the assortment. Furthermore, the results in Table 4.5 also show that the addition of demand effect (f) to APSA-det-(a) results in a model that is not readily solvable by the heuristic. The heuristic approach for this test instance was terminated manually after it had remained at a 0.25% gap for about four hours for one of the single shelf space problems. Meanwhile, the CPLEX solver seems still to be able to terminate successfully for APSA-det-(a)(f). In addition, comparing the results for APSA-det-(a) in Table 4.5 with the results in Table 4.4, shows that the inclusion of the aggregated form of demand effects (b)(d)(e) by going from model APSA-det-(a) to model APSA-det-(Flam), has a significant impact on the performance of both the CPLEX solver and the heuristic. Nevertheless, both approaches still manage to construct a solution within an average optimality gap of 5% and the heuristic approach outperforms CPLEX. Finally, comparing the number of variables and number of constraints of model APSA-det-(Flam)(f) and APSA-det-(Flam) shows that this is unlikely to be the cause for the difference in performance, as the latter model has about three times more variables constraints for Set 2, but is still reasonably solvable. To conclude, the bottleneck for the heuristic seems to be the incorporation of (f) out-of-assortment substitution effects and solving the SSP, whereas for CPLEX the real bottleneck seems to be the joint inclusion of demand effect (f) and the aggregated form of effects (b)(d)(e). Judging from the MIP formulation for APSA-det-(Flam)(f), it is most likely that the difficulty arises from the dependencies between the three decision variables $d_j^{(Flam)(f)}$, $s_{k,j}$, and $I_{5,j}$ in the demand function (Equation 3.16). Recall that the inclusion of $I_{5,j}$ is formulated in Table 3.3. For the heuristic approach, further research is deemed necessary, especially for determining a good quality upper bound and understanding why the single shelf problem is difficult to solve when incorporating the model enhancements considered in this paper.

Table 4.4

Effects on the computational time, optimality gap and change in profit, as a result of reformulating model APSA to adhere to the Newsvendor model setting and consolidating demand effects into an aggregate demand function.

Set	$(\|\mathcal{B}\|,\|\mathcal{N}\|)$	Inst.	APSA-(Flam)						APSA-det-(Flam)					
			CPLEX			Heuristic ($\tau = 4,\ \epsilon = 0.5\%$)			CPLEX			Heuristic ($\tau = 4,\ \epsilon = 0.5\%$)		
			CPU(s)	Gap(%)	ΔProfit	CPU(s)	Gap(%)	ΔProfit	CPU(s)	Gap(%)	ΔProfit	CPU(s)	Gap(%)	ΔProfit
Set 1	(30,240)	1	3,600	4.73	-233.09	3,602	3.63	-253.86	3,600	5.66	-226.83	2,661	2.49	-221.72
		2	3,600	4.62	-221.66	3,601	4.63	-246.65	3,601	3.67	-208.99	3,600	3.48	-218.37
Set 2	(50,400)	1	3,601	6.99	-230.51	3,601	4.07	-258.92	3,600	2.80	-221.86	3,601	2.41	-240.67
		2	3,601	5.62	-212.81	3,601	4.60	-247.88	3,600	2.98	-220.32	3,601	1.73	-218.40

Note. The value of ΔProfit represents the change in profits when plugging the solution of the model into the objective function (profit function) of model APSA-det-(Flam)(f).

Table 4.5

Effects on the computational time, optimality gap and change in profit, when including (f) out-of-assortment effects and the aggregated form (b) space elastic, (d) vertical position-dependent, and (e) horizontal-position-dependent demand effects. Only problem instance 1 of Set 1 is considered.

Model	CPLEX			Heuristic ($\tau = 4,\ \epsilon = 0.5\%$)			Number of	Number of
	CPU(s)	Gap(%)	ΔProfit(%)	CPU(s)	Gap(%)	ΔProfit(%)	Variables	Constraints
APSA-det-(a)	120	0.01	0	1,245	0.14	0	124,560	101,520
APSA-det-(a)(f)	387	0.01	-2.29	n/a	n/a	n/a	124,800	111,990
APSA-det-(Flam)(f)	3,600	114.22	-23.16	8,045	16.91	-101.47	124,800	111,990

Note. The value of ΔProfit represents the change in profits when plugging the solution of the model into the objective function (profit function) of model APSA-det-(Flam)(f). The heuristic for APSA-det-(Flam)(f) was terminated after finishing the single shelf problem (SSP) of Flamand et al. (2018).

5 Conclusion and Discussion

The goal of this thesis paper was to fortify the integration of assortment planning and shelf space allocation, and delivering a practical mixed integer programming model that considers demand effects to solve this intertwined problem. This was done by assessing the current state-of-the-art of assortment planning and shelf space allocation, and evaluating how the latest advancements in solving mathematical programs are keeping up with the practically relevant problem settings. In essence, this paper answered the research question below, followed by its corresponding sub-questions:

- *Which practically relevant considerations are feasible to incorporate when modelling an integrated retail assortment planning and store-wide shelf space allocation problem?*
- *What practically relevant considerations are addressed in the existing literature on assortment planning and shelf space allocation?*
- *Which of these considerations can be incorporated as an extension to the model of Flamand et al. (2018)?*
- *What is the effect of including these considerations on the optimality gap and computational time?*

To address the first sub-question, the existing academic literature has been explored in Section 2 by first considering the assortment selection and shelf space planning problems in isolation. This was mainly done by using the extensive literature reviews of Kök et al. (2015) and Bianchi-Aguiar et al. (2020), respectively. Furthermore, by searching for commonalities between existing integrated models, it has been concluded that almost all integrated models adhere their formulation to the Newsvendor model setting. Moreover, Hübner et al. (2016) noted that common limitations of existing (integrated) models are not considering demand effects, neglecting the cost side, and unsuitability for realistically large problem sizes. Taking into account that Flamand et al. (2018) show promising computational results for their model formulation and heuristic approach, but understate substitution effects and the cost side, their model has been used as starting point for defining a model that incorporates all practically relevant considerations this paper was able to identify in the literature review. The identified practically relevant considerations are the impulse purchase potential of products, space elastic demand effects, cross-space elasticity, position-dependent demand effects, substitution effects, and arrangement-dependent demand effects.

To address the second sub-question, this paper reformulated model APSA of Flamand et al. (2018) to adhere to the Newsvendor model setting in Section 3, as all of the integrated models found in the literature review make use of this model formulation. Consequently, this allowed the incorporation of all the demand effects considered in those existing integrated models, resulting in a state-of-the-art model formulation that combines the most common and best practices of multiple papers. However, it has also been concluded that more advancements in the field of solving mathematical programs are needed before this state-of-the-art model can be of practical use. The main limitations come from the fact that the inclusion of all identified practically relevant considerations would result in a non-convex Mixed Integer Non-linear Program (MINLP), which would need advanced linear and convex relaxation techniques to solve. This thesis paper therefore concluded that only the considerations that would not result in a polynomial expression in the MIP can be feasibly incorporated as an enhancement of the original model APSA formulation. More specifically, these are impulse purchase potential, out-of-assortment substitution effects, and an aggregated form of space elastic demand effects and position-dependent demand effects. Also note that a form of arrangement-dependent demand effects is already considered in Flamand et al. (2018), as their model allows the retailer to define which products should be put next to each other and which absolutely not.

To address the third and final sub-question, a data generation scheme has been constructed based on suggestions on what would be realistic in practice, which are given in the academic literature. Model APSA of Flamand et al. (2018), their heuristic approach, and the APSA models with practically relevant enhancements have been implemented using the Python API of IBM ILOG CPLEX Optimization Studio 20.1.0 and the results of the computational study are outlined in Section 4. Based on these results, the research question can be answered. Even though only a limited amount of test instances could be considered and the study mistakenly considered real time elapsed instead of CPU time, it can still be concluded from the results that the implementation of model APSA-det-(Flam)(f) is not practical. Note that this model corresponds to the model including all the enhancements that were deemed to be feasible. Furthermore, it has been demonstrated that the heuristic approach is not suitable for the incorporation of out-of-assortment substitution demand effects. However, incorporating salvage values and shortage costs by reformulating the original APSA model to adhere to the Newsvendor model setting is deemed feasible and the heuristic approach is shown to outperform CPLEX.

However, further research is suggested to improve the performance of the heuristic. More specifically, the initialization procedure can be improved and the methodology for determining an upper bound is deemed necessary as well. In addition, the computational study in this paper can be extended by simulating and considering more problem instances.

References

Bianchi-Aguiar, T., Hübner, A., Carravilla, M. A., and Oliveira, J. F. (2020). Retail shelf space planning problems: A comprehensive review and classification framework. *European Journal of Operational Research*.

Bonami, P., Lodi, A., Tramontani, A., and Wiese, S. (2015). On mathematical programming with indicator constraints. *Mathematical programming*, 151(1):191–223.

Burer, S. and Letchford, A. N. (2012). Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, 17(2):97–106.

Campo, K. and Gijsbrechts, E. (2005). Retail assortment, shelf and stockout management: issues, interplay and future challenges. *Applied Stochastic Models in Business and Industry*, 21(4-5):383–392.

Campo, K., Gijsbrechts, E., and Nisol, P. (2004). Dynamics in consumer response to product unavailability: do stock-out reactions signal response to permanent assortment reductions? *Journal of Business Research*, 57(8):834–843.

Çömez-Dolgan, N., Fescioglu-Unver, N., Cephe, E., and Şen, A. (2021). Capacitated strategic assortment planning under explicit demand substitution. *European Journal of Operational Research*.

Corstjens, M. and Doyle, P. (1981). A model for optimizing retail space allocations. *Management Science*, 27(7):822–833.

Dreze, X., Hoch, S. J., and Purk, M. E. (1994). Shelf management and space elasticity. *Journal of retailing*, 70(4):301–326.

Edgeworth, F. Y. (1888). The mathematical theory of banking. *Journal of the Royal Statistical Society*, 51(1):113–127.

Eisend, M. (2014). Shelf space elasticity: A meta-analysis. *Journal of Retailing*, 90(2):168–181.

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

Gecili, H. (2020). *Joint Shelf Design and Shelf Space Allocation Problem for Retailers*. PhD thesis, Wright State University.

Geismar, H. N., Dawande, M., Murthi, B., and Sriskandarajah, C. (2015). Maximizing revenue through two-dimensional shelf-space allocation. *Production and Operations Management*, 24(7):1148–1163.

Hansen, P. and Heinsbroek, H. (1979). Product selection and space allocation in supermarkets. *European journal of operational research*, 3(6):474–484.

Hübner, A. (2011). *Retail category management: decision support systems for assortment, shelf space, inventory and price planning*, volume 656. Springer Science & Business Media.

Hübner, A., Düsterhöft, T., and Ostermeier, M. (2021). Shelf space dimensioning and product allocation in retail stores. *European Journal of Operational Research*, 292(1):155–171.

Hübner, A., Kuhn, H., and Kühn, S. (2016). An efficient algorithm for capacitated assortment planning with stochastic demand and substitution. *European Journal of Operational Research*, 250(2):505–520.

Hübner, A. and Schaal, K. (2017). A shelf-space optimization model when demand is stochastic and space-elastic. *Omega*, 68:139–154.

Hübner, A., Schäfer, F., and Schaal, K. N. (2020). Maximizing profit via assortment and shelf-space optimization for two-dimensional shelves. *Production and Operations Management*, 29(3):547–570.

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

Hwang, H., Choi, B., and Lee, M.-J. (2005). A model for shelf space allocation and inventory control considering location and inventory level effects on demand. *International Journal of Production Economics*, 97(2):185–195.

Inman, J. J., Winer, R. S., and Ferraro, R. (2009). The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *Journal of Marketing*, 73(5):19–29.

Irion, J., Lu, J.-C., Al-Khayyal, F., and Tsao, Y.-C. (2012). A piecewise linearization framework for retail shelf space management models. *European Journal of Operational Research*, 222(1):122–136.

Khouja, M. (1999). The single-period (news-vendor) problem: literature review and suggestions for future research. *omega*, 27(5):537–553.

Kök, A. G. and Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55(6):1001–1021.

Kök, A. G., Fisher, M. L., and Vaidyanathan, R. (2015). Assortment planning: Review of literature and industry practice. *Retail supply chain management*, pages 175–236.

Rabbani, M., Salmanzadeh-Meydani, N., Farshbaf-Geranmayeh, A., and Fadakar-Gabalou, V. (2018). Profit maximizing through 3d shelf space allocation of 2d display orientation items with variable heights of the shelves. *Opsearch*, 55(2):337–360.

Smith, S. A. and Agrawal, N. (2000). Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48(1):50–64.

Van Nierop, E., Fok, D., and Franses, P. H. (2008). Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements. *Marketing Science*, 27(6):1065–1082.

Van Woensel, T., Van Donselaar, K., Broekmeulen, R., and Fransoo, J. (2007). Consumer responses to shelf out-of-stocks of perishable products. *International Journal of Physical Distribution & Logistics Management*.

Yücel, E., Karaesmen, F., Salman, F. S., and Türkay, M. (2009). Optimizing product assortment under customer-driven demand substitution. *European Journal of Operational Research*, 199(3):759–768.

A Appendix. Replication results

A.1 Replication of Table 3 in Flamand et al. (2018)

Table A.1

Performance of CPLEX vs. the heuristic.

Set	($\|\mathcal{B}\|$,$\|\mathcal{N}\|$)	Inst.	CPLEX			Heuristic	
						$\tau = 4$, $\epsilon = 0.5\%$	
			$CPU(s)$	$B\&B/C$	$Gap(\%)$	$CPU(s)$	$Gap(\%)$
Set 1	(30,240)	1	3,601	3,818	0.95	124	0.43
		2	3,601	4,703	0.43	53	0.47
		3	3,601	2,619	1.07	105	0.47
		4	3,601	2,649	0.48	86	0.35
		5	3,601	3,884	0.65	314	0.48
		6	3,601	2,811	0.71	57	0.48
		7	3,601	2,680	0.44	154	0.49
		8	3,601	2,567	0.96	96	0.40
		9	3,601	4,117	0.90	111	0.39
		10	3,601	2,873	0.69	95	0.36
Set 2	(40,320)	1	3,602	2,969	0.99	155	0.47
		2	3,602	2,964	0.87	112	0.40
		3	3,601	2,796	0.96	264	0.49
		4	3,601	2,491	0.69	253	0.47
		5	3,601	2,052	1.09	142	0.46
		6	3,601	1,951	1.16	270	0.38
		7	3,601	1,943	1.01	153	0.42
		8	3,601	2,166	2.28	278	0.46
		9	3,601	3,277	0.96	219	0.48
		10	3,601	1,764	1.06	252	0.44
Set 3	(50,400)	1	3,601	1,947	2.40	329	0.43
		2	3,602	2,299	1.69	478	0.45
		3	3,601	3,661	1.90	546	0.45
		4	3,601	2,051	1.83	179	0.48
		5	3,601	1,833	1.71	311	0.44
		6	3,602	1,954	1.66	208	0.45
		7	3,601	2,877	1.25	201	0.42
		8	3,601	1,933	1.66	195	0.48
		9	3,601	2,140	1.73	241	0.40
		10	3,601	1,877	1.73	147	0.37
Set 4	(60,480)	1	3,601	3,534	2.77	315	0.50
		2	3,601	1,986	2.28	274	0.48
		3	3,602	1,806	1.80	641	0.50
		4	3,601	1,829	1.11	607	0.45
		5	3,602	1,849	2.82	805	0.44
		6	3,601	1,736	1.75	464	0.48
		7	3,601	1,972	2.16	277	0.48
		8	3,602	1,966	1.51	311	0.43
		9	3,601	1,916	2.69	399	0.42
		10	3,601	1,903	2.81	383	0.40
Set 5	(100,800)	1	3,608	465	5.07	1,324	0.43
		2	3,606	1,384	3.04	852	0.47
		3	3,603	462	6.14	946	0.45
		4	3,607	499	5.78	1,142	0.40
		5	3,604	511	6.59	990	0.48
		6	3,603	292	6.64	1,040	0.49
		7	3,604	691	7.18	967	0.48
		8	3,605	398	5.64	936	0.48
		9	3,605	501	5.93	1,068	0.48
		10	3,608	1,121	6.42	970	0.50

Note. The number of explored branch&bound/cut nodes is given in column $B\&B/C$.

A.2 Replication of Table 4 in Flamand et al. (2018)

Table A.2

The effect of the neighborhood size τ on the performance of the heuristic.

Set	$(\lvert\mathcal{B}\rvert,\lvert\mathcal{N}\rvert)$	Inst.	CPLEX			Heuristic		Heuristic		Heuristic	
						$\tau = 2$, $\epsilon = 0.5\%$		$\tau = 3$, $\epsilon = 0.5\%$		$\tau = 4$, $\epsilon = 0.5\%$	
			CPU(s)	*B&B/C*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*
Set 1	(30,240)	1	3,601	3,818	0.95	1,003	0.62	75	0.45	124	0.43
		2	3,601	4,703	0.43	1,003	0.56	51	0.48	53	0.47
		3	3,601	2,619	1.07	1,002	0.54	109	0.49	105	0.47
		4	3,601	2,649	0.48	1,001	0.57	62	0.50	86	0.35
		5	3,601	3,884	0.65	1,000	0.76	71	0.41	314	0.48
		6	3,601	2,811	0.71	1,000	0.51	63	0.45	57	0.48
		7	3,601	2,680	0.44	1,001	0.61	117	0.49	154	0.49
		8	3,601	2,567	0.96	497	0.49	98	0.49	96	0.40
		9	3,601	4,117	0.90	1,006	0.54	109	0.50	111	0.39
		10	3,601	2,873	0.69	177	0.48	92	0.48	95	0.36
Set 2	(40,320)	1	3,602	2,969	0.99	1,014	0.54	131	0.48	155	0.47
		2	3,602	2,964	0.87	1,010	0.53	134	0.48	112	0.40
		3	3,601	2,796	0.96	1,015	0.80	218	0.47	264	0.49
		4	3,601	2,491	0.69	1,000	0.56	149	0.49	253	0.47
		5	3,601	2,052	1.09	1,005	0.67	150	0.48	142	0.46
		6	3,601	1,951	1.16	800	0.48	124	0.45	270	0.38
		7	3,601	1,943	1.01	875	0.48	135	0.49	153	0.42
		8	3,601	2,166	2.28	1,005	0.54	108	0.48	278	0.46
		9	3,601	3,277	0.96	1,015	0.58	227	0.43	219	0.48
		10	3,601	1,764	1.06	892	0.46	149	0.47	252	0.44
Set 3	(50,400)	1	3,601	1,947	2.40	1,019	0.56	138	0.46	329	0.43
		2	3,602	2,299	1.69	1,000	0.58	189	0.48	478	0.45
		3	3,601	3,661	1.90	1,014	0.76	255	0.46	546	0.45
		4	3,601	2,051	1.83	1,036	0.92	186	0.50	179	0.48
		5	3,601	1,833	1.71	1,023	0.88	239	0.49	311	0.44
		6	3,602	1,954	1.66	1,035	0.73	227	0.40	208	0.45
		7	3,601	2,877	1.25	1,055	0.68	249	0.41	201	0.42
		8	3,601	1,933	1.66	1,043	0.66	194	0.45	195	0.48
		9	3,601	2,140	1.73	1,021	0.91	217	0.50	241	0.40
		10	3,601	1,877	1.73	740	0.49	141	0.45	147	0.37
Set 4	(60,480)	1	3,601	3,534	2.77	1,002	0.62	275	0.50	315	0.50
		2	3,601	1,986	2.28	1,014	0.58	336	0.44	274	0.48
		3	3,602	1,806	1.80	1,010	0.76	315	0.47	641	0.50
		4	3,601	1,829	1.11	1,022	0.92	268	0.44	607	0.45
		5	3,602	1,849	2.82	1,007	0.88	302	0.50	805	0.44
		6	3,601	1,736	1.75	1,012	0.73	320	0.46	464	0.48
		7	3,601	1,972	2.16	1,010	0.68	330	0.43	277	0.48
		8	3,602	1,966	1.51	1,029	0.66	240	0.46	311	0.43
		9	3,601	1,916	2.69	1,034	0.91	377	0.44	399	0.42
		10	3,601	1,903	2.81	1,001	0.49	364	0.49	383	0.40
Set 5	(100,800)	1	3,608	465	5.07	1,092	0.96	1,165	0.49	1,324	0.43
		2	3,606	1,384	3.04	1,060	0.90	1,096	0.49	852	0.47
		3	3,603	462	6.14	1,096	0.94	1,004	0.50	946	0.45
		4	3,607	499	5.78	1,079	0.91	1,100	0.47	1,142	0.40
		5	3,604	511	6.59	1,130	0.91	1,135	0.50	990	0.48
		6	3,603	292	6.64	1,173	1.68	1,147	0.48	1,040	0.49
		7	3,604	691	7.18	1,099	1.85	1,234	0.49	967	0.48
		8	3,605	398	5.64	1,068	0.96	1,020	0.45	936	0.48
		9	3,605	501	5.93	1,107	1.00	1,212	0.49	1,068	0.48
		10	3,608	1,121	6.42	1,089	0.92	1,177	0.48	970	0.50

A.3 Replication of Table 5 in Flamand et al. (2018)

Table A.3

The effect of Target Optimality Gap ϵ on the performance of the heuristic.

Set	$(\lvert\mathcal{B}\rvert,\lvert\mathcal{N}\rvert)$	Inst.	CPLEX			Heuristic $\tau = 4$, $\epsilon = 1.5\%$		Heuristic $\tau = 4$, $\epsilon = 1\%$		Heuristic $\tau = 4$, $\epsilon = 0.5\%$	
			$CPU(s)$	$B\&B/C$	$Gap(\%)$	$CPU(s)$	$Gap(\%)$	$CPU(s)$	$Gap(\%)$	$CPU(s)$	$Gap(\%)$
Set 1	(30,240)	1	3,601	3,818	0.95	43	0.93	111	0.93	124	0.43
		2	3,601	4,703	0.43	30	0.82	77	0.82	53	0.47
		3	3,601	2,619	1.07	57	0.84	114	0.84	105	0.47
		4	3,601	2,649	0.48	45	0.89	101	0.89	86	0.35
		5	3,601	3,884	0.65	235	1.18	377	0.60	314	0.48
		6	3,601	2,811	0.71	33	0.89	81	0.89	57	0.48
		7	3,601	2,680	0.44	51	1.26	169	0.81	154	0.49
		8	3,601	2,567	0.96	48	0.83	102	0.83	96	0.40
		9	3,601	4,117	0.90	52	0.84	109	0.84	111	0.39
		10	3,601	2,873	0.69	34	1.05	122	0.61	95	0.36
Set 2	(40,320)	1	3,602	2,969	0.99	78	0.90	177	0.90	155	0.47
		2	3,602	2,964	0.87	73	0.84	149	0.84	112	0.40
		3	3,601	2,796	0.96	117	0.94	216	0.94	264	0.49
		4	3,601	2,491	0.69	64	0.92	151	0.92	253	0.47
		5	3,601	2,052	1.09	85	0.84	173	0.84	142	0.46
		6	3,601	1,951	1.16	63	0.72	151	0.72	270	0.38
		7	3,601	1,943	1.01	108	0.76	205	0.76	153	0.42
		8	3,601	2,166	2.28	63	0.82	148	0.82	278	0.46
		9	3,601	3,277	0.96	113	1.04	274	0.79	219	0.48
		10	3,601	1,764	1.06	58	0.73	140	0.73	252	0.44
Set 3	(50,400)	1	3,601	1,947	2.40	84	0.79	198	0.79	329	0.43
		2	3,602	2,299	1.69	89	0.83	204	0.83	478	0.45
		3	3,601	3,661	1.90	160	0.96	304	0.96	546	0.45
		4	3,601	2,051	1.83	98	0.85	218	0.85	179	0.48
		5	3,601	1,833	1.71	103	0.95	226	0.95	311	0.44
		6	3,602	1,954	1.66	115	0.91	245	0.91	208	0.45
		7	3,601	2,877	1.25	128	0.82	264	0.82	201	0.42
		8	3,601	1,933	1.66	117	0.81	249	0.81	195	0.48
		9	3,601	2,140	1.73	160	1.16	369	0.66	241	0.40
		10	3,601	1,877	1.73	236	0.69	204	0.69	147	0.37
Set 4	(60,480)	1	3,601	3,534	2.77	479	0.82	445	0.82	315	0.50
		2	3,601	1,986	2.28	226	0.86	362	0.86	274	0.48
		3	3,602	1,806	1.80	238	0.85	261	0.85	641	0.50
		4	3,601	1,829	1.11	223	0.78	158	0.78	607	0.45
		5	3,602	1,849	2.82	303	0.95	270	0.95	805	0.44
		6	3,601	1,736	1.75	261	0.99	218	0.99	464	0.48
		7	3,601	1,972	2.16	246	0.87	161	0.87	277	0.48
		8	3,602	1,966	1.51	227	0.76	155	0.76	311	0.43
		9	3,601	1,916	2.69	265	0.91	181	0.91	399	0.42
		10	3,601	1,903	2.81	252	0.78	174	0.78	383	0.40
Set 5	(100,800)	1	3,608	465	5.07	829	0.88	743	0.88	1,324	0.43
		2	3,606	1,384	3.04	597	0.82	646	0.82	852	0.47
		3	3,603	462	6.14	664	0.76	795	0.76	946	0.45
		4	3,607	499	5.78	619	0.76	755	0.76	1,142	0.40
		5	3,604	511	6.59	648	0.84	754	0.84	990	0.48
		6	3,603	292	6.64	750	0.85	827	0.85	1,040	0.49
		7	3,604	691	7.18	722	0.85	817	0.85	967	0.48
		8	3,605	398	5.64	695	0.78	795	0.78	936	0.48
		9	3,605	501	5.93	793	0.83	953	0.83	1,068	0.48
		10	3,608	1,121	6.42	618	0.86	821	0.86	970	0.50

A.4 Replication of Table 6 in Flamand et al. (2018)

Table A.4

The effect of affinity relationships for the heuristic. For each relationship set $\in \{\mathcal{L}, \mathcal{H}, \mathcal{H}_2, \mathcal{H}_3\}$, *five product pairs are considered.*

Set	$(\lvert\mathcal{B}\rvert,\lvert\mathcal{N}\rvert)$	Inst.	No Affinity		$\mathcal{L}$		$\mathcal{H}$		$\mathcal{H}_2$		$\mathcal{H}_3$		$(\mathcal{L},\mathcal{H},\mathcal{H}_2,\mathcal{H}_3)$	
			CPU(s)	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*	*CPU(s)*	*Gap(%)*
Set 1	(30,240)	1	124	0.43	111	0.43	131	0.46	92	0.40	103	0.48	143	0.91
		2	53	0.47	41	0.48	72	0.48	55	0.44	60	0.43	123	0.81
		3	105	0.47	94	0.47	122	0.85	95	0.48	131	0.75	122	1.50
		4	86	0.35	73	0.35	84	0.42	64	0.41	72	0.42	123	0.42
		5	314	0.48	219	1.18	132	0.75	122	0.60	165	1.36	131	1.32
		6	57	0.48	57	0.48	57	0.41	55	0.30	67	0.38	128	0.46
		7	154	0.49	133	0.69	150	0.87	126	0.63	123	0.76	149	0.72
		8	96	0.40	80	0.40	124	0.54	78	0.41	101	0.42	127	0.74
		9	111	0.39	104	0.39	122	0.47	94	0.39	141	0.48	132	0.64
		10	95	0.36	79	0.35	120	0.63	75	0.40	108	0.45	126	1.03
Set 2	(40,320)	1	155	0.47	144	0.49	200	0.70	133	0.46	195	0.44	195	0.88
		2	112	0.40	119	0.40	188	0.51	150	0.46	137	0.44	189	0.57
		3	264	0.49	191	0.60	187	0.85	202	0.61	195	0.96	266	1.44
		4	253	0.47	110	0.47	158	0.44	160	0.43	183	0.57	206	0.81
		5	142	0.46	149	0.46	190	0.65	196	0.51	184	0.45	201	0.85
		6	270	0.38	110	0.38	176	0.45	166	0.47	102	0.49	206	0.53
		7	153	0.42	135	0.42	114	0.47	119	0.42	148	0.48	273	0.89
		8	278	0.46	115	0.47	206	0.63	118	0.44	177	0.46	192	0.86
		9	219	0.48	185	0.63	195	1.02	198	0.66	185	0.60	199	0.82
		10	252	0.44	97	0.46	186	0.99	94	0.49	155	0.47	198	0.97
Set 3	(50,400)	1	329	0.43	159	0.43	207	0.49	175	0.45	183	0.48	261	0.45
		2	478	0.45	186	0.48	237	0.46	155	0.48	180	0.49	237	0.48
		3	546	0.45	283	0.46	321	0.54	303	0.46	314	0.61	337	0.98
		4	179	0.48	174	0.49	178	0.46	178	0.42	236	0.40	215	0.48
		5	311	0.44	248	0.45	312	0.52	254	0.47	211	0.49	334	0.73
		6	208	0.45	240	0.45	305	0.59	181	0.37	232	0.40	216	0.47
		7	201	0.42	218	0.46	214	0.42	220	0.42	236	0.48	284	0.40
		8	195	0.48	205	0.48	208	0.43	183	0.43	214	0.48	221	0.48
		9	241	0.40	244	0.40	911	1.28	249	0.40	257	0.44	952	1.44
		10	147	0.37	154	0.37	302	0.57	120	0.47	157	0.50	302	0.70
Set 4	(60,480)	1	315	0.50	326	0.50	388	0.44	310	0.47	308	0.45	493	0.50
		2	274	0.48	274	0.50	337	0.46	304	0.47	274	0.44	360	0.45
		3	641	0.50	303	0.48	527	0.65	359	0.41	485	0.49	512	1.07
		4	607	0.45	281	0.46	268	0.46	247	0.46	352	0.44	407	0.45
		5	805	0.44	421	0.45	333	0.43	390	0.43	405	0.42	498	0.56
		6	464	0.48	432	0.46	441	0.49	400	0.45	517	0.47	545	0.80
		7	277	0.48	395	0.42	288	0.46	263	0.49	425	0.49	554	0.73
		8	311	0.43	323	0.43	417	0.45	525	0.42	307	0.48	569	0.77
		9	399	0.42	401	0.43	399	0.44	490	0.62	459	0.45	502	0.57
		10	383	0.40	374	0.47	507	0.51	571	0.65	323	0.49	500	0.63
Set 5	(100,800)	1	1,324	0.43	1,109	0.42	1,292	0.46	1,096	0.46	1,117	0.50	1,388	0.47
		2	852	0.47	831	0.49	1,068	0.50	1,030	0.48	915	0.47	1,142	0.46
		3	946	0.45	1,126	0.44	1,223	0.48	1,033	0.42	1,506	0.47	2,454	0.49
		4	1,142	0.40	1073	0.45	872	0.45	845	0.43	1,298	0.44	1,042	0.49
		5	990	0.48	1163	0.47	1,029	0.49	990	0.47	1,972	0.41	1,228	0.44
		6	1,040	0.49	937	0.42	1,023	0.47	982	0.48	1,830	0.49	1,439	0.43
		7	967	0.48	1,166	0.47	1,010	0.46	1,074	0.47	1,904	0.42	1,043	0.48
		8	936	0.48	874	0.48	998	0.47	899	0.49	1,128	0.44	1,165	0.48
		9	1,068	0.48	1,188	0.48	1264	0.47	1,050	0.45	1,061	0.45	956	0.48
		10	970	0.50	967	0.49	1,222	0.49	1,044	0.42	1,212	0.44	1,436	0.45

Note. The time limits are 120, 180, 300, 480, and 2400 CPU seconds for the test sets 1-5 respectively, as motivated in Flamand et al. (2018). However, this time limit is only checked after all shelves have been visited in the algorithm.

B Appendix. Algorithms for heuristic approach of Flamand et al. (2018).

The algorithms in this appendix are used for solving the enhanced versions of model APSA. Both Algorithms 1 and 2 are directly taken from (Flamand et al., 2018). For Algorithm 2, a slight adjustment is made in the first step (deriving an upper bound). As motivated in the main text, this is necessary because the upper bound constructed by solving the continuous relaxation of the MIP problem is of insufficient quality.

Algorithm 1: Initialization procedure (Flamand et al., 2018).

1 Input σ. Set $i = 1$, $\mathcal{S} = \emptyset$ ($\mathcal{S}$ refers to the set of selected product categories)
2 Set $i^* \leftarrow \sigma_i$
3 Find an optimal allocation for shelf i^*, regarding product categories in $\mathcal{N} \setminus \mathcal{S}$.
4 Let $\mathcal{N}_{i^*}$ be the set of product categories that are packed into σ_i
5 Set $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{N}_{i^*}$
6 **forall** $((j_1, j_2) \in \mathcal{H}_3)$ **do**
7   **if** $(j_1 \in \mathcal{N}_{i^*})$ **then**
8   Set $\mathcal{S} \leftarrow \mathcal{S} \cup \{j_2\}$
9   **end**
10   **if** $(j_2 \in \mathcal{N}_{i^*})$ **then**
11   Set $\mathcal{S} \leftarrow \mathcal{S} \cup \{j_1\}$
12   **end**
13 **end**
14 **if** $i = m)$ *or* $(\mathcal{S} = \mathcal{N})$ **then**
15   Stop.
16 **end**
17 Set $i \leftarrow i + 1$, Go to Step 3.

Algorithm 2: MIP-based re-optimization procedure (Flamand et al., 2018) with adjusted first step.

Set the upper bound equal to the best bound found by CPLEX after five MIP solutions have been found
Construct an initial feasible solution with a total reward $r^* = \sum_{i \in \mathcal{B}} r_i$ using Algorithm 1
incumbent $\leftarrow r^*$
$\hat{\kappa} \leftarrow (x, y, s)$
$\tau \leftarrow 4$
repeat
  tempset $\leftarrow \mathcal{B}$
  while *(|tempset| > $|\mathcal{B}|$ (mod) τ)* **do**
    Let $\Delta = (\Delta_1, \ldots, \Delta_m)$ denote a permutation of the shelves in tempset that is obtained by sorting them in non-increasing order of their current objective value contribution
    $\Omega = \text{round}(|\Delta|/\tau)$
    forall $k = 1, \ldots, \tau$ **do**
      Randomly select a shelf i between $\Delta_{(k-1)\Omega+1}$ and $\Delta_{k\Omega}$
      tempset $\leftarrow$ tempset - $\{i\}$
    end
    Solve the model to re-optimize the selected subset of shelves by considering all the products currently allocated on them as well as the unassigned products
    incumbent $\leftarrow r^*_{new}$
    $\hat{\kappa}^{new} \leftarrow (x^{new}, y^{new}, s^{new})$
    $\hat{\kappa} \leftarrow \hat{\kappa}^{new}$
  end
until *at least one stopping criterion is met*;

C Appendix. Brief description of code used in computational study

All the code for the computational study has been coded in Python. Algorithms 1 and 2 of Flamand et al. (2018) are coded in separe files and named accordingly. The MIP model is generated in $apsa_flamand$ and $apsa_extension$ for the replication and extension respectively. To produce the results of the paper, the main files need to be run, which are $main_replication$ and $main_extension$ for the replication part and extension part, respectively. Doc strings are provided were deemed necessary.