Inventory Models for Substitutable Products:

Optimal Policies and Heuristics

Mahesh Nagarajan
Operations and Logistics Division
Sauder School of Business
University of British Columbia
Vancouver, BC V6T 1Z2
Canada
mahesh.nagarajan@sauder.ubc.ca

S. Rajagopalan
Information and Operations Management Department
Marshall School of Business
University of Southern California
Los Angeles, CA 90089-1421
Phone: (213) 740-0193
srajagop@marshall.usc.edu

Revised June 2006
Revised February 2007
Abstract

In this paper, we examine the nature of optimal inventory policies in a system where a retailer manages substitutable products. We first consider a system with two products 1 and 2 whose total demand is D and individual demand proportions are \( p \) and \( (1-p) \). A fixed proportion of the unsatisfied customers for 1(2) will purchase item 2 (1), if it is available in inventory. For the single period case, we show that the optimal inventory levels of the two items can be computed easily and follow what we refer to as "partially decoupled" policies, i.e. base stock policies that are not state dependent, in certain critical regions of interest both when D is known and random. Furthermore, we show that such a partially decoupled base-stock policy is optimal even in a multi-period version of the problem for known D for a wide range of parameter values. Using a numerical study, we show that heuristics based on the de-coupled inventory policies perform well in conditions more general than the ones assumed to obtain the analytical results. The analytical and numerical results suggest that the approach presented here is most valuable in retail settings for product categories where there is a moderate level of substitution between items in the category, demand variation at the category level is not too high and service levels are high.

Keywords: Inventory/Production: stochastic multi-period inventory models; stockout-based substitution.
1. Introduction

This paper explores the impact of consumer-driven substitution due to stockouts on the inventory decisions of a product category at a retailer. Consider a retailer who manages the inventory of two items that are partial substitutes, for example Duracell and Energizer size C batteries. Since the demand for these items is random, the retailer may stock out of one or both of them and some unsatisfied customers for an item will purchase the other one if it is available in stock. The retailer has to consider these substitution effects in addition to the traditional tradeoff between overstocking and understocking in determining stocking levels. Several papers in the inventory literature have considered substitution effects, as will be clear later in the literature review. We consider a stylized model that yields new structural results and insights, which are used to develop a simple but effective heuristic to make inventory decisions for substitutable items within a product category. We believe our approach can be useful to retailers of staple products such as batteries, grocery items, and office products. Our goal is to develop some simple approaches that work well in scenarios where there is substitution behavior and to identify conditions when it is necessary to consider such substitution effects.

Consider a scenario with two substitutable products, say 1 and 2, whose total demand in a period is D (D may be random) where D is, say, traffic at a store or demand for a particular product category. The individual demand proportions of the two products are $p$ and $(1-p)$ where $p$ is random and the average value of $p$ is the market share of product 1 at the retailer. A fixed proportion $\gamma$ of unsatisfied customers for a product will purchase the other one if it is available in stock. The retailer has to decide the inventory for the two items. We analyze this model and its variants in both single and multi-period scenarios. The main contributions of the paper are as follows. First, we show that for reasonable parameter values, the optimal inventory (or base stock) level of product 1 is independent of the inventory of 2, when the latter is above its base-stock level—we call these “partially decoupled” policies. Thus, even if product 2’s inventory is very high, it is optimal to raise the inventory of product 1 to its base-stock level. Further, the optimal inventory levels can be computed easily using closed-form formulas for any parameter value if D is known. Second, we identify conditions when such a result holds even if D is random. Third, we show that
the “partially decoupled policies” are optimal in the multi-period two-product model under reasonable conditions on certain parameter values. Finally, we show numerically that a heuristic based on the two-product result is robust and performs well under conditions more general than the ones assumed to obtain the analytical results and when the number of products is more than two. In particular, we use data from numerous academic and industry studies on service levels, substitution behavior, and demand characteristics to evaluate the performance of the heuristic and to point out when our approach works well and when it is necessary to consider such substitution effects. The approach developed here is valuable for many product categories in retail settings, as discussed in detail later.

1.1 Literature review

There is a substantial literature in the area of inventory management. In particular, several papers have investigated inventory policies when there is stockout-induced substitution (Mahajan and Van Ryzin (1998)). The inventory literature with substitution effects and a single-decision maker can be broadly classified into two categories, depending on whether the substitution is driven by the consumer or by the decision-maker. The decision-maker driven substitution literature is also referred to as the transshipment literature since substitution in such a setting may involve substitution of products or inventory transfer or transshipment between different locations. Examples of this literature include Robinson (1991), who examines a multi-product, multi-period transshipment problem and characterizes the structure of the optimal policy and Bassok, Anupindi and Akella (1999), who consider a model with full downward substitution, show that the optimal policy is a base stock policy and provide an algorithm to compute optimal inventory levels. Recent examples of papers in this literature are Rudi, Pyke and Kapur (2001), Van Mieghem and Rudi (2002), Axsater (2003), and Deniz, Scheller-Wolf and Karaesmen (2004).

There is a fundamental difference between the above setting and consumer-driven substitution in retail settings, which is the focus of this paper because when the decision-maker makes substitution decisions, the level of substitution $\gamma$ referred to earlier, is not exogenous. Also, unlike in consumer-driven substitution settings, it is typically optimal to substitute between two products (or locations) fully or not at all, assuming there is sufficient inventory. This is reinforced by Mahajan and Van Ryzin (2001),
“Substitution in retail settings is fundamentally different. Substitution decisions are not directed by the retailer.” So, we focus in this review on papers that consider consumer-driven substitution. In any case, none of the above works identify the type of partially decoupled policies discussed in this paper.

Models with substitution effects tend to be inherently hard. As Mahajan and Van Ryzin (1998) point out, dynamic substitution models, where the substitution behavior depends on the inventory status at the time a customer makes his choice (as in our case), are more realistic but are typically less tractable and the profit functions are in general quite complex and not easily amenable to analysis. McGillivary and Silver (1978) develop heuristic policies for the two-product problem when there is partial substitution. Parlar and Goyal (1984) also study the partial substitution problem and show that the expected profit function is concave. Parlar (1985) considers a perishable inventory model with substitution between new and old items when there is a stockout. Pasternack and Drezner (1991) show some interesting characteristics of optimal inventory levels in a two-product problem with full substitution. Gerchak and Mossman (1992) show that, contrary to intuition, substitution effects may result in an increase in optimal inventory levels under certain cost and demand conditions. The above papers consider single-period models where demands of the products are not correlated, unlike in our work.

There are a few recent papers that study inventory policies with substitution effects when demands are correlated. One of them is Rajaram and Tang (2001), where the impact of partial substitution and correlation in demand is studied in a single-period model. They indicate that characterizing exact optimal policies analytically is complex and develop heuristics to compute inventory levels when demand is a bivariate normal distribution. More recently, Yang and Schrage (2002) show that increased substitution may result in higher inventory levels under certain conditions and discuss other interesting anomalies related to risk pooling. Ernst and Kouvelis (1999) study a problem of packaged goods with two substitutable products that may be sold singly or as a bundle and perform a numerical analysis on the effects of correlation on the optimal policy. Netessine and Rudi (2003) study an N-product substitution problem where N products have random demands that are correlated and a fixed fraction of customers facing a stock out of a product will substitute (if possible) to another product. They show that in general,
this substitution problem need not be concave and perhaps not even unimodal. They analyze models with and without retail competition and show that optimal inventory levels are higher in the competitive scenario. The above papers are restricted to single-period scenarios, unlike our work, and illustrate the difficulty in solving substitution problems with or without demand correlation.

There is another recent stream of literature on assortment planning, representative early works being Smith and Agrawal (2000) and Mahajan and Van Ryzin (2001) which study inventory systems where consumer choice and substitution effects are modeled in great detail and substitution in out-of-stock situations is probabilistic. The focus of this work is on determining both optimal product assortments as well as inventory levels of the items as a function of consumer purchase behavior. In recent work, Kok and Fisher (2004) consider an assortment planning problem taking into account substitution effects and provide structural results and an iterative heuristic that performs very well. They also discuss the implementation of the model at a supermarket chain. Gaur and Honhon (2005) study a single-period assortment planning and inventory management problem that uses a locational choice model of consumer choice. They obtain bounds on profit when customers dynamically substitute and identify conditions under which static substitution serves as a good approximation. Kok, Fisher and Vaidyanathan (2006) provide a comprehensive review of the literature on assortment planning.

While the above papers present numerous insights and some of them consider models that are richer than ours in terms of the modeling of substitution and demand and some of them consider decisions such as product assortment that we do not consider, our work yields insights not found in these works. Specifically, none of the prior works have shown the optimality of “decoupled” inventory policies in single or multi-period models in any type of scenario. Further, we show numerically that the decoupled policies perform well even when some of the underlying assumptions are relaxed.

The organization of this paper is as follows. In section 2, we present the single period, two-product model in great detail. We establish some properties of the profit function and characterize the optimal inventory levels, providing closed form expressions. In section 3, we present our analysis of the finite and infinite horizon two-product model and characterize the optimal policy. In section 4, we present
the results of a computational study and finally in section 5, we summarize and present potential extensions of this research. Proofs of all the results are in the Appendix.

2. The Single Period Model

We begin by considering a single period model where a retailer stocks two products, say 1 and 2, which are partial substitutes. For ease of exposition and to focus attention on the demand substitution effects, we first study a simple model where total demand for the two products, \( D \), is known, but demand for each product is uncertain. Demand for products 1 and 2 are \( p^*d \) and \((1-p^*)D\) respectively where \( p \) is a random variable with a continuous distribution function \( F \) that has a finite support \([a, b]\), \( 0 \leq a < b \leq 1 \). For ease of exposition, we assume \( a = 0 \) and \( b = 1 \). A fixed proportion \( \gamma \) of the customers who come in for a product and do not find it in stock will switch to the other product and buy it if it is available, while the proportion \((1-\gamma)\) will walk away. Thus, the products are substitutes in two respects: their demands are negatively correlated and unsatisfied customers for one product may switch to the other. A typical example of such a scenario is an airline that often knows the total number of passengers on a flight ahead of time and offers two types of lunch entrees but does not know the proportion that will choose each type. Our objective is not to just solve this stylized problem but to obtain results and insights that will be useful for more general problems, as will be clear soon.

We first assume that all the cost and substitution parameters are the same for both products and later show how to relax this condition. Note that Rajaram and Tang (2001) and several other recent works such as Gaur and Honhon (2005) and Kok and Fisher (2004) assume symmetric costs. This is appropriate for horizontally differentiated products; various ice cream or yogurt flavors have identical prices. The retailer purchases the items at a unit cost \( c \) and obtains a revenue \( s \) by selling one unit. The retailer incurs a holding cost \( h \) per unit of unsold inventory, which is equal to the starting inventory (the purchased quantity) less the inherent demand for the product and the demand from customers who switch. Further, a penalty cost \( \pi \) is incurred per instance of an unsatisfied customer. Unsatisfied customers include customers who when faced with a stockout of their first preference were not willing to substitute and
those who were willing to switch but could not find the other product. We assume that $c, \pi, h, s > 0$ and $s > c$, as is typical in the literature. Salvage value for excess inventory can be incorporated easily. There may be a hidden cost for the retailer when consumers do not find their first choice product and are forced to substitute and we can refer to this as a substitution cost. We show later how such a cost can be easily incorporated in the model. The retailer has to decide how much of each product to stock to maximize expected profits. We assume zero starting inventories in our initial exposition but relax it thereafter.

Let the expected profits of the retailer be denoted by $\Pi(Q)$ where $Q$ is the vector $(Q_1, Q_2)$ denoting the order quantities of the products 1 and 2. We then have the following concavity result which is useful in establishing that the first order optimality conditions are also sufficient.

**Theorem 2.1:** The expected profit function $\Pi(Q)$ is concave and continuously differentiable.

To characterize the optimal policy, we first write $\Pi(Q)$ explicitly and then arrive at the policy using first order conditions. The explicit formulation of $\Pi(Q)$ is typically quite complex involving various cases and so we adopt the following approach to simplify the formulation. Let the order quantities of the two products be $Q_1 = \alpha D$ and $Q_2 = \beta D$. We let $D = 1$, unless noted otherwise, without loss of generality. Let

$$\Pi(Q) = \begin{cases} 
\Pi_1(\alpha, \beta) & \text{if } \alpha + \beta \geq 1 \\
\Pi_2(\alpha, \beta) & \text{if } \alpha + \beta < 1.
\end{cases}$$

We split the formulation $\Pi(Q)$ into these two cases because when $\alpha + \beta \geq 1$, it can be shown that stockout-generated substitution demand can be fully satisfied from available inventory and so the derivation of $\Pi_1(\alpha, \beta)$ is straightforward. So, we derive $\Pi_2(\alpha, \beta)$ for the case $\alpha + \beta < 1$ and omit the derivation of $\Pi_1(\alpha, \beta)$. To facilitate the analysis, we define $\hat{\alpha} = \frac{\alpha - \gamma (1 - \beta)}{(1 - \gamma)}$ and $\hat{\beta} = \frac{(1 - \beta - \gamma \alpha)}{(1 - \gamma)}$.

Note that $\hat{\alpha} < \alpha$ and $\hat{\beta} > 1 - \beta$. Intuitively $\hat{\alpha}$ is the $p$ value at which excess inventory of product 1 is exactly equal to the substitution demand from product 2. Similarly, $\hat{\beta}$ is the value of $p$ at which excess inventory of product 2 is exactly equal to substitution demand for product 1. Furthermore, note that
\( \hat{\alpha} = \alpha \) and \( \hat{\beta} = \alpha = 1 - \beta \) when \( \alpha + \beta = 1 \). We consider all possible cases of the realization of the random variable \( p \) to identify the exact formulation.

**Case 1:** \( p < \alpha \). Since \( \alpha + \beta < 1 \), this immediately implies that \( 1 - p > \beta \). Hence \( (1 - p - \beta)D \) is the excess demand of product 2 and \( \gamma(1 - p - \beta) \) customers are willing to substitute to product 1. Note that some or all of these substituting customers may be satisfied by the excess inventory \( \alpha - p \) of product 1. Thus we have to analyze further sub cases.

**Case 1A:** \( p \leq \hat{\alpha} \). All the searching customers are satisfied since \( p \leq \hat{\alpha} = \frac{\alpha - \gamma(1 - \beta)}{1 - \gamma} \) implies that \( \alpha - p \geq \gamma(1 - p - \beta) \). In this case, total sales equals \( p + \beta + \gamma(1 - p - \beta) \) from sale of product 1, sale of product 2 and all the customers who substituted product 1 for product 2. The excess inventory (of product 1) after all substitution has occurred is \( (\alpha - p) - \gamma(1 - p - \beta) \). Further, the lost sales (of product 2) due to the customers who were not willing to substitute is equal to \( (1 - \gamma)(1 - p - \beta) \).

**Case 1B:** \( p > \hat{\alpha} \). In this case, \( \alpha - p < \gamma(1 - p - \beta) \), which implies that not all searching customers are satisfied. So, the total sales is equal to \( p + \beta + (\alpha - p) = \alpha + \beta \). Clearly since all inventory is used up by the originating demands and the searching customers, there is no excess inventory at the end of the period. The lost sales after some minor simplification is \( (1 - (\alpha + \beta)) \).

**Case 2:** \( p > \alpha \) and \( (1 - p) > \beta \Rightarrow p < 1 - \beta \). In this case, both products have excess demand and so there is no excess inventory at the end of the period. Total sales equals \( \alpha + \beta \) and lost sales equals \( 1 - (\alpha + \beta) \).

**Case 3:** \( p > \alpha \) and \( (1 - p) < \beta \Rightarrow p > 1 - \beta \). Here, we have excess demand for product 1 and excess inventory of product 2 and some unsatisfied customers of product 1 are willing to substitute product 2. As in Case 1, we have two sub cases:

**Case 3A:** \( 1 - \beta < p < \hat{\beta} \): In this case, not all searching customers are satisfied. Thus, as in case 1B, the total sales is \( \alpha + \beta \), lost sales is \( (1 - (\alpha + \beta)) \) and there is no excess inventory.
Case 3B: $\hat{\alpha} \leq p \leq 1$: In this case, as in (1A), all searching customers are satisfied. Total sales is $\alpha + (1 - p) + \gamma (p - \alpha)$, lost sales is $(1 - \gamma)(p - \alpha)$ and the inventory left over at the end of the period is $\beta - (1 - p) - \gamma (p - \alpha)$.

Observe that the total sales, lost sales and excess inventory at the end of the period are identical in cases (1B), (2) and (3A) and these three cases can be combined into one scenario where $\alpha \leq p \leq \beta$. So,

$$\Pi_2(\alpha, \beta) = \int_0^\alpha \left[ s[p + \beta + \gamma(1 - p - \beta)] - h[(\alpha - p) - \gamma(1 - p - \beta)] - \pi(1 - \gamma)(1 - p - \beta) \right] dF(p) \quad \text{Case 1A}$$

$$+ \int_\alpha^\beta s(\alpha + \beta) - \pi(1 - (\alpha + \beta)) dF(p) \quad \text{Cases 1B, 2, 3A}$$

$$+ \int_\beta^1 \left[ s[(\alpha + (1 - p)) + \gamma(p - \alpha)] - h(\beta - (1 - p) - \gamma(p - \alpha)) - \pi(1 - \gamma)(p - \alpha) \right] dF(p) \quad \text{Case 3B}$$

$$- c(\alpha + \beta);$$

A similar but somewhat simpler derivation, which is omitted, follows for the case $\alpha + \beta \geq 1$.

$$\Pi_1(\alpha, \beta) = \int_0^\beta \left[ s[p + \beta + \gamma(1 - p - \beta)] - h[(\alpha - p) - \gamma(1 - p - \beta)] - \pi(1 - \gamma)(1 - p - \beta) \right] dF(p)$$

$$+ \int_\beta^1 s[\alpha - p + \beta - (1 - p)] dF(p)$$

$$+ \int_1^\alpha \left[ s[(\alpha + (1 - p)) + \gamma(p - \alpha)] - h(\beta - (1 - p) - \gamma(p - \alpha)) - \pi(1 - \gamma)(p - \alpha) \right] dF(p) - c(\alpha + \beta)$$

The three integral terms correspond respectively to the following three scenarios: (1) Inventory exceeds demand for product 1, demand exceeds inventory for product 2 and substitution demand of 2 is fully satisfied from remaining inventory of 1, (2) Inventory exceeds demand for both products, (3) Inventory exceeds demand for product 2, demand exceeds inventory for product 1 and substitution demand of 1 is fully satisfied from remaining inventory of 2. The function $\Pi(Q)$ was shown to be concave in $Q$. It is
easy to show that $\Pi_i(\alpha, \beta)$ is strictly concave in $(\alpha, \beta)$. Thus the global maximum of $\Pi(Q)$ is uniquely attained. The optimal values of $(\alpha, \beta)$ and $(\hat{\alpha}, \hat{\beta})$, denoted by $(\alpha^*, \beta^*)$ and $(\hat{\alpha}^*, \hat{\beta}^*)$, respectively, can be obtained by solving the corresponding first order conditions of $\Pi_i(\alpha, \beta)$. We then have:

**Theorem 2.2:** Let $\gamma^* = 1 - \frac{2(h + c)}{s + \pi + h}$. If $\gamma \leq \gamma^*$, the optimal inventory level is determined using:

$$F(\alpha^*) = \frac{s + \pi - c - \gamma(s + \pi + h)}{s + \pi + h - \gamma(s + \pi + h)} = 1 - \frac{h + c}{(1 - \gamma)(s + \pi + h)}$$

(2.1)

$$F(1 - \beta^*) = \frac{h + c}{(1 - \gamma)(s + \pi + h)}$$

(2.2)

If $\gamma > \gamma^*$, the optimal inventory level is determined using:

$$F(\hat{\alpha}^*) = \frac{1}{1 + \gamma} - \frac{h + c}{(1 + \gamma)(s + \pi + h)} = \frac{s + \pi - c}{(1 + \gamma)(s + \pi + h)}$$

(2.3)

$$F(\hat{\beta}^*) = \frac{\gamma}{1 + \gamma} + \frac{h + c}{(1 + \gamma)(s + \pi + h)} = \frac{\gamma(s + \pi + h) + h + c}{(1 + \gamma)(s + \pi + h)}$$

(2.4)

where $\hat{\alpha}^*$ and $\hat{\beta}^*$ are functions of both $\alpha$ and $\beta$.

Observe that when $\gamma \leq \gamma^*$, we have a “partially decoupled” inventory policy wherein the optimal base-stock level of a product is independent of the inventory level of the other product. Thus, even if product 2’s inventory (after ordering) is very high, it is optimal to raise the inventory of product 1 to its base-stock level even though the products are substitutable. When $\gamma > \gamma^*$ this is not true because there is so much substitution that high inventory levels of one product significantly diminish the attractiveness of holding an incremental unit of inventory of the other product. Observe that the optimal inventory expressions in (2.1) and (2.2) are similar to the newsboy formula, except for the term $(1 - \gamma)$. This implies that, higher the substitution fraction, lower the optimal inventory level. This effect is also true in the case where $\gamma > \gamma^*$, even though this is not transparent in (2.3) and (2.4). This is consistent with prior results as well as the intuitive reasoning that as items become more substitutable, the retailer needs to carry lower inventories. Even the early papers on both consumer-driven and retailer-driven substitution (McGillivray and Silver 1978, Robinson 1990) have pointed out that increased levels of substitution result in lower
inventories. However, prior literature has not identified an optimal policy for substitutable items wherein one could decouple the inventory decisions without ignoring substitution. Expressions (2.1) and (2.2) and the decoupled policy can also be obtained by reformulating the problem as a single-product problem with a “left-over inventory” function that represents the inventory remaining after meeting all substitution demand. Finally, the closed-form nature of the optimal inventory expressions (2.1-2.4) is attractive as it suggests a simple approach to adjust for substitution that can be implemented easily.

Salvage value for excess inventory can be incorporated easily; if \( v \) represents the salvage value, then we replace \( h \) with \( (h-v) \) in the above formulas. Further, in reality, consumers who have to frequently substitute due to stockouts may stop visiting the retailer and so the retailer may incur a cost due to such substitution. We can incorporate such a substitution or switching cost, say \( b \) per unit substitution, easily. In this case, we add \( -b(1-p-\alpha) \) within the first integral term and \( -b(\gamma(p-\alpha)) \) within the third integral term in the equations for \( \Pi_i(\alpha, \beta) \). The optimality equations will change as follows: the denominator in (2.1) will include a term \( \gamma b \) and in (2.3), \( (1+\gamma) \) in the denominator will be replaced by \( (1+b\gamma/(s+\pi+h)) \). The impact is as one would expect – optimal inventories increase with substitution cost \( b \).

We note that the condition \( \gamma \leq \gamma^* \) is not restrictive and higher the newsboy ratio, higher the value of \( \gamma^* \). For instance, suppose the parameters are such that \( s = 100, c = 20, \pi = 5 \) and \( h = -8 \) (implying a significant salvage value). In this case, \( \gamma^* = 0.75 \), where the newsboy ratio (NR) for each product without substitution is 0.87. So, we find that for a wide range of \( \gamma \) values between 0 and 0.75, the decoupled policy is optimal. We discuss the range of \( \gamma \) values typically observed in practice in section 5.

2.1 Asymmetric costs

We now consider the case where the costs and revenues and the substitution parameter for the two products are different. As before, we study a model where the total demand \( D \) for both products is constant but the demand for each product is uncertain. A fixed proportion \( \gamma_i \) of the customers who come

---

\(^1\) We are grateful to an anonymous referee for pointing this out.
in for item $i$ and do not find it in stock will switch to the other product and buy it if it is available, while the proportion $(1-\gamma_i)$ will walk away.

Multi-product substitution inventory problems with asymmetric costs are difficult to analyze because the objective function may not be concave and sometimes not even unimodal (see for instance Netessine and Rudi (2003)). In addition to the assumptions made earlier, one needs to make certain assumptions on the cost parameters to guarantee concavity (see Parlar and Goyal 1984 and Ernst and Kouvelis 1999). We start by assuming that $\pi_i = \pi_j = 0$, i.e., we ignore the penalty due to a lost sale and assume that the loss of revenue subsumes this potential cost. This condition is not necessary for concavity, but is sufficient and improves ease of exposition. The next condition we require is that $s_i > \gamma_i s_j$. This condition ensures that it is worthwhile to stock product $i$ because, if it were not true, the retailer can earn higher revenue by not stocking item $i$ and having all potential consumers for item $i$ substitute to item $j$. Similarly, it can be shown that we need $s_i + h_i > \gamma_i (s_j + h_j)$; however, this condition is subsumed by the condition $s_i > \gamma_i s_j$ if holding costs of higher-priced products are higher than those for lower-priced products, as is typically true in reality. While Parlar and Goyal (1984) also require $s_i > \gamma_i s_j$ in a two-product substitution problem, holding costs are absent in their formulation.

Using an approach similar to the one used in the previous subsection, we have a partial characterization of the optimal values of $(\alpha, \beta)$ in the following theorem, which we state without proof.

**Theorem 2.3:** Let $\gamma_i^* = \frac{s_i - (h_i + 2c_i)}{s_i + h_i}$ and $\gamma_j^* = \frac{s_j - (h_j + 2c_j)}{s_j + h_j}$. If $\gamma \leq \gamma_i^*, i = 1, 2$, then

$$F(\alpha^*) = \frac{s_i - c_i - \gamma_i (s_j + h_j)}{s_i + h_i - \gamma_i (s_j + h_j)}$$

and

$$F(1 - \beta^*) = \frac{h_j + c_j}{s_j + h_j - \gamma_j (s_i + h_i)}.$$  

(2.5)

As before, it is straightforward to check that when $\gamma \leq \gamma_i^*$, we have $\alpha^* + \beta^* \geq 1$ and thus the above expressions indeed give the optimal values. If $\gamma \leq \gamma_i^*$ for $i = 1, 2$, then substitution levels are high and so it may be optimal to carry very little inventory implying low base-stock levels and so the condition $\alpha^* + \beta^* \geq 1$ may not be satisfied.
Thus, when $\gamma \leq \gamma^*_1$, the “partially decoupled” inventory policy is again optimal. The expression for $F(\alpha^*)$ in (2.5) confirms the intuition that if product 2 is the higher-priced product (with, say, higher $s_2$, $c_2$, and $h_2$), then the optimal base stock level of product 1 will be lower and in fact if $s_2$ and $h_2$ are high enough, the optimal base stock level can be below the mean demand and even close to zero. Higher the substitution parameter $\gamma_1$ (substitution from 1 to 2), greater is the impact of the differential in prices. Thus, the risk-pooling effect due to substitution depresses the inventories of both products as in the symmetric case. However, this impact is more significant on the lower-priced product because the retailer finds it more profitable to hold lower inventory of the lower-priced item and “allow” original consumers of this item to substitute to the higher-priced one. The expressions in (2.5) also clarify the conditions imposed earlier to ensure concavity; observe that if the conditions are not satisfied, the denominator in the expressions in (2.5) can become negative. We note that, in the asymmetric case, the condition that $\gamma \leq \gamma^*_1$ is sufficient, but not necessary. The necessary condition can be obtained as in the symmetric case and is given by

\[
(1 - \gamma_1^* \gamma_2^*) + \frac{(1 + \gamma_1)(s_2 - c_2)}{(s_1 + h_1)} + \frac{(1 + \gamma_2)(s_1 - c_1)}{(s_1 + h_1)} \leq (s_1 - c_1 + s_2 - c_2)\left(\frac{1}{s_2 + h_2} + \frac{1}{s_1 + h_1}\right).
\]

### 2.2 Uncertainty in D

We now consider the case where the total demand $D$ is random. To keep the notations and discussion simple, we assume that costs and $\gamma$ are the same for both products. We assume that $D$ has a compact support given by $[L, H]$ and the distribution and probability density function of $D$ are given by $G$ and $g$ respectively. The discussion in this section is as follows. We first formulate the problem when $D$ is random. Then, we derive the optimal inventory levels and show that the partially decoupled property in the known $D$ case still holds when $D$ is random provided $\gamma$ is less than some $\hat{\gamma}$, which depends on the cost parameters and the distributions of $p$ and $D$. We then derive expressions for $\hat{\gamma}$ for some specific distributions of $p$ and $D$ and demonstrate that $\gamma \leq \hat{\gamma}$ represents a reasonable range, as we did earlier.
Let $Q = (Q_1, Q_2)$ be the order quantities and $\mathcal{R}(Q)$ denote the expected single period profit. As before, we re-cast the decision variables as $\alpha$ and $\beta$, where $Q_1 = \alpha D$ and $Q_2 = \beta D$, with the understanding that $D$ is now a random variable. When $D$ was constant, recall that we split the formulation of the profit function into two cases, namely $\alpha + \beta \geq 1$ and $\alpha + \beta < 1$. Correspondingly, we now have three cases:

Case 1: $\mathcal{R}(Q) = \int_L^\mu \Pi_1 \left( \frac{Q_1}{D}, \frac{Q_2}{D} \right) dG(d)$ when $Q_1 + Q_2 \geq H$, 

Case 2: $\mathcal{R}(Q) = \int_{\hat{Q}_1}^\mu \Pi_1 \left( \frac{Q_1}{D}, \frac{Q_2}{D} \right) dG(d) + \int_{\hat{Q}_1}^\mu \Pi_2 \left( \frac{Q_1}{D}, \frac{Q_2}{D} \right) dG(d)$ when $L \leq Q_1 + Q_2 < H$, 

Case 3: $\mathcal{R}(Q) = \int_L^\mu \Pi_2 \left( \frac{Q_1}{D}, \frac{Q_2}{D} \right) dG(d)$ when $Q_1 + Q_2 < L$, 

where $\Pi_i, i = 1, 2$ are exactly the same functions as in section 2. Note that $\mathcal{R}(Q)$ is jointly concave in $Q = (Q_1, Q_2)$ because if we fix $D$ to be a constant (i.e. a realization of $D$), we know from section 2 that the profit function $\Pi(Q)$ is concave.

For the remaining part of this section, we focus on Case 1, i.e. when $Q_1 + Q_2 \geq H$ since our intent is to identify when the decoupled policy is optimal. So, we look at the optimality conditions corresponding to this case and characterize the region where this is indeed the optimal solution. We have:

$$\int_L^\mu F \left( \frac{Q_1^*}{D} \right) dG(D) = 1 - \frac{h + c}{(1 - \gamma)(s + \pi + h)}$$

and

$$\int_L^\mu F \left( 1 - \frac{Q_2^*}{D} \right) = \frac{h + c}{(1 - \gamma)(s + \pi + h)}.$$

We note that since $\int_L^\mu F \left( \frac{Q_1}{D} \right) dG(D)$ is non-decreasing in $Q_1$ and is zero when $Q_1 = 0$, there is some threshold $\hat{\gamma}$ such that the above equations indeed represent the optimal values of $(Q_1, Q_2)$ whenever $\hat{\gamma} \leq \hat{\gamma}$. Clearly, the exact value of $\hat{\gamma}$ depends on the specific distributions that we choose for $p$ and $D$. Next, we derive this threshold value for a few distributions and observe that $\hat{\gamma}$ is indeed high enough.

The calculations are tedious and can be found in the appendix.
Corollary 2.1: (1) Let $p$ be Uniform $[a, b]$ and $D$ be uniform $[L, H]$. We then have:

$$\hat{\gamma} = 1 - \frac{c + h}{(1 - X)(s + \pi + h)}$$

where if $H > 2bL$ then, $X = \frac{0.5H(1 + \ln(2b)) - aH - L(b - a)}{(H - L)(b - a)}$ and

if $H \leq 2bL$ then $X = \frac{0.5H(\ln(H / L)) - a(H - L)}{(H - L)(b - a)}$.

(2) When $p$ follows the theta distribution i.e. $F(x) = x^\theta, \theta \geq 2$ and $D$ is uniform $[L, H]$ with $H > 4L$,

$$\hat{\gamma} = 1 - \frac{c + h}{(1 - Y)(s + \pi + h)}, \text{ where } Y = \frac{(0.5H)^\theta (H^{\theta-1} - L^{\theta-1})}{(H - L)(\theta - 1)(HL)^{\theta-1}}.$$

To get an idea of what the values of $\hat{\gamma}$ are in these scenarios, we considered various possible values of the parameters. We found that in all cases, as one would expect, $\hat{\gamma}$ decreases with increased variability in total demand $D$ and $\hat{\gamma}$ increases when the newsboy ratios increase. For instance, when $D$ varies between $[500, 1000]$ and $p$ is uniformly distributed between $(0, 1)$ for newsboy values of 0.91, the value of $\hat{\gamma} = 0.72$. Note that the uniform distribution represents much higher variability than the family of theta distributions. In the above scenario, the value of $\hat{\gamma}$ is higher and equals 0.83 when theta equals 2.

We also note that as theta increases, the value of $\hat{\gamma}$ also increases. For instance when theta equals 5, $\hat{\gamma} = 0.88$. Overall, our analysis indicates that for reasonable values of newsboy ratios that one may observe in practice, the decoupled policy is optimal even if $D$ is random but within some reasonable ranges and we can compute the optimal target levels easily.

3. The Multi-period Problem.

In this section, we first present the finite horizon lost sales version of the two-product substitution problem followed by the infinite horizon case. Our assumptions on demand distributions, costs, etc. remain the same as in the single period problem. We focus on the symmetric cost case since the conditions under which decoupled policies are optimal are simpler and weaker than those in the asymmetric case and also for ease of exposition. However, we do point out at the end of the section how the results can be generalized to the asymmetric case. For similar reasons, we assume that total demand $D$
is constant over time. Finally, in all our analysis, we assume that unfulfilled demand (after all possible substitution takes place) at the end of a period is lost and is not backlogged and we also assume a zero lead time for replenishment.

**Finite horizon case**

Next, we formulate the multi-period model and show that it is concave. We then demonstrate that under conditions similar to those identified earlier, it is optimal to follow a base stock policy for each product that is independent of the inventory level of the other product. Thus, the partially decoupled structure obtained in the single period problem holds in the dynamic problem. Moreover, we show that the optimal order up to level is monotonic non-decreasing in the number of periods until the end of the horizon.

Without loss of generality, we assume that the discount factor \( \delta = 1 \). Let \( G_n(I^1_n, I^2_n) \) represent the expected profit when there are \( n \) periods until the end of the horizon, the starting inventory before ordering of the two products are \( I^1_n \) and \( I^2_n \) and an optimal policy is used in all the remaining periods.

The corresponding dynamic program\(^2\) is as follows:

\[
G_1(I^1_1, I^2_1) = \max_{(\alpha_1, \beta_1) \in (I^1_1, I^2_1)} \left\{ \Pi_1(\alpha_1, \beta_1) - c(\alpha_1 + \beta_1 - I^1_1 - I^2_1) \right\}
\]

\[
G_n(I^1_n, I^2_n) = \max_{(\alpha_n, \beta_n) \in (I^1_n, I^2_n)} \left\{ \Pi_n(\alpha_n, \beta_n) - c(\alpha_n + \beta_n - I^1_n - I^2_n) + \int_{1-\beta_n}^{1} G_{n-1}(\alpha_n - p\beta_n - (1-p)\beta_n) dF + \int_{\alpha_n}^{\beta_n} G_{n-1}(0, \beta_n - (1-p)\beta_n) dF \right\}
\]

Let \( G_n(I^1_n, I^2_n) = \max_{(\alpha_n, \beta_n) \in (I^1_n, I^2_n)} \left\{ (X_n(\alpha_n, \beta_n) + c(I^1_n + I^2_n)) \right\} \), where \( X_n(\alpha_n, \beta_n) \) is the expected profit with \( n \) periods until the end of the horizon and the starting inventories of the two products are \( \alpha_n \) and \( \beta_n \) and an optimal policy is used in all the remaining periods. Our first result which we show using ideas used in the proof of Theorem 2.1 and induction, is as follows.

**Theorem 3.1:** \( X_n \) is jointly concave in \( (\alpha_n, \beta_n) \).
The concavity of \( X_n \) immediately implies that if \((\alpha_n^*, \beta_n^*)\) is the maximizer of \( X_n(\alpha, \beta) \),

\[
G_n(I_n^1, I_n^2) = c(I_n^1 + I_n^2) + X_n(\alpha_n^*, \beta_n^*); \quad I_n^1 \leq \alpha_n^*, I_n^2 \leq \beta_n^*,
\]

that is, when the initial inventory of both products are below the optimal target levels, it is optimal to raise the inventory of both products to the optimal level \((\alpha_n^*, \beta_n^*)\). Note that when there is more than one decision variable, in general, concavity does not guarantee the optimality of a simple base stock policy, wherein the order-up-to level of a product is independent of the starting inventory level of the other product. Thus, for instance, even if \( I_n^1 \leq \alpha_n^* \), it may not be optimal to raise the inventory of product 1 to \( \alpha_n^* \) if \( I_n^2 > \beta_n^* \).

However, our next result shows that when \( \gamma \leq \gamma^* \), the optimal policy in the multi-period setting also exhibits the partially decoupled structure. The proof of this result uses two key observations. The first observation shows that in a multi-period setting, when \( \gamma \leq \gamma^* \) it is sufficient to restrict our attention to the case when in every period the optimal total inventory level after ordering is at least equal to \( D \). In the single-period model, as we saw, this is the case. To show that this is true requires the use of joint concavity and the calculation of the value of a marginal unit of inventory. The second observation essentially works out the technical details of how the dynamic program preserves separability, i.e. the decoupled structure of the optimal policy that we observed in the single period problem. We also show in the theorem below that the optimal base stock levels are monotonic non-decreasing in \( n \), a property that is known to be true in certain single product dynamic inventory control problems without substitution.

**Theorem 3.2** When \( \gamma \leq \gamma^* \) we have

\[
(1) \quad G_n(I_n^1, I_n^2) = c(I_n^1 + I_n^2) + X_n(\alpha_n^*, \beta_n^*); \quad I_n^1 \leq \alpha_n^*, I_n^2 \leq \beta_n^* \\
2\quad I_n^1 \leq \alpha_n^*, I_n^2 > \beta_n^* \\
3\quad I_n^1 \leq \alpha_n^*, I_n^2 \leq \beta_n^* \\
4\quad I_n^1 > \alpha_n^*, I_n^2 \leq \beta_n^* \\
5\quad I_n^1 > \alpha_n^*, I_n^2 > \beta_n^*
\]

Note that \( \Pi_1(\alpha, \beta) \) in the above expression is the single period expected profit excluding the purchasing cost and
(2) \((\alpha_{n+1}^*, \beta_{n+1}^*) \leq (\alpha_n^*, \beta_n^*)\)

The above result is quite surprising as when products are substitutable, one would expect to find some type of a state dependent base stock policy to be optimal, where the state contains information about the initial inventory of the products. We note that since we have not made assumptions about the end of the horizon whereby a myopic policy is not optimal (as can be seen from (2) of the above theorem), one cannot simply conclude the property in a dynamic setting from its validity in a single period setting. The intuition for the above result is as follows. In period \(n\), the cost structure is such that it is optimal to have a total inventory level of at least \(D\), after ordering and before demand realization. This immediately implies that every customer who is willing to substitute is satisfied from the inventory. Thus, any incremental inventory more than, say, \(\alpha_n^*\) can be used to satisfy only the original demand of product 1. This partially drives the structure of the policy. This intuition can be seen somewhat clearly in the single period case. In the multi-period problem, the intuition carries through, with several technicalities needed that can be seen in the proof. The optimality of such a decoupled policy has not been shown in single- or multi-period models in the consumer-driven or retailer-driven inventory substitution literature.

To extend these results to the asymmetric cost case, observe that a key result we used in the proof is that \(\alpha_n^* + \beta_n^* \geq 1\) \(\forall n\) if \(\gamma \leq \gamma^*\). Equivalently, in the asymmetric cost case, we have shown that \(\alpha^* + \beta^* \geq 1\) in the single-period model if \(\gamma \leq \gamma^*\). In addition, assuming the same conditions as in the single-period model to ensure concavity, we can use an approach similar to that used in the symmetric case to obtain an optimal solution to the finite horizon problem in the asymmetric case.

**Infinite horizon case**

We continue to restrict our attention to the case when the values of the parameters are such that it is optimal to order \(\alpha + \beta \geq 1\) in every period and focus again on the symmetric case for ease of exposition.

---

is different than the one used in section 2.
Denote by \( G(I_1, I_2) \) the total discounted expected infinite profit obtained by starting with an initial inventory \( I_i \) of product \( i \) and using an optimal policy. Then, \( G \) satisfies a functional equation of the form:

\[
G(I_1, I_2) = \max_{(\alpha, \beta) \in I_1, I_2} \left[ \Pi_i(\alpha, \beta) - c(\alpha + \beta - I_i - I_j) + \delta \int_0^{1-\beta} G(\alpha - p - \gamma(1 - p - \beta), 0) dF 
\right. \\
+ \delta \int_{1-\beta}^{\alpha} G(\alpha - p, \beta - (1 - p)) dF + \delta \int_{\alpha}^{1} G(0, \beta - (1 - p) - \gamma(p - \alpha)) dF \right] 
\]  

(3.1)

We now characterize the solution of the infinite horizon problem. To do so, we rewrite (3.1) as

\[
G(I_1, I_2) = \max_{(\alpha, \beta) \in I_1, I_2} \left[ V(\alpha, \beta) + c(I_1 + I_2) \right]^3 
\]

In the proof of the theorem that follows, we demonstrate that \( G \) is a bounded continuous function. This is established by standard convergence techniques using fixed point operators. The proof is illustrative in that several desirable properties of the finite horizon problem hold for infinite horizon problem. In particular, concavity and the decoupled policy structure continue to hold in the infinite horizon case. This allows us to analyze \( V(\alpha, \beta) \) which leads to the optimal policy of the infinite horizon problem. Let \( \gamma^\circ = 1 - \frac{2(h + (1 - \delta)c)}{s + \pi + h - \delta c} \) and \( \alpha^*_\gamma \) and \( \beta^*_\gamma \) be such that

\[
F(\alpha^*_\gamma) = \frac{(s + \pi - c) - \gamma(s + \pi + h - \delta c)}{(s + \pi + h - \delta c)(1 - \gamma)} \quad \text{and} \quad F(1 - \beta^*_\gamma) = \frac{h - c(1 + \delta)}{(s + \pi + h - \delta c)(1 - \gamma)} 
\]

We then have the following result:

**Theorem 3.3:** Let \( \gamma \leq \gamma^\circ \). Then the optimal policy for the infinite horizon problem is to follow a stationary base-stock policy with order-up-to levels \( \alpha^*_\gamma \) and \( \beta^*_\gamma \).

We find that our earlier result for the single period problem extends to the infinite horizon problem. That is, for certain \( \gamma \) values, we have a partially decoupled inventory policy similar to the one in the single-period model. In fact, the range of \( \gamma \) values is larger in this case (since \( \gamma^\circ \leq \gamma^\circ \)) and can even extend to a value close to 1. For instance, if \( c = 0.2s \), \( \pi = 0.15s \) and \( h = 0.12c \) and \( \delta = 0.95 \), then \( \gamma^\circ = 3 \)

---

\[1\] We note that \( V(\alpha, \beta) \) is the infinite horizon discounted version of the function \( X \) used in earlier discussions on the finite horizon multi-period model.
The decoupling effect is sharpened in an infinite horizon setting, where the optimal inventory level is typically higher than in a single period setting for the same set of cost parameters.

Based on our results so far, one can perhaps make a more general conjecture. We find that if the problem parameters are such that the inventory level of an item is sufficiently high, then all substitution demand is fully satisfied and so the marginal benefit of additional inventory does not come from meeting substitution demand and in turn the optimal inventory level of the item is not impacted by the inventory level of a substitute. One may then expect a similar phenomenon to occur even in more general inventory models with multiple products if the substitution fractions are not too high and total demand $D$ is not highly variable. In that case, one can compute inventory levels of an item using simple newsvendor type equations stated before Theorem 3.3 and in (2.1), and these may be near optimal in the original problem. This conjecture, if validated, can be valuable since substitution problems are in general very complex to analyze.

4. Computational study

In earlier sections, we have pointed out various scenarios where an optimal policy involves a simple modification to the newsboy policy to adjust for substitution. These scenarios might be seen as restrictive since we have conditions on the value of the substitution parameter and on the demands of the products. Further, our analysis was restricted to the two-product case. To explore whether the results obtained are reasonably robust under more general conditions, we conducted a computational study. In particular, we wanted to understand whether the approach performs well when the problem parameters take on values typically found in one or more retail sectors. So, we performed two sets of numerical studies, one with two products and one with six products.

Problem parameters
To explore realistic problem parameters for the study, we researched both the academic and trade literature as will be clear shortly. The manner in which data was generated and the parameters used in the study are described next, focusing first on the two-product case.

**Demand:** In the two-product experiments, the demands for the two products were generated as follows. First, the demand for each product was picked from a uniform distribution so as to test the performance under high levels of variation in demand. The demand values for each product were picked from the range [0,100] or a subset of it such as [10, 30], [20, 80], etc. The range of demand for each product was picked so that the ratio of the mean demand for the two products varied from 1:1 to 1:4 and three different ratios were considered. So, if demand for product 1 is picked from U[10, 70] and product 2 is U[20, 100], then the ratio of mean demand is 40:60 or 1:1.5. Also, two different demand variance cases were considered: high and low. For example, if mean demand for both items is 50, then in the “high variance” case, demand varied between 0 and 100 and in the “low variance” case, demand varied between 30 and 70. Further, the demands for the two products were generated so that their correlation varied over a wide range. Eight different correlation values were considered: +0.53, +0.36, +0.21, 0, -0.14, -0.27, -0.41, -0.56. Specifically, we ran the simulations for 500 periods and the demand for the two products across these 500 periods had the above correlation values.

To identify if we were considering realistic data points for the demand values in the numerical study, we accessed sales and gross margin data from the Dominick’s grocery chain data base, extensively used in the marketing literature (http://www.gsb.uchicago.edu/kilts/research/db/dominicks/). We found that for almost all the product categories, excluding items that had sporadic demand (i.e., many periods with zero demand), raw demand correlation values between items in a product category varied between 0.5 and -0.4. If we controlled for seasonality effects in total store demand (due to holidays, store promotions, etc.), these correlation numbers are lower. Further, the demand ratio between items in a product category for items such as canned tuna, bath tissue, paper towels, etc. at most stores was well

---

4 In addition, we talked to several managers in the grocery industry that participate in a semester-long management training program called the Food Industry Management Program at University of Southern California.
within the range described above. For instance, the average weekly demand at one of the stores for two
brands of canned tuna was 19 and 27 and correlation in their demands was 0.05, bath tissue was 21 and 59
with a correlation of 0.33 and for paper towels, they were 28 and 36 with a correlation of -0.29. Note that
these correlation numbers are raw numbers and are not controlled for seasonal variations in store sales.
Thus, we ensured that the demand variance, correlation and demand ratio values considered in the study
subsumed the values observed in the Dominick’s database.

Recurrent and cost parameters: Data from industry and academic studies (Andersen Consulting 1996,
Gruen, et al. 2002) indicate that out-of-stock percentages range between 1-10% for various product
categories in grocery stores. Hence, the revenue and cost parameters were generated so as to achieve
newsboy ratios for the items in the range of 0.83 to 0.98. The rationale for this parameter choice is as
follows. Assuming probability of no stockout as the service level measure, which is appropriate for
retailers, this range of newsboy ratios directly translates into a range of service level measures consistent
with those found in the studies referenced above. In particular, we used the following values: 0.83, 0.88,
0.93, and 0.98. Further, these parameter choices included the type of revenue and cost data, appropriately
scaled, observed in the Dominick’s data.

2002, Sloot et al. 2005) have looked at consumer responses to stockouts, primarily in grocery or discount
stores, and they find that a significant proportion of the consumers substitute a similar item within the
store. The substitution percentages within a store vary depending on the product characteristics and the
study, with typical values between 30% and 60%. Further, the studies suggest that customers coming in
for products with higher brand equity are less likely to substitute. Also, substitution rates are higher for
perishable products and “necessities” such as bread. Hence, we considered the following three $\gamma$ values in
the study: 0.3, 0.5 and 0.6. There may very well be products and product categories wherein $\gamma$ values are
higher than 0.6 (or smaller than 0.3) but the higher $\gamma$ values are more applicable to commodity-type
products with little brand equity. We explored both identical and different values of $\gamma$ for the two products.

Heuristic approach: The heuristic approach tested was to use the formula in (2.1) and (2.2) for the two-product problems and (4.1) for the six-product problems as the order quantity for the single-period problem. Note that these order quantities, which are optimal in the stylized settings studied earlier, need not be optimal for the problems tested in the study. We compared the heuristic solution to the average profit for the optimal solution. We are able to obtain the optimal solution because we discretized the state space and computed the profit values for every possible set of inventory values. This was feasible because the demand values were generated so that they were in the range 0 to 100 in the two-product case and so we have to evaluate at most 10,000 possible combinations of inventory values since we consider only integer values. In the 6-product problems, we restricted the study to products with identical parameters so that we could restrict the search to optimal solutions with identical inventories for all the products, which substantially reduces the state space. We also compared the heuristic with a simple base stock policy that ignores substitution, i.e., we simply compute the target base stock level for each item and raise the inventory in each period to this level. Such a base stock policy may be representative of current best practice in the retail grocery sector.

For each set of demand, revenue, cost and substitution parameters, the average profit over 500 demand realizations were computed for the heuristic, optimal solution and the simple base stock policy.

Six product experiments: The procedure for generating parameters was largely the same as in the two-product case except for the following differences. The average demand values for each product was taken to be 25 and two cases were considered: high variance where demand is generated uniformly from the range [0, 50] and low variance where the range is [15, 35]. The demands for the six products were generated so that demand for product $i$ was correlated with sum of the demand for products 1 through ($i - 1$). Five different correlation values were used: 0.55, 0.27, 0.05, -0.17, -0.48. We also considered three

---

5 The grocery managers we talked to confirmed this and indicated that they aim for around 95% in-stock service levels for the items stocked but the actual service levels varied from 90-98% for most items.
possible scenarios for substitution behavior: customers who do not find their first-choice item in stock will make one, two, or three substitution attempts to satisfy their demand. The revenue parameters and substitution parameter ($\gamma$) values considered were the same as in the two product experiments except that we considered identical values for all the products. Since the parameters are identical for all the products, we restricted the search for the optimum to solutions where inventory is identical for all the six products.

**Performance of the solution approach**

The main results of the two-product experiments are presented in Table 1, which presents the percentage gap relative to the optimal solution for both the heuristic and the base stock policy ignoring substitution. The results for the six-product experiments are in Table 3. Based on the results, the heuristic seems to perform very well for a wide range of parameter values. The average optimality gap of the heuristic is 0.22% in the two-product problems and 0.25% in the six-product problems. We have the following observations on the performance of the heuristic as a function of the problem parameters.

- The performance of the heuristic is better at higher values of the newsboy ratio. If the newsboy ratios of both products are 0.88 or higher, the average optimality gap is 0.14% (0.19%) and below 0.8% (1.39%) in the 2-product (6-product) problems.

- In the 2-product problems, if the newsboy ratios are low, the performance of the heuristic is better at low substitution parameters. At lower newsboy ratios and high $\gamma$ values, the heuristic tends to hold too little inventory and this hurts its performance when demand correlation is highly positive. If the newsboy ratio is high, the performance is slightly worse when $\gamma$ is lower and demand correlation is highly negative. This is clear from Table 2, which provides detailed results for various demand correlation values for two different newsboy ratios and $\gamma$ values.

- In the 6-product problems, the impact of substitution parameters is moderated by the number of substitution attempts and the correlation values. So, the worst performance of the heuristic is when the substitution parameter is high ($\gamma = 0.6$) and maximum number of substitution attempts is 1. When the maximum number of substitution attempts is 3, the heuristic performs very well and the worst optimality gap is 1.18% for any problem instance. This is because when consumers are willing to make more attempts to search and substitute, carrying lower inventory is more desirable.

- In 2-product problems where newsboy ratios are different for the two products, it becomes attractive to hold less inventory of the product with the lower newsboy ratio and deliberately stock out of that item to move customers to the higher priced product. But this is not necessarily a reflection of the
poor performance of the heuristic but is instead an indication of the relative unattractiveness of meeting a product’s demand from that product’s inventory relative to stocking out of the product and meeting that demand from a substitute. But we are not trying to model this type of scenario.

The detailed results (not reported due to space limitations) indicate that the optimality gap is larger when the demand variance is higher. This is to be expected as higher variance implies that the heuristic inventory may be less than optimal especially if demand is positively correlated. However, we do not find the optimality gap to be sensitive to the ratio of mean demand of the two products. We solved several problems with different $\gamma$ values for the two products but did not find the results to be very sensitive to differences in $\gamma$ values. Instead, the results are primarily sensitive to the level of both $\gamma_i$ values.

The heuristic policy carries lower inventory than the optimum in most scenarios but sometimes does carry higher inventory. On the hand, the simple base stock policy that ignored substitution always carries higher inventory than the optimum. In the 6-product problems, the inventory determined by the heuristic is on average 3% above the optimum inventory while the simple base stock policy carries on average 11% higher inventory. The heuristic inventory varies between 0.9 to 1.14 times the optimum inventory while the simple base stock policy inventory varies between 1 to 1.4 times the optimum inventory. (We have not provided detailed results on inventories due to space limitations.) When the substitution fraction is high but the maximum number of substitution attempts is low, the heuristic typically carries insufficient inventory relative to the optimum. However, if the newsboy ratio is high, substitution fraction is low (0.3) and the number of substitution attempts is high, the heuristic policy may carry higher inventory than the optimum.

Comparing the performance of the heuristic with the base stock policy ignoring substitution, we find that the profit of the heuristic policy is on average 1.5% (0.7%) better than the policy that ignores substitution in the 2-product (6-product) problems. Note that these are improvements in gross profits and may translate into a significant increase in net profits.\(^6\) Small percentage savings in profits are often used to justify the implementation of technology that is quite expensive. For example, implementing RFID
technology is expected to improve inventory costs by 5% and also see a reduction in out-of-stock items resulting in a recurring annual benefit of $700,000 per $1 billion in annual sales for retailers (A.T. Kearney 2006). Unlike these expensive systems, our heuristic is a simple modification of the newsboy ratio and requires a minor adjustment to existing ordering systems and is easy to implement. Moreover, the heuristic is far more consistent in its performance than the base stock policy. For instance, in the 6-product problems, the heuristic outperforms the simple base stock policy in over 97% of the problem instances. Further, the maximum optimality gaps are significantly larger for the base stock policy as compared to the simple heuristic. However, in the 2-product problems, there is little difference between the two approaches when newsboy ratios are very high (0.98). Indeed, the optimality gap is very small for either policy and in such scenarios, the retailer can essentially ignore stockout-induced substitution effects in making inventory decisions.

Thus, we find that the heuristic performs well under realistic conditions. For most branded products, retailers and manufacturers aim for high service levels, generally around 90-95%. In these cases, the optimality gap is likely to be small (less than 1%) as is clear from Tables 1 and 3. Thus, our approach is likely to be ideal for items that have high brand equity and are non-perishable that have lower substitution fractions and higher newsboy ratios or when the consumer who is willing to substitute is likely to make two or three substitution attempts. Overall, the heuristic approach proposed here appears to be quite robust and is likely to be valuable in most retail environments selling branded products that are likely to have high service levels (over 88%) and moderate substitution proportions (0.3-0.6) if stocked out. If service levels are very high (98% or more), then our results suggest that the retailer can essentially ignore substitution in making inventory decisions. Clearly, our numerical study is not comprehensive and so the results have to be interpreted and used with some caution.

An important input to using the model and heuristic outlined here is the demand for each item and our approach implicitly assumes that the original demand for a product is known. In reality, we observe only sales and this comprises of both original and substitution demand. However, Anupindi, Dada and Gupta

---

6 For instance, Kroger, the largest US grocery chain had gross profits of $15 billion and net profits of $1 billion in fiscal 2005.
(1998) and Kok and Fisher (2004) have proposed and tested empirical methods for estimating original demand and substitution fractions from sales data when there is stockout-based substitution. These methods can be used together with our heuristic to determine inventory levels.

5. Conclusions and future research

We have derived the optimal inventory policy for certain reasonable problem parameters for a single decision-maker in a model with two substitutable products whose demands are negatively correlated and that are partial substitutes in stockout situations in both single-period and multi-period scenarios. We show that the optimal inventory level for a product is partially decoupled from the inventory level of the other product in such scenarios. The two-product model and analysis may appear stylized but is valuable because warehouse stores such as Smart and Final and Costco and convenience stores such as Seven-Eleven typically carry two brands or two substitutable items in many product categories. Even grocery stores carry only two potentially substitutable items in many categories (e.g., batteries or coffee filters in certain sizes). In any case, the numerical study shows that the approach and results developed in the paper are robust and perform well under a variety of conditions not explicitly modeled, and in particular under scenarios observed in reality. The approach developed in the paper is most valuable to a retailer making inventory decisions of items within a category having the following characteristics: substitution across the items is moderate, demand of the category is not highly variable and desired service levels are high (over 85%) but not too high (>98%). Numerous branded product categories and items in the retail sector have these characteristics.

There are natural extensions of this work. First, we would like to extend the analytical results to the N-product case. Second, we could explore experimentally how the decoupled policies perform in multi-product, multi-period problems with replenishment lead times, random demand D and general substitution behavior. Finally, we would like to investigate the contractual aspects of a single decision-maker managing the inventories of the two products versus two independent players managing them. The closed-form nature of inventory policies we have derived may facilitate this analysis.
Acknowledgment: We are grateful to Bill Lovejoy for his valuable suggestions and to the AE and referees for their comments that have improved the paper significantly.

References


Table 1: Results of the computational study: 2 products

<table>
<thead>
<tr>
<th>Newsboy ratios of items 1 and 2</th>
<th>Substitution fraction for the two items ($\gamma_1 = \gamma_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>Heuristic</td>
<td>Base stock</td>
</tr>
<tr>
<td>0.83, 0.83</td>
<td>0.09, 0.33</td>
</tr>
<tr>
<td>0.83, 0.88</td>
<td>0.13, 0.37</td>
</tr>
<tr>
<td>0.83, 0.93</td>
<td>0.12, 0.52</td>
</tr>
<tr>
<td>0.83, 0.98</td>
<td>0.11, 0.49</td>
</tr>
<tr>
<td>0.88, 0.88</td>
<td>0.23, 0.61</td>
</tr>
<tr>
<td>0.88, 0.93</td>
<td>0.23, 0.5</td>
</tr>
<tr>
<td>0.88, 0.98</td>
<td>0.25, 0.56</td>
</tr>
<tr>
<td>0.93, 0.93</td>
<td>0.22, 0.45</td>
</tr>
<tr>
<td>0.93, 0.98</td>
<td>0.21, 0.42</td>
</tr>
<tr>
<td>0.98, 0.98</td>
<td>0.05, 0.11</td>
</tr>
</tbody>
</table>

Note: The numbers (x, y) in each cell represent, respectively, the average and maximum percentage optimality gap over the 48 problems (3 demand ratios, 8 demand correlation values and 2 demand variance values) for each combination of newsboy ratios and substitution fraction. The “Base stock” policy ignores substitution, unlike the heuristic approach.
Table 2: Optimality gap of heuristic for different demand correlation values (2 products)

<table>
<thead>
<tr>
<th>Newsboy ratio ((\gamma_1 = \gamma_2))</th>
<th>Substitution fraction</th>
<th>Demand correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>0.93, 0.93</td>
<td>0.6</td>
<td>0.11</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>0.88, 0.83</td>
<td>0.6</td>
<td>1.91</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note: In these problems, the demand variance is high and the ratio of mean demands is 1:1.

Table 3: Results of the computational study: 6 products

<table>
<thead>
<tr>
<th>Substitution Fraction</th>
<th>Max # of substitution Attempts</th>
<th>Policy</th>
<th>Newsboy Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>Heuristic</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>0.94</td>
</tr>
<tr>
<td>0.3</td>
<td>2</td>
<td>Heuristic</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.10</td>
</tr>
<tr>
<td>0.3</td>
<td>3</td>
<td>Heuristic</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.16</td>
</tr>
<tr>
<td>0.45</td>
<td>1</td>
<td>Heuristic</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.19</td>
</tr>
<tr>
<td>0.45</td>
<td>2</td>
<td>Heuristic</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.53</td>
</tr>
<tr>
<td>0.45</td>
<td>3</td>
<td>Heuristic</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.69</td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>Heuristic</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>1.48</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>Heuristic</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>2.07</td>
</tr>
<tr>
<td>0.6</td>
<td>3</td>
<td>Heuristic</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Base stock</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Note: The Mean and Max values represent, respectively, the average and maximum percentage optimality gap over 10 problems (5 demand correlation values and two demand variance values) for each combination of newsboy ratio, substitution fraction and maximum number of substitution attempts.