CATEGORY AND INVENTORY MANAGEMENT

Yasin Alan
Vishal Gaur
Johnson Graduate School of Management, Cornell University, Ithaca, New York

SHELF SPACE ALLOCATION

Problem Description

Retail stores have a limited amount of space and many products to display. The amount of shelf space allocated to an item affects its frequency of replenishment, incidence of stockouts, and demand rate. Therefore, finding the optimal amount of shelf space to allocate to each item becomes a key factor for success. Effective shelf space allocation leads to higher profits by increasing sales and customer satisfaction, creating better product visibility and brand exposure, and reducing inventory-related costs and stockouts [1–3]. Indeed, effective shelf space allocation has become harder and more critical in recent years because of increases in product variety and competition. For instance, on average, the number of consumer-packaged stock keeping units (SKUs) in the marketplace increased by 16% per year between 1985 and 1992, whereas shelf space expanded by only 1.5% per year during the same period [4]. A modern conventional supermarket offering major food departments, nonfood grocery, and limited general merchandize products has 20,000 to 30,000 sq. ft of floor space and it carries 20,000 to 40,000 SKUs [[1], p. 40].

Shelf space allocation is the process of apportioning the amount of space to each product in order to maximize the total store profit or another well-defined objective function subject to limited store space and other financial and operational constraints. Shelf space is usually measured in linear terms such as the number of *facings* and the length of each facing. The depth of a shelf and the dimensions of an item are also used in order to compute the shelf capacity allocated to each item.

Researchers from marketing, operations management, and other related fields have been developing tools and techniques for more effective shelf space allocation [5]. From a marketing perspective, the visual appeal of the display, the amount of product variety, and the location of each item in the store are important in order to maximize total sales. In particular, researchers define a metric called the *shelf space elasticity* of demand as the ratio of relative change in unit sales to relative change in shelf space. They measure shelf space elasticity for various types of products to evaluate the effect of display on sales. The key factors from an operations perspective are demand estimation, shelf replenishment frequency, inventory holding cost, the cost of stockouts, replenishment-related labor time and cost, and the impact of case pack sizes on inventory management policies. The operations perspective is more relevant for supermarkets, convenience stores, drug stores, and discount retailers because operational efficiency can lead to lower costs, which, in turn, increases competitiveness. The marketing issues are dominant for other types of retailers such as apparel stores in which the visual appeal of the display has a substantial demand-stimulating effect. Conflicting interests of the marketing and operations strategies also pose challenging questions.

In this article, we describe the measurement of shelf space elasticity of demand, the mathematical models and heuristic approaches to allocate shelf space, and commercial packages used in the industry.

Measurement of Shelf Space Elasticity

The amount of shelf space allocated to a product influences its unit sales. In one of the earliest studies on shelf space allocation, Cairns [[6], p. 43] states that “the more space

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

allocated to an item, the more likely it is to be seen by a shopper and, hence, the more likely it is to be purchased." However, the shelf space elasticity of demand can vary across products. Brown and Tucker [7] identify the following three classes of products with varying shelf space elasticity of demand:

1. Unresponsive products, for which changes in shelf space allocation have no impact on sales rate (e.g., salt and spices). These products are also generally price inelastic.
2. General use products, for which increasing shelf space leads to increase in sales but at a diminishing rate (e.g., breakfast foods and canned fruit and vegetables).
3. Occasional purchase products, for which shelf space has a step-function or threshold effect on sales. Sales increase slowly with shelf space at first, until a large display causes a steep increase in sales to a point of diminishing returns. These include impulse buys (e.g., candy).

Following Brown and Tucker [7], researchers have sought to measure shelf space elasticity for various types of products. Both retailers and manufacturers are interested in this measurement, but for different reasons. Retailers care about total profits across all products. They benefit because they can allocate different amounts of shelf space to different products depending on their space elasticities and gross margins. Manufacturers care about the profit from their own products. They benefit because they can build shelf space allocation into their merchandizing discussions with retailers. For example, it has been found that private-label or store-brand products have higher shelf space elasticity than competing national brand products [8]. Thus, an easy way for a retailer to shift sales from national to store brands is to give more space to store brands on its shelf display.

Among early works in this area, Kotzan and Evanson [9] have conducted an experiment at a drug store chain to evaluate the impact of changes in shelf space on unit sales. Their experiment was conducted on four products that met certain criteria related to demand uncertainty and availability of inventory. These products were tested in three stores for three weeks, where the number of facings allocated to each product was changed each week from 1, 2, or 3 facings. The authors discovered that three of the four products, a family-size Crest toothpaste, Hook's Red Mouth Wash, and Johnson and Johnson Assorted Band Aids, had statistically significant positive shelf space elasticities. For example, the sales of Crest toothpaste with 1, 2, and 3 facings were 219, 291, and 294 tubes, respectively. The results for the fourth product (Preparation H Suppositories) were inconclusive.

Kotzan and Evanson did not investigate why shelf space elasticity varied across these products. However, the academic literature suggests many reasons for such variation. For instance, Curhan [8] relates shelf space elasticity to product characteristics. He hypothesizes that items with smaller package sizes, lower prices, smaller market shares, private-label brands, and higher sales rates (i.e., fast moving items) would have higher shelf space elasticities. He also hypothesizes that greater product variety, more availability of substitutes, and lower repurchase frequency would lead to higher shelf space elasticities. He tests these hypotheses by conducting an experiment for grocery products in supermarket stores. In this experiment, shelf space is changed for 500 items and their unit sales are observed for 5 weeks before the change and 12 weeks after the change. Shelf space elasticity is measured as the ratio of percent change in unit sales to the percent change in shelf space. Curhan [8] obtains very low R^2 values. He concludes that the product characteristics do not satisfactorily explain the observed variation in shelf space elasticities because the impact of shelf space on unit sales is very small relative to the effects of other environmental variables, leading to a failure of the model. Nonetheless, the average space elasticity across all items in his dataset is 0.212, showing a positive effect of shelf space on sales. In another study, Frank and Massy [10] show that environmental variables such as store size, number of shelf rows, and shelf levels have a substantial impact on sales. Cox [11] also conducts an in-store experiment and provides evidence that staple product brands (e.g., salt brands) and impulse product brands that have low consumer acceptance are unresponsive to changes in shelf space. On the contrary, sales of an impulse product brand that has high consumer acceptance (e.g., Coffeemate) increase in shelf space.

Various conclusions are drawn in the literature from this early research. One is that shelf space elasticity is difficult to measure because a retail store is a dynamic environment and it is almost impossible to control factors such as retail prices, advertising, and addition and deletion of products, which have direct effects on sales [12,13]. Another is that shelf space elasticity is not large enough in magnitude to be managerially relevant. Instead, shelf space allocation should emphasize operational considerations such as the labor cost of restocking shelves and avoidance of stockouts. On the positive side, it is well recognized that shelf space elasticity varies across products, and is more important for private-label products and impulse-purchase items.

Research has also addressed how shelf space elasticity can be used by a retailer to increase sales and profits. For instance, Anderson [14] models the relationship between a product's market share and its share of shelf space using a logistic regression in order to find the profit-maximizing shelf space allocation. Dreze *et al*. [15] conduct experiments comparing two types of shelf management at a supermarket chain. In the first of these experiments, they change the shelf space allocation for each product to be proportional to the historical sales rate of the product in similar stores. Thus, in this method, they customize shelf space allocation in each store according to its historical sales. They note that this contrasts with the existing practice, which allocates shelf space in all stores of the chain identically regardless of differences in their sales mix. In these authors' second experiment, they reorganize the planogram for the store to facilitate cross-category merchandizing by placing complementary product categories closer to each other. The experiment shows a 4% increase in sales and profits due to customized shelf space and 5–6% increase due to planogram reorganization. They use the results of their experiment in an optimization model and estimate that there is a potential for 15% increase in sales by optimizing the shelf space allocation to each item using the estimated parameters. They conjecture that the increase in sales is driven by customers increasing their share of purchases at the subject supermarket store when they are presented with a better shelf space allocation.

Thus, we see that experiments have been widely used to study shelf space allocation. Next, we turn our attention to using the results of such experiments in optimization models.

Optimization Models

We explain the optimization of shelf space using a model given by Corstjens and Doyle [13]. This model has been widely used and improved upon since 1981. We summarize some of the later developments after presenting the model.

Consider a retailer with total available shelf space S^*. The retailer seeks to allocate this space among K products in order to maximize its total profit. Let s_i be the shelf space allocated to product i, β_i be the direct elasticity of sales of product i with respect to its shelf space s_i, and δ_{ij} be the cross-space elasticity of the sales of product i with respect to the shelf space s_j allocated to product j. δ_{ij} can be positive or negative, and need not be equal to δ_{ji}. Then, the total sales for product i are written as

$$q_i = \alpha_i s_i^{\beta_i} \prod_{j=1,\ldots,K,j\neq i} s_j^{\delta_{ij}}. \tag{1}$$

The gross margin from product i is written as $w_i q_i$, where w_i is the percent gross margin, and the variable store expense for product i given the sales quantity is written as $C_i = \gamma_i q_i^{\tau_i}$, where τ_i is the operating cost elasticity associated with the sales of product i. The retailer seeks to maximize its total profit, which is equal to the difference between the total gross margin and the total variable store expense. This problem is formulated as the following constrained nonlinear program:

$$\max \sum_{i=1}^{K} w_i \left[\alpha_i s_i^{\beta_i} \prod_{j=1,\ldots,K, j\neq i} s_j^{\delta_{ij}} \right] - \sum_{i=1}^{K} \gamma_i \left[\alpha_i s_i^{\beta_i \tau_i} \prod_{j=1,\ldots,K, j\neq i} s_j^{\delta_{ij}\tau_i} \right] \quad (2)$$

subject to

$$\sum_{i=1}^{K} s_i \leq S^*, \quad (3)$$

$$\alpha_i s_i^{\beta_i} \prod_{j=1,\ldots,K, j\neq i} s_j^{\delta_{ij}} \leq Q^* \quad i = 1, \ldots, K, \quad (4)$$

$$s_i^L \leq s_i \leq s_i^U \quad i = 1, \ldots, K, \quad (5)$$

$$s_i \geq 0 \quad i = 1, \ldots, K. \quad (6)$$

Here, the first constraint represents the upper limit on available shelf space in the retail store. The second constraint represents an upper limit on the amount of sales that can be achieved for each product in the store. The third constraint restricts the amount of shelf space that can be allocated to each item to lie between two control limits.

This problem is extremely difficult to solve because the objective function and one of the constraints are nonlinear. Corstjens and Doyle [13] use a geometric programming method to solve this problem. They illustrate their model using data from a retail chain selling various types of candy, ice cream, and greeting cards. First they obtain data on sales and facings for each product across 140 stores. They fit equation (1) to each product category to estimate the direct and cross-shelf space elasticities. They also obtain cost data from the management to estimate gross margin and variable store expenses. They then optimize shelf space using these estimated parameters. The results lead to substantial changes in shelf space allocation from the existing allocations. For example, the results show that the optimal allocations for large and small stores differ from each other due to variation in sales mix. Thus, the model incorporates direct and cross-space elasticities, profit margins, and operating costs to improve shelf space allocation. Corstjens and Doyle [16] later introduced a dynamic version of this model. It is a multiperiod model that takes into account anticipated changes in customer preferences and sales growth and decline of products. This model presents a long-term strategic view that encourages retailers to sacrifice short-term profits in order to maximize profits in the long run by taking shelf space away from declining products and allocating it to products with high growth potential. The optimization model of Corstjens and Doyle [13] has been further improved upon in subsequent research using many techniques, such as marginal analysis [17], dynamic programming [18], generalized Lagrange multiplier methods [19], and multistage optimization [20].

Most of the mathematical shelf space allocation models are nonlinear, mixed-integer problems that are computationally expensive for moderate-sized cases; even linear versions of this problem are NP-hard [2]. These difficulties have generated interest in heuristic solution approaches for these problems. For example, Urban [21] proposes a greedy heuristic to solve a nonlinear mixed-integer shelf space allocation problem. His algorithm starts with an initial solution including all items in the assortment plan. It then iteratively removes one item at a time from the assortment based on the greatest improvement in net profit estimated using a generalized reduced gradient. The algorithm stops when net profits cannot be improved any further by removing another item. Other techniques such as genetic algorithms [21–23], knapsack heuristics [24], local search methods [2], goal programming [25], simulated annealing [26], and greedy heuristics [27] are also used to solve shelf space allocation problems. In short, methods that are used to solve nonlinear programs can be tailored to solve shelf space allocation problems.

Another important consideration in shelf space allocation is its effect on inventory costs. Two items differing in shelf facings and/or service level requirements will have different inventory costs. For instance, an item with less shelf space will have more frequent stockouts and replenishments, which will lead to higher stockout and labor costs, *ceteris paribus*. The shelf space allocation models that we have discussed thus far assume that the shelves are always fully stocked. Thus, they do not address the role of inventory costs on shelf space allocation. Freund and Matsuo [28] study this aspect by modeling inventory replenishment as a periodic order-up-to policy and explicitly defining holding costs, review costs, and stockout risk. They show that the importance of considering inventory costs increases as the desired service level increases. Furthermore, they conclude that higher service level requirements force the retailer to have a smaller assortment because the net profit suffers due to high operating costs when there is a wide product variety.

In many retail stores, inventory is split into "display inventory" and "backroom inventory." Urban [21] distinguishes between the backroom and displayed inventories by tracking them separately. In his model, demand is a function of displayed inventory, whereas the backroom inventory allows the retailer to achieve economies of scale by ordering more than the shelf capacity and storing the excess units in the backroom. Thus, a backroom allows the retailer to stock more SKUs. Maiti and Maiti [23] adopt a similar framework. They provide a contractive mapping genetic algorithm to simultaneously solve the inventory management and shelf space allocation problems. See Urban [29] for an overview of the interdependencies between the inventory and shelf space allocation decisions.

Hwang *et al*. [22] analyze a model in which the demand rate is a function of the display location and quantity displayed. Their model determines shelf space allocation, order quantities, and the location where each brand should be displayed. Recent research by Kök and Fisher [30] has focused on integrating not only shelf space allocation and inventory decisions but also product assortment decisions in a single model. They determine what products should be selected to be put in the assortment and how much shelf space should be allocated to each item given inventory costs, product substitution, and shelf space constraints. They implement their results at a supermarket chain. It is noteworthy that earlier researchers have also sought to integrate shelf space allocation and assortment decisions into one model. For example, Anderson and Amato [31] determine product assortment and shelf space allocation decisions jointly, but without considering inventory costs.

Commercial Software

Early commercial software such as PROGALI, OBM, CIFRINO, SLIM, COSMOS, and HOPE are based on simple rules of thumb such as allocating more space to products with the highest average sales or profit margins. None of these tools explicitly considers elasticities [16–18]. The next-generation commercial solutions are hybrid knowledge-based systems that can integrate human expertise and algorithmic techniques. One of the earliest hybrid knowledge-based decision support systems (DSS) used for shelf space allocation is Resource-opt [32]. It takes past sales, market research information, and managerial intuition as user input and extends Corstjens and Doyle [13] to provide store managers with a user-friendly DSS. It has been utilized to redesign a hypermarket in France, and perform shelf space allocation for three departments of a Scandinavian store as well as an oil company's 2000 store franchise in Europe.

Presently it is common practice to use software assistance to generate visual diagrams (planograms) that show where every product in a retail store should be placed. For instance, PC-based systems such as Appollo (IRI) and Spaceman (Nielsen) are widely used at a strategic level. However, their operational level use is still very restricted due to their limited decision support functionality [20]. Generic optimization software packages such as LINGO (LINDO Systems) can be used at a tactical and/or operational level once the shelf space allocation problem is modeled as a nonlinear program [33].

Discrepancies occurring due to the use of planograms pose another problem as store managers have a tendency to deviate from recommended allocations. Most large retailer chains design planograms at the corporate level and give local store managers some degree of freedom to modify the recommended design. Incentives for store managers are frequently tied to stockouts on the shelf and discipline in adhering to the shelf space allocation decided by the corporate office. However, managers also need to consider other issues such as the sales effect of fully stocked shelves and holding labor and replenishment costs (most of which are not captured by high level planograms). Furthermore, they need to react to local campaigns run by competitors. van Woensel *et al*. [34] analyze the possible causes for discrepancies between the recommended planograms and actual allocations and discuss the negative impact of these deviations on marketing efforts and operational efficiency.

Moreover, since shelf space is a scarce resource, it is natural for competing suppliers to try to influence the retailers' allocation decisions via negotiations and contract terms in order to obtain more shelf space. Martin-Herran *et al*. [35] model the interaction between two suppliers and a retailer as a Stackelberg game where two suppliers are leading and competing for the follower's (i.e., the retailer's) shelf space. They focus on the effect of the suppliers' advertisement strategies and show that the Stackelberg open-loop equilibrium is time-consistent.

In conclusion, researchers have been addressing various aspects of the shelf space allocation problem for more than 40 years. However, other issues remain to be addressed, such as the impact on shelf space management of the introduction of new tools and techniques for inventory management (e.g., RFID tags, contracting, and vendor managed inventory systems), poor processes for replenishment from backroom to shelf, inventory data inaccuracy, and misplaced SKUs [36,37].

MODELS WITH INVENTORY DEPENDENT DEMAND

Problem Description

In classical inventory models, we minimize inventory-related costs under the assumption that the demand is exogenous. However, in retailing, where the inventory is visible to the end consumer, the rate of demand is often increasing in the amount of inventory stocked. As mentioned in the previous section, displayed stocks can stimulate demand [13]. Similarly, Wolfe and Little [38] provide evidence that the sales of style merchandise goods such as women's dresses are proportional to the amount of inventory displayed. In fact, the concept of psychic stock, defined as retail display inventory for stimulating demand, is motivated by the prevalence of this effect [39]. The demand-stimulating effect of inventories has motivated research to determine optimal inventory levels by solving the trade-off arising from the benefits and costs of holding more inventories. Early studies in the literature focus on developing inventory models for items with inventory-level-dependent rates. Recent literature has started to address strategic and tactical issues.

Economic Order Quantity Type Models and their Extensions

Economic order quantity (EOQ) type models are used to determine the optimal inventory policy when the demand is deterministic and influenced by inventory. We follow the research paper by Balakrishnan *et al*. [40] to introduce such models. Let the demand rate vary continuously as a function of the current inventory. If I denotes current inventory, then the demand rate is written as a nondecreasing concave function of I, denoted as $\lambda(I)$. For example, $\lambda(I) = \alpha(I + \phi)^{\beta}$, where $\alpha > 0$ is a scaling constant, $0 < \beta \leq 1$ is called the inventory elasticity of demand, and ϕ is a nonnegative base demand factor.

The mechanics of the model are as follows. Let $\hat{I}$ denote the inventory at the beginning of a replenishment cycle and T denote the length of the replenishment cycle. Inventory depletes at the rate of demand during the cycle. The inventory level at any time t during the cycle is obtained by integrating the inventory balance equation, $\mathrm{d}I(t)/\mathrm{d}t = -\lambda(I(t))$. For example, if $\lambda(I) = \alpha I^{\beta}$, then the inventory at time t is equal to $I(t) = \left(\hat{I}^{1-\beta} - \alpha t\right)^{1/(1-\beta)}$. Let $D(\hat{I}, T)$ denote the total demand during the replenishment cycle. It is equal to the difference between $\hat{I}$ and $I(T)$. The length of the replenishment cycle must be smaller than the time taken to deplete all the inventory. Let $T_{\text{runout}}(\hat{I})$ denote the time at which inventory will run out if not replenished.

The firm seeks to determine the order-up-to level $\hat{I}$ and the replenishment interval T in order to maximize its average profit per unit time. Let h denote the inventory holding cost rate, S the ordering cost, and r the per unit profit contribution. The profit per unit time is given by

$$\pi(\hat{I}, T) = \frac{1}{T}\left\{rD(\hat{I}, T) - S\right\} - h\overline{I}(\hat{I}, T), \quad (7)$$

where $\overline{I}(\hat{I}, T)$ is the average inventory per unit time and can be written as

$$\overline{I}(\hat{I}, T) = \frac{1}{T}\int_0^T I(t)\mathrm{d}t = \hat{I} - \frac{1}{T}\int_0^T D(\hat{I}, t)\mathrm{d}t.$$

The firm's demand-stimulating inventory problem, thus, involves maximizing (7) subject to the constraint that $T \le T_{\text{runout}}(\hat{I})$. This problem differs from the EOQ problem because it entails profit maximization rather than cost minimization and it allows replenishment to take place when the inventory is not fully depleted. Indeed, Balakrishnan *et al.* [40] distinguish two types of products depending on the optimal policy: early-replenishment products are those for which $T < T_{\text{runout}}(\hat{I})$ and runout-replenishment products are those for which the constraint is binding.

We illustrate the optimal solution obtained by Balakrishnan *et al.* [40] for a specific demand function which they call the reference demand model. In this model, the inventory elasticity of demand is equal to 0.5 and the demand rate is specified as $\lambda(I) = \alpha\sqrt{I + \phi}$. This yields the cumulative demand function $D(\hat{I}, t) = \alpha t\sqrt{\hat{I} + \phi} - (\alpha t/2)^2$ and $T_{\text{runout}}(\hat{I}) = (2/\alpha)(\sqrt{\hat{I} + \phi} - \sqrt{\phi})$. Substituting these expressions into the profit function and solving for the optimal $\hat{I}$ and T, we find that the early-replenishment strategy gives the optimal solution if the ordering cost S is below a certain threshold and the runout-replenishment strategy is optimal otherwise. The optimal replenishment cycle and order-up-to level for early-replenishment products are obtained as

$$T^* = 2\left(\frac{3S}{\alpha^2 h}\right)^{1/3}$$

and

$$\hat{I}^* = \frac{\alpha^2}{4}\left(\frac{r}{h} + \left(\frac{3S}{\alpha^2 h}\right)^{1/3}\right)^2 - \phi.$$

The optimal order-up-to level for runout-replenishment products is obtained similarly.

Many insights can be obtained from this solution. As in the classical EOQ model, in this model T^* depends on the ratio S/h. Moreover, T^* does not depend on r. Thus, as in the EOQ model, the length of the replenishment cycle is determined by the trade-off between the ordering cost and the holding cost. The optimal order-up-to level, however, depends on r/h. Hence, it is determined by the trade-off between the contribution and the holding cost. When the contribution increases, the replenishment cycle remains unchanged, but the firm carries more inventory and places its orders at higher reorder points. Balakrishnan *et al.* [40] also construct a heuristic and refer to it as the adaptive EOQ policy. In this heuristic, the firm uses the EOQ formula to determine the order quantity in each cycle, but recalibrates the demand rate parameter λ using the observed average demand rate in the previous cycle. This heuristic differs from the optimal policy because it does not allow orders to take place before the inventory runs out. However, it enables the firm to learn about the dependence of demand on inventory from historical data. Balakrishnan *et al.* [40] show that this heuristic converges to an equilibrium order quantity. However, it performs poorly because (i) it waits for the inventory to run out before placing a reorder and (ii) in each cycle, it orders too little and too frequently. For example, their analysis shows that, if $\lambda(I) = \alpha I^{0.5}$ and the product is an early-replenishment product, then the profit from the adaptive EOQ policy is always less than 40% of the optimal profit.

The demand function $\lambda(I) = \alpha I^{\beta}$ is appealing because it is similar to functions used to model the dependence of demand on shelf space [13]. This functional form makes the parameter estimation relatively easy because we can do a logarithmic transformation and fit a linear regression model to estimate α and β. It is possible to use many similar functional forms in a model to capture the demand-stimulating effect of inventory. Gupta and Vrat [41] assume that the demand rate depends only on the initial inventory, and remains constant at this rate. Baker and Urban [42] and Urban [43] allow demand to vary continuously with inventory and show that it may not be optimal to wait to place an order until the inventory level reaches zero. Urban [44] gives a recent survey of the literature on this topic.

Modeling the inventory dependent demand for perishable items require an additional structure since demand for a perishable item decreases not only with inventory but also due to loss of product freshness and/or an approaching expiration date. Such models have been studied in the literature [45–48]. For example, Balkhi and Benkherouf [48] model demand as $\lambda(I,t) = I^{\beta}G(t)$, where I is the current inventory, t denotes time, and $G(t)$ is an increasing function of time, and the inventory is allowed to deteriorate continuously over time at a fixed rate θ. Instead of deterioration of inventory, another feature to consider in certain situations is that the value of the unsold inventory may decrease over time [49]. This happens for seasonal or fashion products.

Strategic and Tactical Implications

Strategic consumers indirectly observe retailers' inventory policies over time. If a consumer's observations lead her to believe that she has a low probability of finding a desired product at a given retailer, she might choose to shop at a competitor instead. For instance, Anderson *et al*. [50] present empirical evidence showing that consumers who experience a stockout are less likely to place an order; these consumers also order fewer items, and spend less. These authors also show that the retailer might be able mitigate the cost of a stockout if an item is out of stock due to high popularity as consumers are more willing to backorder scarce items. On the contrary, consumers may be willing to pay a premium price for consistently high service rates. For instance, Dana [51] describes a small experiment on video stores and argues that Blockbuster's advertised claims of high availability allow the retailer to charge higher prices. These empirical results show that the retailer's stocking decisions and ability to manage stockouts shape its reputation which affects its future demand and pricing power.

Dana and Petruzzi [52] present a model in which consumers form beliefs about the service level offered by a retailer. They then decide whether to visit the retailer based on these beliefs. This model describes a single period, which can be interpreted as a steady-state representation of a repeated game between the consumers and the retailer. Their analysis shows that, when the retailer recognizes the effect of its service level on demand, it stocks higher inventory levels. As a consequence, the retailer attracts more customers and earns a higher expected profit. Similarly, many models have been developed in the recent research describing consumers who behave strategically by deciding when to visit a given retailer. Their strategic behavior can be affected by product availability, expectation of markdowns, and so on. We do not describe these models in this article, since they form a large body of literature by themselves.

Overlooking the demand-stimulating effect of inventory also creates incentive misalignment issues in retail operations, as most of the automated inventory replenishment systems try to minimize inventory-related costs whereas store managers are assessed on revenues. van Donselaar *et al*. [53] empirically show that retail stores managers have a tendency to deviate from order advices generated by an automated inventory replenishment system. Their analysis illustrates that the store managers improve the automated replenishment system's performance by systematically modifying the order advices because they are able to capture the demand-stimulating effect of inventories which is ignored by the system. Similarly, Kesavan *et al*. [54] estimate the effect of gross margin and inventory on each other and empirically show that more accurate firm level sales forecasts can be achieved by incorporating its relationship with inventory and gross margin.

In conclusion, capturing the demand-stimulating effect of inventory is necessary for effective inventory management. However, it is not sufficient to achieve firm level success. Identifying the relationship between demand and inventory should be seen as a part of a broader action plan called demand-based (supply chain) management [55]. This approach seeks to maximize the total value (i.e., profits and/or other well-defined objectives of the entire supply chain) by addressing the interdependencies between inventory, demand, pricing, and marketing. Overlooking this approach and tackling the demand inventory relationship independently of pricing and marketing leads to suboptimal solutions. Put differently, pricing, promotional marketing, and/or other demand manipulation tools should be used to mitigate the negative consequences of the relationship between demand and inventory.

MODELS OF RETAIL COMPETITION

The actions of one retailer directly affect not only its own demand but also the demand faced by its competitors. Promotions, price discounts, and increases in service level can increase its demand due to customers switching their purchases from competitors to its stores. On the contrary, core service failures (e.g., stockouts, billing mistakes, and service catastrophes), service encounter failures (e.g., impolite and unknowledgeable employees), and pricing failures (e.g., high prices and deceptive pricing strategies) can lead to customer losses [56]. These changes in demand can be temporary or permanent. Therefore, a retailer should determine its operational policies taking into account their competitive implications.

Retail competition can be modeled in many ways of varying complexity depending on the extent of competitive interaction among firms in the marketplace. For example, we may consider

1. *single-period models*, in which consumers switch from one retailer to another in the immediate period upon experiencing a stockout;
2. *multiperiod models*, in which consumers switch from one retailer to another in a future period upon experiencing a stockout;
3. *models of full or partial information*, in which consumers have some prior knowledge of inventories at competing retailers and choose whether to visit a retailer upon evaluating their chances of finding a product in stock;
4. *learning models*, in which consumers do not have knowledge of product availability but learn about their probability of finding a product in stock based on their own historical experience with a retailer, and then modify their future shopping behavior accordingly;
5. *multidimensional models*, in which consumers select the retailer to visit based on many service dimensions, such as price, promotions, and service level.

At the simplest level (in (1) above), consider a model studied by Lippman and McCardle [57]. In this model, there are two retailers competing with each other in a single period. The price and cost parameters are given and are equal across retailers. The retailers compete only on the basis of their inventory levels y_i. The total demand is a random variable denoted as D. It is allocated between the retailers via some splitting rule; let D_i denote the initial allocation of demand to firm i. If the demand allocated to firm j is greater than its inventory, that is, if $D_j > y_j$, then a fraction a_i of the excess demand is reallocated to firm i. Thus, the effective demand at firm i, including its initial demand and reallocation, is given by $R_i = D_i + a_i \max\{0, D_j - y_j\}$, where $0 \leq a_i \leq 1$. Each retailer seeks to determine the inventory level that maximizes its expected profit. Note that each retailer's choice of inventory affects the demand faced by its competitor because of substitution or reallocation of excess demand.

Lippman and McCardle [57] describe many ways by which the initial demand for each retailer may be obtained. For example, a deterministic splitting is one in which total demand is allocated between the retailers in a fixed proportion. A random splitting can be obtained in many ways: if each customer flips a coin to choose a retailer, then it is called an incremental random splitting of demand; if the first customer flips a coin to choose a retailer and each subsequent customer follows the previous customer, then it leads to a simple random splitting or herd behavior, that is, all of the demand is allocated to one or the other retailer with probability 0.5; finally, the initial demand at the two retailers may be given by independent random variables. Lippman and McCardle [57] present a beautiful example of a tourist bus visiting a chateau to illustrate the demand models. They show that there exists a pure strategy Nash equilibrium in inventory levels in this competitive game. The equilibrium need not be unique. Moreover, competition can be detrimental to the firms because it can drive down the industry profit to zero and drive up the total inventory level in the industry to be higher than the monopolist's optimal inventory, that is, the optimal inventory quantity if there were a single firm serving the entire demand.

Whereas single-period models allow us to study how competition affects the demand faced by a retailer in the current time period, multiperiod models (in (2) above) are useful to study the effect on demand in subsequent time periods. A multiperiod model of competition can allow us to measure the future goodwill cost of losing a customer if the service level in the current period is reduced. Following Hall and Porteus [58], suppose that there are two firms competing in a marketplace over a finite time horizon of T periods. Let D_{it} be the demand and y_{it} be the inventory level for firm i in period t. The unsatisfied demand is then given by $\max\{0, D_{it} - y_{it}\}$, and represents the number of customers who experience a stockout. Suppose that a fraction γ_i of these customers switch to the competitor in the next time period. Thus, if a firm were to increase its inventory, it would lose fewer customers to its competitor. The objective of the firm in the model is to determine its inventory levels in order to maximize the total expected profits over the time horizon. With some additional simplifying assumptions, Hall and Porteus [58] show that this problem yields a unique subgame perfect Nash equilibrium. Their solution gives an imputed lost goodwill cost for each firm, which is a function of the present value of an additional customer in the next time period and the probability of losing that customer due to a stockout. The retailer with less sensitive customers, that is, a smaller value of γ_i, faces a lower imputed goodwill cost. Such a retailer will provide a lower service level but still enjoy less defections and a larger market share.

A different way to model competition due to inventory is to define fill rate or service level as a dimension of quality. Fill rate is defined as the fraction of demand satisfied by a retailer from stock. It is equal to the ratio of expected sales to mean demand; if y is the amount of inventory stocked by a retailer and D is the random demand faced by it with mean $E[D] = \mu$, then fill rate is equal to $E[\min\{D, y\}]/\mu$. The numerator in this expression is the expected sales of the firm, which is given by the minimum of demand and inventory. When the fill rate is high, fewer customers experience stockout, and therefore fewer customers are likely to be dissatisfied with the retailer. Therefore, the demand faced by a retailer becomes a function of its fill rate as well as the fill rates of its competitors. This concept is usually expressed in models of type (3)–(5) listed above by defining the market share of a retailer as

$$\frac{f_i}{\sum_j f_j},$$

where f_i is the service level offered by the subject retailer. In models of full information, customers know the inventory level stocked by each retailer and choose which firm to visit by weighing their probability of finding the product in stock [51]. In models of partial information, customers do not know the inventory levels of the retailers, and use other cues such as price to form expectations about the fill rates provided by retailers [51,59]. Learning models are multiperiod models in which customers learn about the fill rates from their past shopping experience at each firm [60,61].

Consumers' reaction to price and service quality variations might be asymmetric. That is, consumers might weigh negative experiences (losses) more than equivalent positive experiences (gains). For instance, Hardie *et al*. [62] argue that loss aversion and the position of brands relative to multiattribute reference points (e.g., price and quality) influence the brand choice. They empirically support this claim by estimating and comparing gain and loss coefficients for price, quality, loyalty, and the presence of advertising for orange juice sales using scanner data. These analyses can be utilized to model the effect of asymmetric consumer behavior on the new product introduction and price promotions. Similar asymmetry also arises in services based on satisfying and unsatisfying service experiences. For instance, consumers react more strongly to a stockout of a necessity items such as bread because they always expect to see such items in stock (negative bias). On the other hand, they might have positive bias for prestige products such as designer suits because they expect to search for these items. Gaur and Park [61] show that the effect of competition on total inventory levels and total industry profits depends on the type of bias exhibited by consumers. Moreover, when retailers have different costs, the difference in market shares of the retailers also depends on the type of bias. The lower cost retailer enjoys greater market share and profit differential when consumers have a negative bias, whereas a positive bias tends to attenuate the effect of competition for the higher cost retailer.

Competition does not necessarily occur in a single dimension. For instance, two retailers selling an identical product might compete on selling price and service. Service can be measured with a single proxy such as the fill rate [63] or it can be a performance measure which aggregates many aspects of the shopping experience such as promotions, advertising, and customer relations into a single decision variable [64]. In both cases, the product can be treated as a bundle of two attributes, price and service. Tsay and Agrawal [64] consider a single period setting in which two competing retailers obtain a product from a common manufacturer, and discover that there are cases under which both retailers prefer an increase in competitive intensity because adding a small amount competition in one dimension mitigates the competitive intensity in the other. Bernstein and Federgruen [63] characterize an infinite-horizon, stochastic general equilibrium model for competing firms under different competition scenarios and demand processes.

Competition can affect firms in many other ways. A retailer might be willing to offer a monetary incentive to a customer in order to convince her to backorder instead of switching to another retailer [65]. The option to backorder reduces competition because unsatisfied demand is not necessarily lost. Another aspect of retail competition is the speed of delivery. If firms compete on the delivery time, then holding inventories is utilized as a tactic to reduce the customer waiting time and increase sales at the expense of high inventory holding costs. In fact, firms are more likely to switch from a make-to-order policy to a make-to-stock policy when the number of competitors increases. This type of competition increases consumer welfare and reduces retailers' profits [66]. Factors such as the firm and consumer characteristics, service quality, and searching costs also affect retail competition. For instance, McGahan and Ghemawat [67] study the relationship between firm sizes and competition to retain customers in a two stage game. In the first stage, firms try to build up loyalty among existing customers. In the second stage, they compete on price. Their analysis shows that large firms are likely to exhibit greater customer retention rates than their small rivals in equilibrium. Lastly, if customer search costs are low in an oligopolistic price competition setting, profit-maximizing firms may choose to have occasional stockouts to reduce competition. Reduced competition allows the firms to charge higher prices which might offset the effect of lost sales due to stockouts [68].

In conclusion, retailer competition has various aspects including, but not limited to, quantity, price, and service quality. Taking competition into account helps retailers create more accurate models that can capture market dynamics and consumer choice, which will lead to more effective pricing and stocking decisions.

SUBSTITUTION AND TRANSSHIPMENT MODELS

Adoption of modern information technology tools such as ERP (enterprise resource planning) systems and web-based inventory tracking applications has led to remarkable improvements in retail supply chain transparency. Nowadays, it is possible to obtain real-time information regarding on-hand and in-transit inventory quantities for each location in a retailer's supply chain. In some systems, it is even possible to track lost sales due to shortages so that more accurate demand forecasts can be made. Reduced information and transportation costs and shorter transportation lead times have enabled companies to move items not only from an upper installation (e.g., a warehouse) to a lower installation (e.g., a store) but also between any two lateral points in the system (e.g., from one store to another) [69].

These technologic advancements have allowed companies to utilize two risk pooling techniques more effectively: product substitution and transshipment. Substitution redistributes demand from a stocked-out product to another product with excess inventory. Lateral transshipment redistributes inventory from stores with excess on-hand inventory to stores facing shortages or low inventory levels [70]. These techniques are complementary because substitution can be utilized when the consumer is willing to purchase a similar item instead of waiting for her favorite item to be restocked, whereas lateral transshipment is more appropriate when the consumer is willing to delay her purchase. For instance, a consumer might be willing to purchase a 13.5-oz shampoo of the same brand when she could not find the 25.4-oz size. However, she might be willing to wait for designer shoes if her size is temporarily out of stock. The mathematical models to compute solutions that provide substitution and transshipment capabilities are similar. Below, we first summarize the effect of substitution on inventory management. Then, we describe transshipment policies. The topic of assortment planning is similar to substitution models and is discussed in a separate section. Typically, substitution models have a fixed number of available products, whereas the number of products to stock is a decision variable in assortment planning models.

Substitution

The implications of substitution on retail profits and inventory levels are important to study because consumers are often willing to purchase substitute items when they face stockouts. According to a survey conducted by Food Marketing Institute, more than 80% of the survey participants would be willing to buy a substitute item if their favorite item were not available [71]. Although demand substitution mitigates the effect of lost sales by switching demand from one item to another, it complicates inventory management because the sales of each item now depend not only on its own inventory and demand but also on the inventory and demand of all other items. Therefore, inventory policies developed without taking the substitution effect into account can lead to large profit losses.

There are two types of substitution phenomena, retailer driven and consumer driven. Under the first scenario, the retailer satisfies the demand for one product using another, possibly higher quality, product to mitigate stockouts. For instance, a downward substitution takes place when a high quality item can be downgraded and used as a substitute for a low quality item but not vice versa. Downward substitution can be understood using the model in Bassok *et al.* [71] as follows. Suppose that there are N products and N demand classes. The demand from class i can be satisfied using inventory of any product j such that $j \leq i$, thus representing downward substitution. The retailer earns revenue p_i for meeting demand of type i, incurs backorder cost π_i for shortages in class i, and incurs a cost of substitution b if demand for class i

is satisfied using inventory of some other product $j < i$. Product i can be purchased at cost c_i and its excess inventory can be salvaged at s_i. The retailer's profit maximization problem can be decomposed into two parts. In the first stage, the retailer determines the amount of inventory y_i of each product to stock. Then, a random amount of demand of each class is received. Given this demand, the retailer determines how to allocate the available inventory of various products among the demand classes in order to maximize its profit. The second stage problem can be formulated as the following linear program:

$$G(y_1, \ldots, y_N, D_1, \ldots, D_N) = \max_{u_i, v_i, w_{ji}} \sum_{i=1}^{N} \left[p_i w_{ii} + \sum_{j=1}^{i-1} (p_j - b) w_{ji} + s_i v_i - \pi_i u_i \right],$$

subject to

$$\begin{aligned} u_i + \sum_{j=1}^{i} w_{ji} &= d_i \quad i = 1, \ldots, N \\ v_i + \sum_{j=1}^{N} w_{ij} &= y_i \quad i = 1, \ldots, N \\ w_{ij} &\geq 0 \quad i,j = 1, \ldots, N \\ u_i, v_i &\geq 0 \quad i = 1, \ldots, N. \end{aligned}$$

The decision variables in this formulation are w_{ij}, which is the amount of product i used to satisfy demand class j; u_i, the amount of shortage in demand class i; and v_i, the excess inventory of product i. The objective function consists of the revenue from meeting demand, the salvage value of excess inventory, and the cost of shortages. The constraints implement the balance of flow between supply and demand within the downward substitution structure. Given the solution of the second stage problem for any inventory level and demand realization, the first stage consists of determining the inventories to maximize total expected profit, that is,

$$\max_{y_1, \ldots, y_N} E[G(y_1, \ldots, y_N, D_1, \ldots, D_N)] - \sum_{i=1}^{N} c_i y_i.$$

Bassok *et al*. [72] derive a solution for this problem by showing that the profit function in the first stage problem is concave and submodular in inventory levels under mild assumptions on the price and cost parameters. Besides being suitable for many kinds of applications, this problem formulation has also been used to give an upper bound on the profit function of the substitution problem under consumer-driven substitution [73].

In consumer-driven substitution, the excess demand from one product is reallocated to other products according to some fixed substitution rule. We describe this model using Netessine and Rudi [74]. Consider the same notation as above. However, now, demand for class i must be satisfied first by available inventory of product i. Then, unsatisfied demand of class i is allocated to all the other products in a fixed proportion. Let a_{ij} denote the fraction of unsatisfied demand of class i allocated to product j, where $a_{ij} \in [0, 1]$ and $\sum_{j=1}^{N} a_{ij} \leq 1$. Therefore, the effective demand for product i is equal to

$$D_i + \sum_{j \neq i} a_{ji} (D_j - y_j)^+,$$

and the leftover inventory of product i is

$$\left(y_i - D_i - \sum_{j \neq i} a_{ji} (D_j - y_j)^+ \right)^+ .$$

Here, $(D_j - y_j)^+$ denotes $\max\{0, D_j - y_j\}$. Thus, the retailer seeks to maximize its expected profit, which is a function of the total sales revenue, the salvage value of leftover inventory, and the purchasing cost of the inventory across all products. Note that the substitution is up to "first-level," that is, when excess demand of product i is reallocated to product j, then a stockout of product j does not lead to further reallocation of the leftover demand to other products. Even so, this problem is extremely difficult to solve. Netessine and Rudi [74] derive first-order necessary optimality conditions for this problem, but show that there may be multiple local maxima. For simpler two- and three-product problems, the profit function is, however, concave [75,76].

Substantial savings in inventory-related costs can be achieved when items in the product portfolio are highly substitutable, that is, the probability of accepting another item is relatively high for a customer that cannot find her favorite item [77]. However, demand substitution does not necessarily lead to decrease in the total inventory level. For instance, for a two-product case, it has been shown that the total inventory level may increase when only one item is substitutable [78].

Transshipment

The purpose of transshipment is to redistribute inventory so that the right quantities are available in the right location [79]. Lateral transshipment can be divided into emergency lateral transshipment (ELT) and preventive lateral transshipment (PLT) [80]. ELT mandates emergency transfers from a retailer with excess stock to a retailer that has a stockout [81,82]. This policy responds to stockout incidents after the realization of demand. On the other hand, under a PLT policy, items are transferred among locations in order to balance inventories in anticipation of stockout [83]. PLT has a nonmyopic view which tries to reduce the risk of future stockouts by redistributing the inventory [70]. In both ELT and PLT policies, the initial inventory at each location is planned with the view to allowing transshipment in the future. In the most primitive case, if the retailer does not conduct any transshipment, the inventory decision for each location can be made independently. Thus, transshipment requires the inventories at all locations to be determined jointly, adding considerable complexity to the problem. The benefits are that it provides risk pooling across locations, and generally reduces the total inventory requirement.

Most of the early literature is concentrated on ELT policies for repairable items because they have low demand rates and high backorder costs and thus could benefit the most from transshipment. The models used in this context are one-for-one continuous-review models similar to the seminal METRIC model of Sherbrooke [84]. For instance, Lee [81] studies a one-for-one multiechelon model for repairable items which allows lateral transshipments between identical retailers. He derives approximations for the commonly used performance measures such as the backorder levels and the number of lateral transshipments. These approximations are used to determine the optimal stocking levels. He shows that the use of lateral transshipment leads to large cost savings. Axsäter [82] extends these analyses under a scenario which the retailers are not identical. Archibald *et al*. [85] utilize Markov decision processes to study a two-location, multi-period, multi-item transshipment problem subject to capacity constraints. They show that the order-up-to policy is optimal when the demand for each item arises according to independent Poisson processes at each location. Although it is, in theory, possible to represent an inventory system as a Markov model and derive the steady-state probabilities, this approach may not always be computationally feasible since the state space grows exponentially with system size. Dada [86] focuses on developing a fast procedure to approximate the steady-state expected performance of a two-echelon system with transshipment. His model provides tight bounds on the system performance.

Due to the complexity of the transshipment problem, it is useful to devise simple heuristics. For instance, when shortage costs differ among locations, a simple yet effective technique is to allow unidirectional transshipments, that is, transshipments from locations with lower shortage cost to locations with higher shortage cost, but not in the opposite direction [87]. Another heuristic takes reorder points and batch quantities as given and tries to fulfill the excess demand at a retailer by a lateral transshipment from the retailer with most stock on hand [88]. This heuristic is useful under complex decision situations because it does not jointly optimize inventories across locations.

ELT policies for a supply chain with multiple retailers with different cost structures and demand parameters have also been studied under a continuous-review one-for-one inventory policy [89] as well as for a periodic-review system with a base stock inventory policy [90].

A PLT policy is modeled by Das [83] for a two-location stochastic inventory problem under centralized decision making. He implements this policy by setting a predetermined time point within each period at which the decision maker can move items from the overstocked location to the understocked location. He shows that a base stock policy with a transfer to the understocked location to bring its inventory level closest to its base stock level without decreasing the inventory level of the overstocked location below its base stock level is optimal.

Lee *et al*. [80] study a periodic-review model and develop a transshipment policy which can be classified as a combination of ELT and PLT. Namely, they define predetermined and fixed upper, lower, and target service levels which are used to compute the lateral transshipment quantities. At the end of each review period, retailers with inventory levels exceeding the corresponding upper service level send their excess inventories to retailers with low inventory levels or stockouts. A retailer with a low inventory level determines the amount of inventory that it can receive using the difference between the low and target service levels. Thus, this policy performs inventory balancing as well as emergency transshipments.

Two extreme transshipment policies are to never transship and always transship when there is a shortage at one location and stock is available at another location. It has been observed that choosing the better of these two extreme policies leads to a performance which is almost as good as a complex policy that takes the future impact of a stock transfer into account [91].

When a supply chain is centralized, the objective of transshipment is to optimize the overall system performance. However, when locations within a supply chain belong to different organizations, transfer prices affect each location's profitability and willingness to participate in a transshipment activity. In general, when each location tries to maximize its own profits, the resulting Nash equilibrium will not maximize joint profits. However, it is possible to set transshipment prices to create supply chain coordination such that the decisions made by each location are consistent with joint-profit maximization [92]. Furthermore, in decentralized supply chains, it might be necessary to offer some incentives to supply chain partners in order to prevent free-riding and implement effective inventory distribution and transshipment policies [69]. Dong and Rudi [93] study a transshipment model in which an external supplier sells to multiple retail stores owned and operated by the same firm. They show that stores' order quantities are less sensitive to the wholesale price under a transshipment policy due to risk pooling. Hence, the manufacturer benefits from transshipments at the expense of the retailers because it can charge a higher wholesale price. Zhao *et al*. [94] consider a decentralized system with a large number of independent retailers and prove that a threshold requesting and rationing policy is optimal. Under this policy, there exist thresholds $Z_i \leq K_i \leq S_i$ for each retailer i, denoting the optimal requesting, rationing, and base stock levels, respectively, for retailer i. It is optimal to send a transshipment request to another retailer only if the inventory level is below the requesting level Z_i, and to fill a received transshipment request only if the inventory level is above the rationing level K_i.

To sum up, lateral transshipment is a way to perform risk pooling and satisfy demand especially for low demand, high stockout cost items. However, one should take the transshipment and replenishment lead times; related ordering, holding, transportation, and stockout costs; structure of the supply chain (e.g., centralized vs decentralized); alternative risk pooling techniques (e.g., substitution); and the underlying demand distributions into account in order to develop an effective transshipment policy. Chiou [95] presents an extensive survey of the academic literature in this area.

REFERENCES

1. Levy M, Weitz BA. Retailing management. 6th ed. New York: McGraw-Hill/Irwin; 2007.
2. Lim A, Rodrigues B, Zhang X. Metaheuristics with local search techniques for retail shelf-space optimization. Manage Sci 2004;50(1):117–131.
3. Buttle F. Retail space allocation. Int J Phys Distrib Mater Manage 1984;14(4):3–23.
4. Quelch JA, Kenny D. Extend profits, not product lines. Harvard Bus Rev 1994; 72(5):153–160.
5. Corstjens J, Corstjens M. Store wars: the battle for mindspace and shelf space. Chichester, UK: Wiley; 1995.
6. Cairns JP. Allocate space for maximum profits. J Retailing 1963;39(2):41–45.
7. Brown W, Tucker WT. The marketing center: vanishing shelf space. Atlanta Econ Rev 1961;11(10):9–13.
8. Curhan RC. The relationship between shelf space and unit sales in supermarkets. J Mark Res 1972;9(4):406–412.
9. Kotzan JA, Evanson RV. Responsiveness of drug store sales to shelf space allocations. J Mark Res 1969;6(4):465–469.
10. Frank RE, Massy WF. Shelf position and space effects on sales. J Mark Res 1970;7(1):59–66.
11. Cox KK. The effect of shelf space upon sales of branded products. J Mark Res 1970; 7(1):55–58.
12. Curhan RC. Shelf space allocation and profit maximization in mass retailing. J Mark 1973; 37(2):54–60.
13. Corstjens M, Doyle P. A model for optimizing retail space allocations. Manage Sci 1981;27(7):822–833.
14. Anderson EE. An analysis of retail display space: theory and methods. J Bus 1979; 52(1):103–118.
15. Dreze X, Hoch SJ, Purk ME. Shelf management and space elasticity. J Retailing 1994;70(4):301–326.
16. Corstjens M, Doyle P. A dynamic model for strategically allocating retail space. J Oper Res Soc 1983;34(10):943–951.
17. Bultez A, Naert P. SHARP: shelf allocation for retailers' profit. Mark Sci 1988;7(3):211–231.
18. Zufryden FS. A dynamic programming approach for product selection and supermarket shelf-space allocation. J Oper Res Soc 1986;37(4):413–422.
19. Hansen P, Heinsbroek H. Product selection and space allocation in supermarkets. Eur J Oper Res 1979;3(6):474–484.
20. Yang MH, Chen WC. A study on shelf space allocation and management. Int J Prod Econ 1999;60:309–317.
21. Urban TL. An inventory-theoretic approach to product assortment and shelf-space allocation. J Retailing 1998;74(1):15–35.
22. Hwang H, Choi B, Lee MJ. A model for shelf space allocation and inventory control considering location and inventory level effects on demand. Int J Prod Econ 2005;97(2):185–195.
23. Maiti MK, Maiti M. Multi-item shelf-space allocation of breakable items via genetic algorithm. J Appl Math Comput 2006;20(1–2):327–343.
24. Yang MH. An efficient algorithm to allocate shelf space. Eur J Oper Res 2001; 131(1):107–118.
25. Reyes PM, Frazier GV. Goal programming model for grocery shelf space allocation. Eur J Oper Res 2007;181(2):634–644.
26. Borin N, Farris P, Freeland J. A model for determining retail product category assortment and shelf space allocation. Decis Sci 1994;25(3):359–384.
27. Bai R, Burke EK, Kendall G. Heuristic, metaheuristic and hyper-heuristic approaches for fresh produce inventory control and shelf space allocation. J Oper Res Soc 2008;59(10):1387–1397.
28. Freund R, Matsuo H. Retail inventory and shelf space allocation. Working paper. Department of Management, University of Texas at Austin; 1997.
29. Urban TL. The interdependence of inventory management and retail shelf management. Int J Phys Distrib Logist Manage 2002;32(1):41–58.
30. Kök AG, Fisher ML. Demand estimation and assortment optimization under substitution: methodology and application. Oper Res 2007;55(6):1001–1021.
31. Anderson EE, Amato HN. A mathematical model for simultaneously determining the optimal brand collection and display-area allocation. Oper Res 1974;22(1):13–21.
32. Singh MG, Cook R, Corstjens M. A hybrid knowledge-based system for allocating retail

space and for other allocation problems. Interfaces 1988;18(5):13–22.

33. Hariga MA, Al-Ahmari A, Mohamed AA. A joint optimisation model for inventory replenishment, product assortment, shelf space and display area allocation decisions. Eur J Oper Res 2007;181(1):239–251.
34. van Woensel T, Broekmeulen R, van Donselaar K, *et al*. Planogram integrity: a serious issue. Available at http://home.tm.tue.nl/tvwoense/files/PlanogramIntegrity.pdf. Accessed July 3, 2010.
35. Martin-Herran G, Taboubi S, Zaccour G. A time-consistent open-loop Stackelberg equilibrium of shelf-space allocation. Automatica 2005;41(6):971–982.
36. DeHoratius N, Raman A. Inventory record inaccuracy: an empirical analysis. Manage Sci 2008;54(4):627–641.
37. Raman A, DeHoratius N, Ton Z. Execution: the missing link in retail operations. Calif Manage Rev 2001;43(3):136–152.
38. Wolfe HB, Little AD. A model for control of style merchandise. Ind Manage Rev (now Sloan Manage Rev) 1968;9(2):69–82.
39. Larson PD, DeMarais RA. Psychic stock: an independent variable category of inventory. Int J Phys Distrib Logist Manage 1999;29:495–507.
40. Balakrishnan A, Pangburn MS, Stavrulaki E. Stack them high, let'em fly: lot-sizing policies when inventories stimulate demand. Manage Sci 2004;50(5):630–644.
41. Gupta R, Vrat P. Inventory model for stock-dependent consumption rate. Oper Res 1986;23(1):19–24.
42. Baker RC, Urban TL. A deterministic inventory system with an inventory-level-dependent demand rate. J Oper Res Soc 1988;39(9):823–831.
43. Urban TL. An inventory model with an inventory-level-dependent demand rate and relaxed terminal conditions. J Oper Res Soc 1992;43(7):721–724.
44. Urban TL. Inventory models with inventory-level-dependent demand: a comprehensive review and unifying theory. Eur J Oper Res 2005;162(5):792–804.
45. Padmanabhan G, Vrat P. EOQ models for perishable items under stock dependent selling rate. Eur J Oper Res 1995;86(2):281–292.
46. Giri BC, Pal S, Goswami A, *et al*. An inventory model for deteriorating items with stock-dependent demand rate. Eur J Oper Res 1996;95(3):604–610.
47. Sarker BR, Mukherjee S, Balan CV. An order-level lot size inventory model with inventory-level dependent demand and deterioration. Int J Prod Econ 1997;48(3):227–236.
48. Balkhi ZT, Benkherouf L. On an inventory model for deteriorating items with stock dependent and time-varying demand rates. Comput Oper Res 2004;31(2):223–240.
49. Giri BC, Chaudhuri KS. Deterministic models of perishable inventory with stock-dependent demand rate and nonlinear holding cost. Eur J Oper Res 1998;105(3):467–474.
50. Anderson ET, Fitzsimons GJ, Simester D. Measuring and mitigating the costs of stockouts. Manage Sci 2006;52(11):1751–1763.
51. Dana JD. Competition in price and availability when availability is unobservable. RAND J Econ 2001;32:497–513.
52. Dana JD Jr, Petruzzi NC. Note: the newsvendor model with endogenous demand. Manage Sci 2001;47(11):1488–1497.
53. van Donselaar KH, Gaur V, van Woensel T, *et al*. Ordering behavior in retail stores and implications for automated replenishment. Manage Sci 2010;56(5):766–784.
54. Kesavan S, Gaur V, Raman A. Do inventory and gross margin data improve sales forecasts for U.S. public retailers?. Manage Sci 2010;56(9):1519–1533.
55. Lee HL. Ultimate enterprise value creation using demand-based management. Working paper. Stanford Global Supply Chain Management Forum. Stanford University; 2001.
56. Keaveney SM. Customer switching behavior in service industries: an exploratory study. J Mark 1995;59(2):71–82.
57. Lippman SA, McCardle KF. The competitive newsboy. Oper Res 1995;45(1):54–65.
58. Hall J, Porteus E. Customer service competition in capacitated systems. Manuf Serv Oper Manage 2001;2(2):144–165.
59. Deneckere R, Peck J. Competition over price and service rate when demand is stochastic: a strategic analysis. RAND J Econ 1995;26(1):148–162.
60. Gans N. Customer loyalty and supplier quality competition. Manage Sci 2002;48(2):207–221.
61. Gaur V, Park Y. Asymmetric consumer learning and inventory competition. Manage Sci 2007;53(2):227–240.
62. Hardie BGS, Johnson EJ, Fader PS. Modeling loss aversion and reference dependence effects on brand choice. Mark Sci 1993; 12(4):378–394.

63. Bernstein F, Federgruen A. A general equilibrium model for industries with price and service competition. Oper Res 2004; 52(6):868–886.
64. Tsay A, Agrawal N. Channel dynamics under price and service competition. Manuf Serv Oper Manage 2000;2(4):372–391.
65. Netessine S, Rudi N, Wang Y. Inventory competition and incentives to back-order. IIE Trans 2006;38:883–902.
66. Li L. The role of inventory in delivery-time competition. Manage Sci 1992;38(2):182–197.
67. McGahan AM, Ghemawat P. Competition to retain customers. Mark Sci 1994;13(2):165–176.
68. Balachander S, Farquhar PH. Gaining more by stocking less: a competitive analysis of product availability. Mark Sci 1994;13(1):3–22.
69. Grahovac J, Chakravarty A. Sharing and lateral transshipment of inventory in a supply chain with expensive low-demand items. Manage Sci 2001;47(4):579–594.
70. Tagaras G. Pooling in multi-location periodic inventory distribution systems. Omega 1999;27:39–59.
71. Variety or duplication: a process to know where you stand. Washington, DC: The Research Department, Food Marketing Institute; 1993.
72. Bassok Y, Anupindi R, Akella R. Single-period multi-product inventory models with substitution. Oper Res 1999;47(4):632–642.
73. Gaur V, Honhon D. Assortment planning and inventory decisions under a locational choice model. Manage Sci 2007;52(10): 1528–1543.
74. Netessine S, Rudi N. Centralized and competitive inventory models with demand substitution. Oper Res 2003;51(2):329–335.
75. Parlar M, Goyal SK. Optimal ordering decisions for two substitutable products with stochastic demands. Opsearch 1984;21:1–15.
76. Ernst R, Kouvelis P. The effects of selling packaged goods on inventory decisions. Manage Sci 1999;45(8):1142–1155.
77. McGillivray A, Silver EA. Some concepts for inventory control under substitutable demands. INFOR 1978;16:47–63.
78. Pasternack B, Drezner Z. Optimal inventory policies for substitutable commodities with stochastic demand. Naval Res Logist 1991;38:221–240.
79. Needham PM, Evers PT. The influence of individual cost factors on the use of emergency transshipments. Transp Res E 1998; 34:149–160.
80. Lee YH, Jung JW, Jeon YS. An effective lateral transshipment policy to improve service level in the supply chain. Int J Prod Econ 2007;106:115–126.
81. Lee HL. A multi-echelon inventory model for repairable items with emergency lateral transshipments. Manage Sci 1987;33(10): 1302–1316.
82. Axsäter S. Modeling emergency lateral transshipments in inventory systems. Manage Sci 1990;36(11):1329–1338.
83. Das C. 1975. Supply and redistribution rules for two-location inventory systems: one period analysis. Manage Sci 1975;21(7): 765–776.
84. Sherbrooke CC. METRIC: a multi-echelon technique for recoverable item control. Oper Res 1968;16(1):122–141.
85. Archibald TW, Sassen SA, Thomas LC. An optimal policy for a two depot inventory problem with stock transfer. Manage Sci 1997; 43(2):173–183.
86. Dada M. A two-echelon inventory system with priority shipments. Manage Sci 1992;38(8):1140–1153.
87. Axsäter S. Evaluation of unidirectional lateral transshipments and substitutions in inventory system. Eur J Oper Res 2003;149: 438–447.
88. Axsäter S. New decision rule for lateral transshipments in inventory systems. Manage Sci 2003;49(9):1168–1179.
89. Kukreja A, Schmidt CP, Miller DM. Stocking decisions for low-usage items in multilocation inventory system. Manage Sci 2001;47(10):1371–1383.
90. Robinson LW. Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. Oper Res 1990;38(2):278–295.
91. Minner S, Silver EA. Evaluation of two simple extreme transshipment strategies. Int J Prod Econ 2005;93–94:1–11.
92. Rudi N, Kapur S, Pyke D. A two-location inventory model with transshipment and local decision making. Manage Sci 2001;47(12):1668–1680.
93. Dong L, Rudi N. Who benefits from transshipment? Exogenous vs. endogenous wholesale prices. Manage Sci 2004;50(5): 645–657.

94. Zhao H, Deshpande V, Ryan JK. Emergency transshipment in decentralized dealer networks: when to send and accept transshipment requests. Naval Res Logist 2006;53:547–567.

95. Chiou CC. Transshipment problems in supply chain systems: review and extensions. In: Kordic V, editor. Supply chain, theory and applications. Vienna: I-Tech Education and Publishing; 2008.