1

An Inventory-Theoretic Approach to Product Assortment and Shelf-Space Allocation

TIMOTHY L. URBAN
University of Tulsa

The purpose of this research is to generalize and integrate existing inventory-control models, product assortment models, and shelf-space allocation models. We first generalize the inventory-level-dependent demand inventory model to explicitly model the demand rate as a function of the displayed *inventory level. We then investigate the product assortment and shelf-space allocation problems by extending this model into the multi-item, constrained environment. A greedy heuristic and a genetic algorithm are proposed for the solution to the integrated problem.*

INTRODUCTION

It has long been acknowledged that displayed inventory has an effect on sales for many retail products. Over forty years ago, Whitin (1957) noted:

> For retail stores the inventory control problem for style goods is further complicated by the fact that inventory and sales are not independent of one another. An increase in inventories may bring about increased sales of some items. On the other hand an increased inventory might lead to a decrease in sales of other items.

Timothy L. Urban, Quantitative Methods and MIS Department, The University of Tulsa, 600 South College Avenue, Tulsa, OK 74104-3189, 918-631-2230 <urban@utulsa.edu>.

Journal of Retailing, Volume 74(1), pp. 15–35, ISSN: 0022-4359
Copyright © 1998 by New York University. All rights of reproduction in any form reserved.

Wolfe (1968) also presented empirical evidence that sales of style merchandise, such as women's dresses or sports clothes, are proportional to the amount of inventory displayed. Levin et al. (1972) note that the presence of inventory has a motivational effect on the customer. Schary and Becker (1972) state that the availability of the product is one of the roles of the distribution function in stimulating demand. Silver and Peterson (1985) observe that sales at the retail level tend to be proportional to inventory displayed. Larson and DeMarais (1990) call this phenomenon "psychic stock" and state that "psychic stock is retail display inventory carried to stimulate demand."

The marketing community has recognized this relationship and incorporated it into product assortment models and shelf-space allocation models. These models formulate the demand rate as a function of the shelf space allocated to the product and, sometimes, to the shelf space allocated to competing, substitute, and/or complementary products. Urban (1969) developed a model to identify which products should be included in a firm's product line; the model formulated the demand rate as a polynomial function of price, advertising, and distribution (represented by the number of shelf facings in the empirical application), considering both main and cross elasticities of the marketing variables, and was solved using an iterative search routine.

Corstjens and Doyle (1981) developed a shelf-space allocation model in which the demand rate is a function of shelf space allocated to the product, also applying a polynomial functional form of demand with main and cross-elasticities of shelf space; they utilized signomial geometric programming to solve the model. Zufryden (1986) suggested the use of dynamic programming to solve the shelf-space allocation problem, as it will allow the consideration of general objective-function specifications and will provide integer solutions. Bultez and Naert (1988) presented a model similar to that of Corstjens and Doyle, and utilized marginal analysis on a general theoretical formulation.

Anderson and Amato (1974) developed a model for the simultaneous determination of the product assortment and shelf-space allocation problems; they considered only the main effects and were able to formulate it as a knapsack problem. Borin, Farris, and Freeland (1994, 1995) presented the most comprehensive model to date; their integrated product assortment and shelf-space allocation problem incorporated cross-elasticity effects of substitute items as well as the effect on the demand of products when other products are not included in the assortment, and suggested simulated annealing as a solution methodology.

One drawback of the existing product assortment and shelf-space allocation models is that they give no explicit consideration to inventory-related decisions. Some include an operating-cost factor, which assumes that the costs are proportional to the sales of the product; however, these costs are independent of the order size, inventory levels, or frequency of ordering. Borin et al. (1994) use return on inventory as the objective and incorporate stockouts, but do not include the conventional inventory-control decisions as variables.

Another shortcoming is that this approach of modeling demand implicitly assumes that the shelves are always kept fully stocked. While some of the existing models use the number of facings to calculate demand and shelf inventory to determine how much demand can be satisfied, the effect of shelves that are not fully stocked is not explicitly reflected in the demand formulation. Consider, for example, the shelf display in Figure 1. If the demand of each of the two products is a function of the allocated shelf space, the demand realized by each of the products will depend on which product is allocated the empty space between

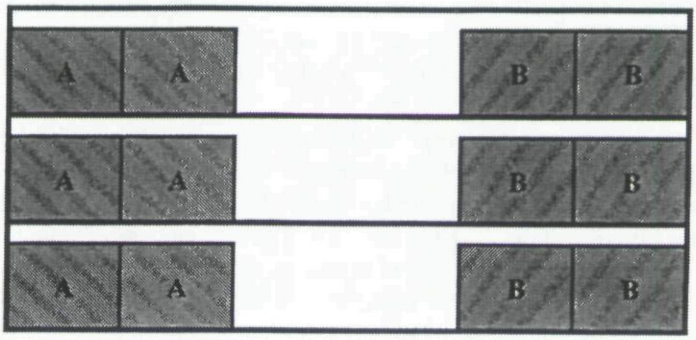

FIGURE 1

A Representative Shelf Display

the products, even if this allocation is not apparent to the customer. Using this rationale, a product with twelve depleted facings (no product on the shelf) allocated to it would realize a higher demand rate than if it only had six depleted facings. Even though the modeling of the problem becomes more complex, it would be more appropriate to model the demand rate as a function of the actual displayed inventory level.

The operations management literature has not been as receptive to modeling the effect of displayed inventories on demand. Baker and Urban (1988a) introduced a class of inventory models in which the demand rate is a function of the instantaneous inventory level of an item. Since then, this model has been extended to include the effect of deteriorating items (Mandal and Phaujdar, 1989; Padmanabhan and Vrat, 1995), random yield (Bar-Lev et al., 1994), two classes of customers (Datta and Pal, 1990; Urban, 1992), and the single-period model (Baker and Urban, 1988b). However, these models implicitly assume that all of the inventory (i.e., the entire order quantity) is displayed. This may be appropriate for some applications, such as used car sales, where the entire inventory can be seen by the customer. But many organizations (e.g., many retail outlets) have a backroom inventory or a warehouse in which the order is received, before being placed in the showroom. Thus, there is a limited amount of displayed inventory that has an effect on sales; much of the inventory is not in the customers' view and has no impact on sales. No inventory model has been developed to reflect this type of situation.

In summary, neither the existing inventory-theory literature nor the existing product assortment and shelf-space allocation literature adequately reflect the effect that displayed inventory has on the demand rate of many retail items. Furthermore, neither of these streams of research explicitly differentiate between the backroom inventory and the displayed inventory. And perhaps the greatest shortcoming of both streams of research is that they tend to ignore the other. Bregman (1995) discussed the need to integrate the marketing, operations, and purchasing functions to compete in today's marketplace. Silver (1981) enumerated several potential inventory-management research problems whose solution would have a major beneficial impact on the practice of inventory management; one of these is the allocation of shelf space in which sales are directly affected by the allocation. Also, Hayya (1991) states that some inventory management problems, "such as the interaction with marketing, as in the allocation of shelf space in supermarkets, remain a murky area."

Therefore, this research has two major objectives. First, we will present a generalization of the inventory-level-dependent demand inventory models that will explicitly model the demand rate as a function of the displayed inventory level—making an explicit distinction between the backroom and displayed inventories. The special case of full-shelf merchandising, in which the product is replenished as soon as the backroom inventory is depleted, will also be presented. We will then consider the product assortment and shelf-space allocation problems by extending this model into a multi-item, constrained environment. The model will incorporate the various inventory decisions, will distinguish between backroom and shelved inventory, and will reflect the effect on demand when the shelves are not kept fully stocked.

THE DISPLAYED-INVENTORY MODEL

In this section, we will generalize the inventory-level-dependent demand inventory model to explicitly consider the fact that the demand rate of many retail products is a function of the *displayed* inventory level. Thus, the demand rate of the product will be constant as long as the inventory level exceeds its shelf-space allocation; during this time, the backroom and/or warehoused inventory is being depleted. Once the backroom inventory level reaches zero, then the demand rate will decrease as the inventory level (i.e., the displayed inventory) decreases.

Assumptions and Notation

The model presented is a deterministic, continuous-review model of an inventory system in which the demand rate is a function of the displayed inventory level. The inventory is replenished instantaneously (i.e., the entire order is received simultaneously—it is not broken up into several orders and it does not come from a production facility at a continuous rate) with a known and constant lead time. The orders received are initially placed into "backroom" storage before being brought to the display area. There is a finite amount of display area dedicated to the product, and there is an associated cost of display space. With this type of model, there are three decision variables under the decision maker's control: the order quantity, the display quantity (the shelf-space allocation), and the reorder point; although the shelf-space allocation is frequently predetermined when the inventory and shelf-space decisions are made independently. We will use the following notation:

q is the order quantity,
r is the reorder point,

- s is the shelf space allocated to the item,
- i is the instantaneous inventory level of the entire system, including both the backroom storage and the displayed inventory,
- ϕ is the quantity of displayed inventory,
- d_ϕ is the demand rate of the item, a function of the displayed inventory,
- p is the selling price of the product,
- c is the unit cost of the product,
- C_h is the holding cost, based on the average inventory level, $\bar{i}$,
- C_o is the fixed procurement cost,
- C_s is the shelf-space cost, based on the space allocated to the product,
- π is the average net profit of the product,
- t is the time elapsed from the start of the inventory cycle,
- τ is the time at which the backroom inventory is depleted, and
- T is the cycle time; that is, the time at which the inventory is replenished.

Our analysis will concentrate on the situation in which the demand function is of a polynomial functional form:

$$d_\phi = \alpha\phi^\beta \qquad \alpha > 0, 0 < \beta < 1 \tag{1}$$

where α and β are the scale and shape parameters, respectively, of the demand function. This formulation is consistent with the existing product assortment literature (Urban, 1969), the existing shelf-space allocation literature (Corstjens and Doyle, 1981), as well as the existing inventory control literature (Baker and Urban, 1988a). The unit of measure of the product can either be in terms of objects (individual items, packages, cases) or in terms of space (facings, linear feet, area), as long as all of the cost and demand parameters reflect the appropriate measure.

Note that the shape parameter of the demand function reflects a constant elasticity, measuring the responsiveness of the demand rate to changes in the level of the displayed inventory, *ceteris paribus*; it does not measure an average response over some period of time, such as an initial, fully-stocked period with subsequent, partially-depleted periods. Also note that by constraining this parameter to be non-negative, we imply that there is a direct relationship between the displayed inventory and the demand rate; it may be conceivable that in some retail situations, partial depletion of inventory may actually stimulate demand.

Model Development

With this model, the amount of displayed inventory will be equal to the shelf-space allocation unless the inventory level falls below this level; that is, $\phi = \min\{s,i\}$. As shown in Figure 2, the inventory level will decrease linearly (from an initial level of $q + r$) until time τ, at which time the inventory level will not be sufficient to fully stock the allocated shelf space. The demand rate will then decrease, and the inventory level will decrease at a

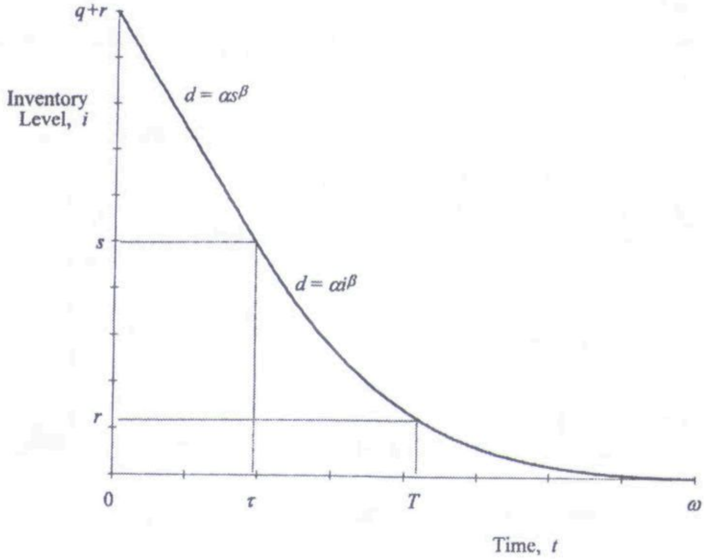

FIGURE 2

Inventory Level over Time—General Model

decreasing rate. The mathematical form of the inventory level over time is (the detailed derivation is given in Appendix I):

$$i_t = \begin{cases} (q + r) - \alpha s^{\beta} t & \text{for } 0 \le t \le \tau \\ \left[s^{1-\beta} - \alpha(1-\beta)(t-\tau)\right]^{1/(1-\beta)} & \text{for } \tau \le t \le \omega \\ 0 & \text{for } t \ge \omega \end{cases} \tag{2}$$

where

$$\tau = \frac{q + r - s}{\alpha s^{\beta}}$$

and

$$\omega = \tau + \frac{s^{1-\beta}}{\alpha(1-\beta)}$$

Note, ω is the time at which the inventory level would reach zero if no subsequent order is received. We make the following claim concerning the timing of the reorder:

Claim: *The optimal end of cycle, T^*, will satisfy the following condition: $\tau \le T^* \le \omega$.*

It has been shown that in the case of a deterministic, constant demand rate and instantaneous replenishment, it is optimal to let the inventory level reach zero before reordering (Silver and Peterson, 1985, p. 176). This same logic can be extended to the proposed model for $t \leq \tau$, since the demand rate is constant during this time; thus, we are assured that the reorder point will not be larger than the shelf-space allocation. We cannot extend this logic beyond τ (i.e., for $t > \tau$), since the demand rate is no longer constant; it is decreasing as the displayed inventory decreases. It may be desirable to reorder before the inventory level reaches zero, to avoid the effect of the decreased demand rate. Furthermore, we will not allow any time to expire with zero inventory as long as the item is profitable; it will always be preferable to place an order and have available stock for sale. Therefore, we can ignore backorders or lost sales in the deterministic model, although there are opportunity costs (essentially lost sales from the diminished demand rate) when the display area is not fully stocked.

The objective of the model will be to maximize the average net profit, π, comprised of the gross revenues less the unit item cost, the procurement cost, the holding cost, and the display cost. Profit maximization is appropriate (as opposed to the typical cost-minimization objective of most inventory models) since the decision variables directly impact the demand rate; the minimization of costs would lead a decision maker to try to reduce demand, which would be counterproductive for a profitable item. The profit function for the displayed-inventory model is:

$$\pi_{q,s,r} = \frac{(p-c)q}{T} - \frac{C_o}{T} - C_h \bar{i} - C_s s \tag{3}$$

As previously mentioned, we assume that the holding cost is based on the average inventory, including both the backroom or warehoused inventory and the displayed inventory; it is determined as follows:

$$\begin{aligned} \bar{i} &= \frac{\int_0^\tau [(q+r) - \alpha s^\beta t]\,dt + \int_\tau^T [s^{1-\beta} - \alpha(1-\beta)(t-\tau)]^{1/(1-\beta)}\,dt}{T} \\ &= \frac{(2-\beta)(q+r)^2 + \beta s^2 - 2r^{2-\beta}s^\beta}{2\alpha(2-\beta)s^\beta T} \end{aligned} \tag{4}$$

By replacing $\bar{i}$, τ, and T (Equations 4, I-2, and I-4, respectively) into Equation 3, the profit function can be expressed in terms of the decision variables:

$$\pi_{q,s,r} = \frac{\alpha(1-\beta)s^\beta[(p-c)q - C_o] - \frac{C_h(1-\beta)}{2(2-\beta)}[(2-\beta)(q+r)^2 + \beta s^2 - 2r^{2-\beta}s^\beta]}{(1-\beta)(q+r) + \beta s - r^{1-\beta}s^\beta} - C_s s \tag{5}$$

There will likely be several constraints associated with this problem. First, we must define the structural constraints from the model development to ensure the shelf-space allocation exceeds the reorder point and the reorder takes place before time ω ($s \geq r \geq 0$) and to ensure the replenishments bring the inventory at least up to the shelf-space allocation ($q + r \geq s$). There is also likely to be a maximum amount of on-hand inventory due to storage, production, or availability limits ($q \leq Q^{\max}$) or a minimum order quantity required from a supplier ($q \geq Q^{\min}$). Retailers frequently require a minimum amount of presentation stock for making the display look attractive, for providing an "image" for the retailer, or for a new product which needs a chance to make an impact ($s \geq \ell$). Finally, upper bounds may be set for products at a later stage of the life cycle to keep the store up-to-date ($s \leq \mu$). The ability to handle a variety of circumstances (different delivery cycles, minimum pack-outs, backroom space, etc.) makes this model quite appropriate in a number of retail applications.

With this set of constraints, a nonlinear programming solution methodology becomes necessary. While the solution to this problem appears quite prohibitive at first glance, it is a simple three-variable objective function (or, as previously mentioned, s is frequently predetermined leaving only two decision variables) with a linear constraint set. Baker and Urban (1988a) suggested the use of separable programming for this type of problem, as all of the functions are in polynomial form and easily separated with appropriate transformations. However, it was found that a generalized reduced gradient method is quite effective in finding the optimal solution to this model.

Example

To illustrate this process, consider the example from Baker and Urban (1988a) with the following parameters:

$\alpha = 0.5$ units per time period	$\beta = 0.4$
$p = \$20$ per unit	$c = \$10$ per unit
$C_o = \$10$ per order	$C_h = \$0.50$ per unit per time period

Also assume that we have a cost associated with the shelf space allocated to the product equal to $C_s = \$0.35$ per unit per time period. No constraints on maximum or minimum shelf space requirements or product availability were made. The optimal solution was found to be:

Order quantity, $q = 7.52$ units	Shelf-space allocation, $s = 4.65$ units
Reorder point, $r = 1.26$ units	Average profit, $\pi = \$3.23$ per time period

Since the problem is so small, we can easily require integer solutions to the problem as advocated by Zufryden (1986). The solution in this case would be q=7 units, s=4 units, and r=1 unit, resulting in an average profit of π=\$3.21, a decrease of less than one percent over the continuous solution.

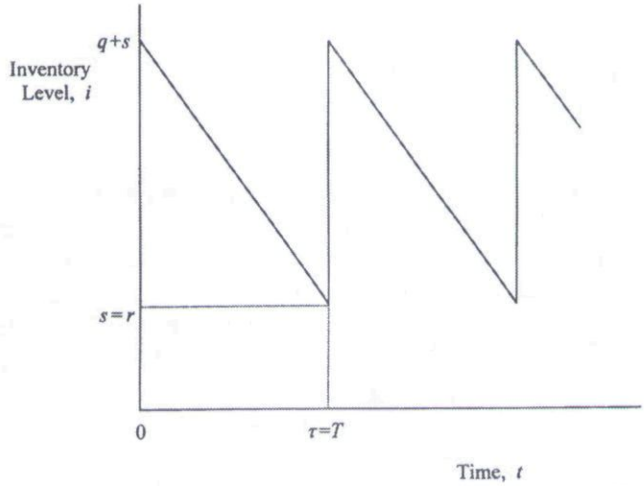

FIGURE 3

Inventory Level over Time—Full-shelf Merchandising Model

Full-shelf Merchandising Policy

Now consider the situation in which the retailer follows a "full-shelf merchandising" policy (see, e.g., Larson and DeMarais, 1990); that is, the display area is always kept fully stocked, so the inventory is replenished as soon as the backroom inventory reaches zero. In this case, the retailer is not willing to live with reduced sales; the displayed inventory will always be at its maximum value (i.e., the shelf space allocated to the product). In this situation, the demand is a function of the shelf space—no longer of the inventory level—so the demand will be constant once the shelf-space allocation has been determined, $d_s = \alpha s^{\beta}$, and the inventory level will decrease at a constant rate (see Figure 3). Note that this is essentially the underlying policy of existing shelf-space allocation models, as they formulate the demand rate in this manner.

We only have two decision variables under this scenario—the shelf-space allocation and the order quantity—as the reorder point is now defined to be equal to the shelf-space allocation. This policy simplifies the model considerably as the average inventory is now $\bar{i} = s + q/2$, the cycle time is $T = q/\alpha s^{\beta}$, and the average net profit function is:

$$\pi_{q,s} = \alpha s^{\beta}\left[(p-c) - \frac{C_o}{q}\right] - \frac{C_h q}{2} - (C_h + C_s)s \tag{6}$$

We can easily find the optimal solution to this problem; it is:

$$q = \left[\frac{2\alpha s^{\beta} C_o}{C_h}\right]^{1/2} \tag{7}$$

and

$$s = \left[\frac{\alpha\beta}{C_h + C_s}\left(p - c - \frac{C_o}{q}\right)\right]^{1/(1-\beta)}$$

This requires the solution of two simultaneous equations. Examination of the second derivatives shows this solution to be the global optimal if $(p - c)q > C_o$; that is, if the gross profit exceeds the ordering cost. If this does not occur, we would likely not be selling this item; in fact, we would also expect the gross profit to cover the holding and shelf-space costs. Also note that the order quantity has regressed to the typical EOQ solution under this policy.

Looking again at the example presented above, if we were to go to a full-shelf merchandising policy, the optimal solution would be an order quantity of q=5.57 units and a shelf-space allocation (reorder point) of s=r=2.99 units, resulting in an average profit of π=\$2.42, a reduction of nearly 25 percent. Obviously, a retailer should carefully weigh the consequences of going to this type of policy. Of course, profit maximization is not always the appropriate objective; retailers may be willing to realize a loss on some items for extended periods of time to build traffic for other items.

SHELF-SPACE ALLOCATION MODEL

We now extend the displayed-inventory model to the multi-item, constrained shelf space situation to incorporate the product assortment and shelf-space allocation decisions. In this situation, we will have several items competing for the limited shelf space. We will generalize the existing product assortment and shelf-space allocation models to incorporate inventory costs and decisions, to distinguish between displayed and backroom inventories, and to represent the demand as a function of displayed inventory levels (relaxing the implicit assumption of full-shelf merchandising made in previous models). The assumptions of the model will be as follows:

Assumption 1: *The inventory system involves a set of items, N={1, … , n}, and one stocking point with limited shelf space and a limited backroom capacity.*

Assumption 2: *Replenishments to the system are independent for each item (no joint replenishments), are sent directly to the backroom inventory, and are instantaneous with a known and constant lead time.*

Assumption 3: *The display area is partitioned into dedicated areas for each product—hence, no products use the area dedicated to another product, even if the inventory is insufficient to fill it completely—and is continuously replenished from the backroom inventory.*

Assumption 4: *The demand rate of the item is deterministic and is a known function of its displayed inventory level and of the displayed inventory level of other products to account for substitute or complementary products.*

Assumption 5: *The selling price and the unit cost of each item are known and constant (no price discounts and no effects of inflation).*

Assumption 6: *All relevant costs of each item (holding, procurement, and shelf-space costs) are known and constant.*

Assumption 7: *No backorders are allowed.*

Existing multi-item inventory models treat the capacity constraint in one of two ways (e.g., see Rosenblatt and Rothblum, 1990). The first approach assumes independent cycle times for each of the items and, since all products will eventually reach a simultaneous peak, planning is such that the available capacity is not exceeded at that point in time. The second approach assumes that the orders are phased within the cycle, and that they either have identical cycle times or that they are integer multiples of a basic cycle time. Our analysis is more in line with the first approach concerning the limited shelf space, since we are dedicating shelf space to the products—our model assumes that no other products use the space dedicated to an item. So, the shelf-space capacity constraint will be a function of the shelf-space allocation level (consistent with the existing shelf-space allocation models) and not the instantaneous inventory. For consistency, we assume independent cycle times, so the backroom capacity constraint will be a function of the maximum inventory level of each item $(q+r)$.

Model Development

The typical demand formulation used in space-allocation models is the polynomial form; the demand for any product j is represented by:

$$d_j = \alpha_j \phi_j^{\beta_j} \left[\prod_{k \in N^+} \phi_k^{\delta_{jk}} \right] \left[1 + \sum_{i \in N^-} (1 - \lambda_{ij}) f(\alpha_i, \delta_{ij}) \right] \tag{9}$$

$$\alpha_j > 0,\ 0 < \beta_j < 1,\ 0 \le \lambda_{ij} \le 1$$

where β_j is the space elasticity for product j with respect to a unit of displayed inventory, δ_{jk} is the cross space elasticity of products j and k, λ_{ij} represents the resistance of customers to switch their purchase choice from an unstocked item to one included in the product assortment, and N^+ (N^-) is the set of products included (excluded) in the assortment. The value of δ_{jk} can be positive for complementary products and negative for substitute products ($\delta_{jj}=0$) and, while it can theoretically take on any value, is expected to be fairly small, certainly in the range from -1 to $+1$. Note the last term (in brackets) accounts for additional demand realized from those consumers that are willing to purchase the item if their preferred item is not included in the assortment. Borin et al. (1994) note that this 'acquired demand' is likely a function of the potential demand (α_i) and the degree of substitutability with the missing item (δ_{ij}); studies such as that of Emmelhainz (1991) have investigated consumer response to unavailable products.

Since d_j is a function of the displayed inventory level of all items included in the assortment, the demand rate of an individual item will change for every instance of the inventory level of any other item falling below its shelf-space allocation. To simplify the evaluation of d_j, we will approximate it such that it is a function of the shelf-space allocation, not the instantaneous inventory level, of all other items in the assortment by replacing ϕ_k with s_k in Equation (9). This implies that the demand of an item is a function of its displayed inventory given the shelf-space allocation of all other items in the assortment, $d_j = f(\phi_j \mid s_k, k \in N^+)$. Utilizing this estimation, we can then follow the procedure of the displayed-inventory model from the previous section to determine the inventory function, average inventory, etc. The average net profit of each item would be:

$$\pi_j = \frac{\alpha_j(1-\beta_j)s_j^{\beta_j}[(p_j - c_j)q_j - C_{o,j}]\left[\prod_{k \in N^+} s_k^{\delta_{jk}}\right]\left[1 + \sum_{i \in N^-}(1-\lambda_{ij})f(\alpha_i, \delta_{ij})\right]}{(1-\beta_j)(q_j + r_j) + \beta_j s_j - r_j^{1-\beta_j} s_j^{\beta_j}}$$
$$- \frac{\frac{C_{h,j}(1-\beta_j)}{2(2-\beta_j)}[(2-\beta_j)(q_j + r_j)^2 + \beta_j s_j^2 - 2r_j^{2-\beta_j} s^{\beta_j}]}{(1-\beta_j)(q_j + r_j) + \beta_j s_j - r_j^{1-\beta_j} s_j^{\beta_j}} - C_{s,j} s_j \quad (10)$$

If we assume that all of the items will be in the product assortment (as did Corstjens and Doyle, 1981), then the objective will simply be the sum of the individual profit functions. By incorporating the product assortment decision, however, we must also formulate the model to determine whether a particular item should remain in the assortment. Since we have now incorporated the inventory decisions into the model, we cannot simply let the decision variables go to zero to remove items from the product assortment, as the procurement cost will get very large for very small values of q. To account for the assortment decision, we introduce binary variables to represent whether or not a particular product is included:

$$z_j = \begin{cases} 1 & \text{if product } j \text{ is included in the product assortment} \\ 0 & \text{otherwise} \end{cases}$$

The nonlinear programming formulation of the inventory-theoretic approach to the product assortment and shelf-space allocation problem can then be expressed as follows:

$$\text{Maximize} \qquad \pi_N = \sum_{j \in N} z_j \pi_j \quad (11)$$

$$\text{subject to} \quad \sum_{j \in N} b_j s_j \le B_s \tag{12}$$

$$\sum_{j \in N} b_j (q_j + r_j) \le B_b \tag{13}$$

$$B_s z_j \ge s_j \qquad \forall j \tag{14}$$

$$B_b z_j \ge r_j \qquad \forall j \tag{15}$$

$$B_b z_j \ge q_j \qquad \forall j \tag{16}$$

$$q_j \le Q_j^{\max} \qquad \forall j \tag{17}$$

$$q_j \ge Q_j^{\min} \qquad \forall j \tag{18}$$

$$r_j \le s_j \le q_j + \lambda \qquad \forall j \tag{19}$$

$$\ell_j \le s_j \le \mu_j \qquad \forall j \tag{20}$$

$$z_j = \{0,1\}, q_j, s_j, r_j \ge 0 \qquad \forall j \tag{21}$$

where π_N is the total profit of the system, π_j is the average net profit of item j, b_j is the amount of shelf space required per unit of product j, B_s is the total availability of shelf space, and B_b is the total capacity of the backroom inventory. The use of the parameter, b_j, allows different units of measure (if desired) for the shelf-space and backroom capacities than that used for the product. This formulation contains $3n$ continuous and n binary variables. For even moderately-sized problems, a mixed-integer nonlinear program can be quite prohibitive to solve, particularly if we require integer solutions for q_j, s_j, and r_j. Furthermore, the necessity of finding the optimal solution to this model is dubitable, since "managers typically do not have access to error-free estimates of the parameters required for the model" (Borin and Farris, 1995). We therefore turn our attention to heuristic solution methodologies for this problem.

Heuristic Solution Methodologies

In developing heuristics for the product assortment and shelf-space allocation problem, we will focus on the search for the optimal product assortment. Since there are 2^n possible assortment combinations for any given problem, we obviously need efficient methods to identify the most profitable product assortments. For example, a 40-item problem will have over one trillion possible combinations. Fortunately, for a given $Z=(z_1,z_2,\ldots z_n)$, the profit curve appears to be quite well behaved. Thus, once a particular assortment is selected, techniques such as the generalized reduced gradient (GRG) algorithm are quite effective in determining appropriate values of the shelf-space allocation, the order quantity, and the reorder point for each of the items in the assortment. We present two heuristics for solving the product assortment and shelf-space allocation problem: a greedy heuristic and a genetic algorithm.

The first approach that we will consider is a greedy search heuristic. The logic behind this heuristic is that we will remove products from the assortment as long as doing so improves the overall profitability. Thus, an initial solution is obtained using the GRG algorithm with all items included in the product assortment. Subsequent product combinations are determined by removing items, one at a time, that contribute the least (or have the greatest net loss) to the total profit, and the GRG algorithm is solved for the subsequent product assortment. This iterative process is continued as long as an improvement in profit is realized. The greedy heuristic can be expressed as follows:

Step 1: *Input data:*

- Demand parameters: $\alpha_j, \beta_j, \delta_{jk}, \lambda_{ij}$
- Revenue and cost parameters: $p_j, c_j, C_{h,j}, C_{o,j}, C_{s,j}$
- Space parameters: $b_j, B_s, B_b, \mu_j, \ell_j$

Step 2: Solve the displayed-inventory problem using the GRG algorithm with *all* items included in the product assortment.

Step 3: Remove the item with the smallest profit from the product assortment.

Step 4: Solve the displayed-inventory problem using the GRG algorithm with the resulting assortment.

Step 5: Repeat *Steps 3* and *4* until the total profit decreases.

This heuristic is designated a greedy heuristic since we are only considering the removal of the least profitable item at each iteration. Another approach would be a "steepest-ascent" type of search in which we would evaluate every item to determine the effect on profitability of removing that item from the product assortment or replacing those items not included in the assortment. This approach, however, would require considerably more computational effort as the GRG method would need to be solved n times at each iteration. For larger problems, this would require considerable computational effort.

One of the major advantages of this heuristic is that it can be easily implemented for small- to moderately-sized problems. Since the GRG algorithm is included in several spreadsheet packages, no special software is required. It will require the solution of $n - n^* + 2$ GRG problems (where n^* is the number of items in the product assortment of the final solution). As the size of the problem increases, however, this method can become fairly tedious to solve.

The second heuristic that we will present for the product assortment and shelf-space allocation problem is a genetic algorithm (GA). Genetic algorithms are search methods based on natural selection and evolution. From a set of initial solutions, the 'survival-of-the-fittest' notion leads us to superior solutions that dominate the population. Thiel and Voss (1994) note that "one of the most crucial ideas for a successful implementation of a GA is the representation of an underlying problem by a suitable scheme…the binary representation is considered to be the most general." Thus, we would expect the use of genetic algorithms to be quite appropriate for this type of problem, as the decision variables for product assortment decision can be represented as a binary string; for example, Z_1=(1, 0, 1, 1, 0, 1) indicates stocking products 1, 3, 4, and 6.

In the proposed GA implementation, each individual will represent a given product assortment combination; that is, a given Z-vector. To initialize the algorithm (and to calculate the fitness of an individual, represented by the total profit), starting values of q_j, s_j, and r_j are needed. Existing rules of thumb for space allocations and inventory decisions could be used as a starting point but, as we will see, these rules tend to perform very poorly. We suggest finding a starting point by solving the GRG algorithm with all of the items in the product assortment (or by solving n single-item problems) while ignoring the shelf-space and backroom constraints. To calculate the fitness (total profit) for each individual in the population, these values can then be proportionately scaled down to satisfy the constraints for a particular product assortment. Alternatively, we can solve a GRG algorithm for each individual incorporating the shelf-space and backroom constraints; however, it is generally not worth the incremental calculations required due to the insensitivity of the model to the values of q_j, s_j, and r_j for a given product assortment.

For brevity, we will not discuss the particulars of genetic algorithms; Hurley et al. (1995) discuss the fundamentals of genetic algorithms and present several marketing applications. The procedure for the genetic algorithm can be expressed as follows:

Step 1: *Initialization.* Input data—demand parameters (α_j, β_j, δ_{jk}, λ_{ij}), revenue and cost parameters (p_j, c_j, $C_{h,j}$, $C_{o,j}$, $C_{s,j}$), and space parameters (b_j, B_s, B_b, μ_j, ℓ_j). Determine initial values of q_j, s_j, and r_j. Randomly create an initial population of P solutions, each consisting of a binary string of size n.

Step 2: *Fitness.* Calculate the total profit for each individual in the population. If the constraints are satisfied for the given Z-vector, use the initial values of q_j, s_j, and r_j. If the shelf-space constraint is violated, multiply each s_j by a factor of $B_s/\Sigma b_k s_k$. If the backroom constraint is violated, multiply each q_j and r_j by a factor of $B_b/\Sigma b_k(q_k + r_k)$.

Step 3: *Reproduction.* Calculate the relative fitness of each individual ($\pi_j/\Sigma\pi_k$). If any individuals represent a net loss ($\pi_j < 0$), assign a profit of zero. Retain the individual with the highest relative fitness, and use the relative fitness values to randomly select the offspring for the remainder of the next generation.

Step 4: *Crossover.* Randomly choose two individuals from the pool. Randomly choose a point at which the two strings will be cut (from 1 to $n - 1$). With a given probability (crossover rate), randomly determine whether a crossover will be performed. If so, exchange the tails of the two strings. Repeat for all strings.

Step 5: *Mutation.* With specified probability (mutation rate), randomly select a particular bit of a particular string and change its value—from 0 to 1, or from 1 to 0.

Step 6: *Termination criterion.* Repeat *Steps 2–5* until the population is fully converged (i.e., all individuals are identical). Alternatively, we may repeat for a given number of generations without improvement or for a given total number of generations. If desired, use the GRG algorithm with the final product assortment to ensure optimal values of q_j, s_j, and r_j.

At first glance, there appears to be a considerable amount of randomness in the algorithm. This is one of the distinguishing features of genetic algorithms and helps to avoid convergence to a local minimum. Still, those individuals (solutions) with the greatest profit will tend to dominate the process—the natural selection process will evolve to a superior solution.

Example

To illustrate this process, we will consider the six-product example of Borin et al. (1994). The demand and shelf-space parameters from their paper and values for the other revenue and inventory cost parameters are as follows:

			$\delta_{j,k}$						
Item, j	α_j	β_j	1	2	3	4	5	6	c_j
1	28.53	.1532	—	–.0630	–.0100	–.0089	–.0101	–.0250	3.00
2	23.62	.2273	–.0480	—	–.0159	–.0303	–.0101	–.0010	1.50
3	25.59	.2089	–.0232	–.0463	—	–.0504	–.0280	–.0120	4.25
4	22.40	.2143	–.0242	–.0606	–.0628	—	–.0300	–.0240	1.75
5	15.62	.2955	–.0130	–.0571	–.0165	–.0296	—	–.0580	6.00
6	10.50	.3104	–.0125	–.0543	–.0221	–.0239	–.0740	—	0.75

μ_j = 12 units
Price markup of 80 percent ($p_j = 1.8c_j$)
B_s = 24 facings
C_{oj} = \$50 per order
C_{hj} = 25 percent of unit cost per time period ($C_{hj} = 0.25c_j$)
ℓ_j = 0 units (=1 if included in the assortment)
b_j = 1 facing per unit
B_b = 240 facings
C_{sj} = \$1.00 per unit per time period
λ_{ij} = 1

For a six-item problem, there are 64 possible product combinations. Conducting an exhaustive search (solving the GRG algorithm for all 64 combinations) provides the optimal solution of stocking items 1, 3, and 5; resulting in a total profit of \$138.04. While this problem is fairly insensitive to values of q_j, s_j, and r_j for a given product assortment, it is considerably more sensitive to the product assortment decision (the values of z_j). Only three other assortment combinations result in a solution within ten percent of optimal—again emphasizing the need for heuristics to focus on the product-assortment decision—and each of these solutions simply adds a fourth item to the assortment. Both the greedy heuristic and the genetic algorithm provide the optimal solution to this problem.

In order to further test the proposed heuristics, additional problems with 18 and 54 items were generated. Of the existing product assortment and shelf-space allocation papers, Anderson and Amato (1974) only considered main effects and Borin et al. (1994) only considered substitute goods. Thus, we will also extend our analysis to consider problems in which the composition includes both substitute and complementary products. The specific parameters used in the genetic algorithm implementation are as follows: crossover rate = 0.6; mutation rate = 0.001; initial population, $P = 20$, 100, and 200 for the $n = 6$, 18, and 54 problem sizes, respectively; and the number of generations to termination = 500 (however, in all cases, the solution was obtained well before this limit, and the populations fully converged within 31 iterations for the $n = 6$ problems).

As seen from Table 1, the genetic algorithm provided the best-known solution to five of the six problems. On the problem in which it did not find the best solution, it was within 0.02 percent. The greedy heuristic also did very well, finding the best solution on four of the six problems; on the problems that it did not find the best solution, it was within 0.40 percent. Given the likely accuracy of the parameter estimates, these methods provide exceptional solutions.

These results are also compared to a common rule of thumb for shelf-space allocation, in which the space allocated to a particular item is proportional to its sales. Since this rule does not determine the product assortment, items incurring a net loss were removed from the product assortment. As illustrated in the table, this rule provides profits from 15 to 25 percent below the solutions to the proposed heuristics. Obviously, following such a strategy could result in substantial losses for the retailer. The use of the proposed methodologies can provide considerable improvement in these important retailing decisions.

CONCLUSION

Many researchers are recognizing the need for integrative research in marketing and operations management (e.g., Karmarkar, 1996). This research generalizes and integrates two distinct streams of research—inventory management and shelf-space allocation—that are quite interdependent in retail organizations. This represents a contribution to the operations management literature by developing inventory models that incorporate the effect of displayed inventory on demand, explicitly distinguish between backroom and displayed inventories, and extend to the multi-item, constrained environment. It advances the marketing literature by incorporating inventory costs and decisions into shelf-space allocation decisions, including the effect of less-than-full shelf stocking on demand, as well as distinguishing between backroom and displayed inventories. And perhaps the most important contribution is the integration of these two important elements of retailing, developing relevant and applicable models and providing effective solution methodologies to the integrated problem. Given the investments generally tied up in retail inventories, the effective management and control of inventory is essential for retailers.

While this model is a significant generalization of existing models, there are a number of areas in which the modeling and analysis can be improved. For example, the incorporation of stochastic demand would make the model more realistic—particularly for the single-product model—and help retailers better determine when a product should be shifted from warehouse to distributor and evaluate various trade-offs, such as smaller delivery quanti-

TABLE 1

Relative Performance of the Proposed Heuristics and a Common Rule of Thumb

		Greedy Heuristic			Genetic Algorithm			Rule of Thumb ($s \approx d$)		
Composition	Size of Problem	No. of Items in Final Assortment	Total Profit	Deviation from Best Solution	No. of Items in Final Assortment	Total Profit	Deviation from Best Solution	No. of Items in Final Assortment	Total Profit	Deviation from Best Solution
Substitutes only	6	3	138.04	0.00%	3	138.04	0.00%	3	103.42	25.08%
	18	15	701.57	0.36%	14	704.12	0.00%	14	598.08	15.06%
	54	35	1627.18	0.00%	35	1626.87	0.02%	32	1333.08	18.07%
Substitutes and Complements	6	3	166.47	0.00%	3	166.47	0.00%	3	130.60	21.55%
	18	13	564.04	0.00%	13	564.04	0.00%	12	461.70	18.14%
	54	34	1578.23	0.03%	34	1578.74	0.00%	30	1308.87	17.09%

ties. Also, the analysis of different types of inventory models, such as periodic review models, in addition to the continuous review model would be appropriate. It would be beneficial to incorporate pricing or other marketing variables into the model, as well as the effect of trade and consumer promotions on shelf space. Concerning the heuristics, we present a basic—albeit effective—implementation of genetic algorithms; further work may focus on a more state-of-the-art approach, such as Balakrishnan and Jacob's (1996) use of GAs for product design. Other heuristics, such as simulated annealing may prove to be advantageous. Finally, and perhaps most importantly, further empirical research is warranted to overcome the obstacles in obtaining good estimates of the parameters in practice (Borin et al. (1994) provide a discussion of these difficulties) and to provide a more empirically rich context of application.

APPENDIX I
DERIVATION OF THE INVENTORY FUNCTION

For $0 \leq t \leq \tau$:

The rate of change of the inventory level will be equal to $-d_\phi$ which, at this point, is constant since the shelf space is fully stocked:

$$\frac{di}{dt} = -d_\phi = -\alpha s^\beta$$

$$\int di = \int -\alpha s^\beta dt$$

$$i = -\alpha s^\beta t + k$$

when $t = 0$, $i = q + r$;
so $k = q + r$, and

$$i = (q + r) - \alpha s^\beta t \tag{I-1}$$

At time τ, the inventory level will be s, so:

$$s = (q + r) - \alpha s^\beta \tau$$

and

$$\tau = \frac{q + r - s}{\alpha s^\beta} \tag{I-2}$$

For $\tau \leq t \leq \omega$:

During this time, the inventory in the backroom will be depleted, the inventory in the display area will then be decreasing, and the demand will now be a function of the instantaneous inventory level:

$$\frac{di}{dt} = -d_{\phi} = -\alpha i^{\beta}$$

$$\int i^{-\beta} di = \int -\alpha dt$$

$$i^{1-\beta} = -\alpha(1-\beta)t + k$$

when $t = \tau$, $i = s$;
so $k = s^{1-\beta} + \alpha(1-\beta)\tau$, and

$$i = [s^{1-\beta} - \alpha(1-\beta)(t-\tau)]^{1/(1-\beta)} \tag{I-3}$$

At time T, the inventory level will be r, so:

$$r = [s^{1-\beta} - \alpha(1-\beta)(T-\tau)]^{1/(1-\beta)}$$

and

$$T = \tau + \frac{s^{1-\beta} - r^{1-\beta}}{\alpha(1-\beta)} \tag{I-4}$$

Furthermore, at time ω, the inventory level will be 0; so:

$$0 = [s^{1-\beta} - \alpha(1-\beta)(\omega-\tau)]^{1/(1-\beta)}$$

and

$$\omega = \tau + \frac{s^{1-\beta}}{\alpha(1-\beta)} \tag{I-5}$$

REFERENCES

Anderson, Evan E. and Henry N. Amato. (1974). "A Mathematical Model for Simultaneously Determining the Optimal Brand-Collection and Display-Area Allocation," *Operations Research*, **22**(1): 13-21.

Baker, R. C. and Timothy L. Urban. (1988a). "A Deterministic Inventory System with an Inventory-Level-Dependent Demand Rate," *Journal of the Operational Research Society*, **39**(9): 823-831.

______. (1988b). "Single-Period Inventory Dependent Demand Models," *Omega*, **16**(6): 605-607.

Balakrishnan, P. V. and Varghese S. Jacob. (1996). "Genetic Algorithms for Product Design," *Management Science*, **42**(8): 1105-1117.

Bar-Lev, Shaul K., Mahmut Parlar, and David Perry. (1994). "On the EOQ Model with Inventory-Level-Dependent Demand Rate and Random Yield," *Operations Research Letters*, **16**(3): 167-176.

Borin, Norm, Paul W. Farris, and James R. Freeland. (1994). "A Model for Determining Retail Product Category Assortment and Shelf Space Allocation," *Decision Sciences*, **25**(3): 359-384.

Borin, Norm and Paul Farris. (1995). "A Sensitivity Analysis of Retailer Shelf Management Models," *Journal of Retailing*, **71**(2): 153-171.

Bregman, R. L. (1995). "Integrating Marketing, Operations, and Purchasing to Create Value," *Omega*, **23**(2): 159-172.

Bultez, Alain and Philippe Naert. (1988). "SH.A.R.P.: Shelf Allocation for Retailers' Profit," *Marketing Science*, **7**(3): 211-231.

Corstjens, Marcel and Peter Doyle. (1981). "A Model for Optimizing Retail Space Allocations," *Management Science*, **27**(7): 822-833.

Datta, T. K. and A. K. Pal. (1990). "A Note on an Inventory Model with Inventory-Level-Dependent Demand Rate," *Journal of the Operational Research Society*, **41**(10): 971-975.

Emmelhainz, Margaret A., Jan

Copyright of Journal of Retailing is the property of Elsevier Science Publishing Company, Inc.. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.