Int. J. Production Economics 145 (2013) 78–87

Contents lists available at SciVerse ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Retail inventory management with stock-out based dynamic demand substitution

Baris Tan*, Selcuk Karabati

College of Administrative Sciences and Economics, Koç University, Rumeli Feneri Yolu, Sariyer, 34450 Istanbul, Turkey

ARTICLE INFO

Article history:
Received 23 February 2012
Accepted 5 October 2012
Available online 17 October 2012

Keywords:
Inventory control
Substitution

ABSTRACT

We consider an inventory management problem of a product category in a retail setting with Poisson arrival processes, stock-out based dynamic demand substitution, and lost sales. The retailer uses a fixed-review period, order-up-to level system to control the inventory levels. We present a computational method to determine the order-up-to levels that maximizes the expected profit with profit margins, inventory holding and substitution costs subject to service-level constraints. Determining expected sales, average inventory levels, and number of substitutions between all products for given demand rates, substitution probabilities, and order-up-to levels is not tractable when there are more than two products. Therefore we present efficient and accurate approximations to approximately compute the same performance measures. The approximate approaches are then used to solve the optimization problem by using a genetic algorithm. In a computational study, we discuss the impact of profit margins, inventory holding and substitution costs, and service level constraints on the order-up-to levels and the expected profits. We show that a retailer can increase its expected profits by incorporating substitution among different products.

1. Introduction

This paper studies an inventory management problem in a retail setting with stock-out based substitutions and multiple items in a product category and proposes an approximate solution to determine the order-up-to levels to maximize the expected profit subject to service level constraints. The method uses demand parameters including the substitution probabilities estimated from the point-of-sales data. As a result, the method provides a practical tool for retailers to manage their inventory.

The literature on inventory management under stock-out based substitutions studies the supplier-(or manufacturer-)controlled and customer-driven substitution schemes. In the supplier-controlled substitution scheme, in a stock-out instance, the supplier decides whether to fulfill the demand of the customer with another product. The inventory management (and/or production planning) problem is usually studied in a "one-way substitution" setting, where a higher-graded product can be substituted for a lower-graded product. The primary objective is to minimize the sum of production, inventory holding, and, in some cases, product conversions costs. A detailed discussion of the relevant literature on supplier-controlled substitution is presented in Hsu et al. (2005) and Rao et al. (2004).

In this paper, inventory management under the customer-driven substitution scheme is studied. In the customer-driven substitution scheme, when the first-choice product of the customer is not available on the shelf, the customer may purchase, with a certain probability, another product in the same category in lieu of her first-choice product. Although the retailer can only indirectly affect customers' decisions through his inventory management decisions, ignoring product substitutions in managing the inventories may result in sub-optimal performance: Mahajan and van Ryzin (2001) analyze a single-period, stochastic inventory problem with substitutable products, and show that "substitution effects can have a significant impact on an assortment's gross profits." Ernst and Kamrad (2006) study a two-product problem with customer-driven substitution in a newsvendor setting, and conclude that "using a Newsboy Model framework without regard to substitutions can be sub-optimal."

In the single-period models, it is usually assumed that the demand realizes at the end of the period. A single-product problem can be analyzed under this assumption; however in a multi-product/customer-driven substitution setting, the dynamics of the problem is quite different because of the customer arrival process. Smith and Agrawal (2000) and Mahajan and van Ryzin (2001) present models, with underage and overage costs, that account for customers' arrival order in finding the optimal order quantities. Hopp and Xu (2008) approximate the dynamic substitution behavior with a fluid network model, and study inventory, price, and assortment decisions in centralized and decentralized settings.

The inventory management problem with static substitutions has been extensively studied in the literature. McGillivray and Silver (1978) study a periodic review system with substitutable items having the same unit variable cost and shortage penalty,

* Corresponding author. Tel.: +902123381459; fax: +902123381393.
E-mail addresses: btan@ku.edu.tr (B. Tan), skarabati@ku.edu.tr (S. Karabati).

0925-5273/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2012.10.002

and develop an upper bound on the inventory and shortage costs savings that could be achieved when the product substitution is taken into account in choosing the order-up-to-levels. Parlar (1985) generalizes the newsvendor problem with a product that perishes in two periods, and assumes that the one-period-old and fresh products are substitutable. Parlar (1985) presents an infinite horizon Markov decision model to find the optimal ordering policy. Avsar and Baykal-Gursoy (2002) analyze the competition of two retailers that offer substitutable products, and present a two-person stochastic game to characterize the Nash equilibrium. Rajaram and Tang (2001) study a multi-product newsvendor problem with substitutability, and analyze the impact of demand uncertainty on order quantities and expected profits. Netessine and Rudi (2003) study a single-period problem where unsatisfied demand for a product flows to other products in deterministic proportions, and present analytically tractable solutions for comparing the profits of the centralized and competitive inventory management settings. Nagarajan and Rajagopalan (2008) study a two-product problem with negatively correlated demands. The substitution proportions from the first to the second and from the second to the first product are assumed to be identical. Nagarajan and Rajagopalan (2008) first show that, in a single-period setting and when the substitution proportion is not very large, the optimal base-stock levels are not state-dependent. In a computational study, they also show that a heuristic based on the solution of the two-product problem performs well with multiple products and under general conditions.

A closely related research stream studies customer-driven substitution in the context of assortment planning. Kök and Fisher (2007) study an assortment planning model with substitutable products, develop a procedure for estimating substitution parameters, and present a heuristic for solving the assortment planning problem. Yücel et al. (2009) study assortment and inventory planning problems under customer-driven substitution in retail operations. They show that ignoring substitutability of products or shelf space limitations may result in sub-optimal assortments. Detailed reviews of the literature on assortment planning have been presented by Kök et al. (2008) and Mahajan and van Ryzin (1998).

Following up on our earlier work on the estimation of substitution probabilities (Karabatı et al., 2009), the objective of this paper is to develop an easily implementable method to determine the optimal order-up-to levels that maximize the expected profit of the system. The inventory management method we propose incorporates the effects of stock-out based dynamic substitutions. The method we developed attempts to answer the question how much the total profit of a product category can be increased by setting the order-up-to levels in a way that captures the effects of substitution and profitability of the products. For example, an inventory plan may force customers of products with low profit margins to substitute with higher profit margin products by setting the order-up-to levels intentionally low to cause stock outs. Our first attempt to analyze this problem is given in the thesis of Helvacıoğlu (2009) where different approximation methods are used together with mathematical optimization techniques.

Our paper contributes to the literature discussed above by considering a multi-period multi-product problem that incorporates stock-out based dynamic substitutions, substitution costs, and service level requirements. Combining with the method developed to estimate the substitution probabilities, the proposed method works directly with the point-of-sales data and suggests order-up-to-levels to the retailers.

This paper is organized as follows. Section 2 provides a description of the problem. In Section 3, an exact analysis of inventory system's performance for the two-product case is presented. Section 4 presents deterministic and probabilistic approaches to approximately compute the performance measures of interest. Section 5 provides a computational analysis of the approximation approaches. The optimization of the order-up-to levels under stock-out based dynamic substitution is investigated with numerical results in Section 6. Section 7 concludes the paper.

2. Problem description

We consider a retailer that stocks and sells N products in a category. Demand for Product i is a Poisson random variable with rate λ_i, $i=1,\dots,N$. If a customer, whose first-choice product is Product i, cannot find it on the shelf, she may substitute it with Product j with probability α_{ij}. The substitutions probabilities, which can be estimated with methods discussed in Anupindi et al. (1998) and Karabatı et al. (2009), are an input of our problem. We assume that the customers make only one substitution attempt, and the demand is lost if their second-choice product is not available either. Kök et al. (2008) state that it is possible to approximate a multiple-substitution attempt model with a single-attempt model by adjusting the parameters. Furthermore when service levels are reasonably high, most customers find their first- or second-choice product on the shelf, eliminating the possibility of a second substitution attempt.

The retailer uses a fixed review period, order-up-to level system to control the inventory. The review period is equal to T time units, and the order-up-to level for Product i is Q_i, $i=1,\dots,N$. The demand of Product i during the review period is denoted by D_i, and is a Poisson random variable with rate $\lambda_i T$.

2.1. Performance measures

The performance measures we are interested in are the expected sales (total, direct, and through substitution) of products, the expected service and inventory levels, and system's expected profit during a review cycle:

Expected sales: The expected total sales of Product i during a review cycle is denoted by $S_i, i=1,\dots,N$. The expected number of units of Product i sold to the customers of Product j during a review cycle, who substituted Product j with Product i due to the unavailability of Product j, is denoted by S_{ji}, $j,i=1,\dots,N; j\neq i$. Therefore, the expected number of substitution sales of Product i during a review cycle is equal to $\sum_{j\neq i} S_{ji}$. The expected number of units of Product i sold during a review cycle to the customers of Product i, i.e., direct sales of Product i, denoted by S_{ii}, $i=1,\dots,N$, is then equal to $S_i - \sum_{j\neq i} S_{ji}$.

Service levels: In a multi-item retail setting with dynamic demand substitution, service levels can be measured in different ways. Measuring service level as the ratio of expected sales to expected demand is harder to interpret, since this measure can go over 100% as a result of substitutions. In addition, this measure does not differentiate between sales to customers who buy these products as their first choices and sale to customers who buy these products as a result of substitution due to not finding their first choice in stock. Accordingly, we use a service level measure, SL_i, $i=1,\dots,N$, that is defined as the ratio of total direct sales of Product i to the total demand of Product i during a review cycle:

$$SL_i = \frac{S_{ii}}{\lambda_i T}. \tag{1}$$

Inventory level: The average inventory level for product i is denoted by $\bar{I}_i$, $i=1,\dots,N$.

2.2. Optimization problem

The per unit profit of Product i is $\pi_i = p_i - c_i$, $i = 1,2,\dots,N$ where p_i (c_i) is the retail price (cost) of product i. The inventory carrying cost rate per unit per review time is h% of a product's cost, and the per unit substitution cost when a customer of Product i substitutes Product i with another product is s_i, $i = 1,2,\dots,N$.

The expected total profit Π obtained per unit time can be expressed as

$$\Pi = \frac{1}{T}\sum_{i=1}^{N}\left(\pi_i S_i - \bar{I}_i c_i h - s_i \sum_{j\neq i} S_{ij}\right). \tag{2}$$

We use maximizing the expected total profit per unit time while maintaining a minimum service level for the direct customers for each product as the main optimization criterion:

$$(OOLPS)\ \underset{Q_1,\dots,Q_N}{\text{Max}}\ \Pi S_{ii} \geq T\lambda_i\gamma_i, \quad i = 1,2,\dots,N, \tag{3}$$

where $0 \leq \gamma_i < 1$ is the minimum service level for Product i.

Since we consider stock-out based substitution, the requirement to maintain a minimum service level for direct customers makes it sure that the retailer carries this product in its assortment. We note that the inventory management problem considered in this study is different from the assortment planning where a retailer decides on the assortment of products offered to the customers. When the stock-out based substitution effects are taken into consideration, it is possible to come across a situation where it might be possible to increase expected profit by deliberately keeping a very low (or no) inventory for a particular product. In this case, realized substitution to a higher-margin product from the customers who demanded this product can make up for lost sales and even increase profits. Even in this case, a retailer may want to carry this item in its assortment for strategic reasons. The minimum service level requirement captures this trade off and forces the order-up-to levels to be set at the minimum levels that yield the desired direct service levels to customers. Obviously as the minimum service level requirement decreases, the expected profit increases.

We note that s_i may be used to capture long term effect of substitutions, such as changes in the repeat visits, especially when customers are forced to substitute by the inventory management policy of the retailer. A decision maker can use s_i as a way to control the proportion of direct sales of a given product to the total sales of that product. In order to show this, note that Eq. (2) can be rewritten as

$$\Pi = \frac{1}{T}\sum_{i=1}^{N}\left(\pi_i S_{ii} + (\pi_i - s_i)\sum_{j\neq i} S_{ji}\right) - \bar{I}_i c_i h.$$

If $s_i = 0$, then there is no additional cost of substitution and direct and indirect sales are of equal importance. However, if $s_i = \pi$ then indirect sales do not contribute to the profit and the optimization yields a solution that sets S_{ii} as high as possible. Accordingly, a decision maker can control the relative importance of direct sales by setting $0 \leq s_i \leq \pi$.

The optimization problem given in Eq. (3) is a very difficult one. The main difficulty arises from performance evaluation. Namely, *evaluating* the objective function and the constraints for given values of decision variables are not analytically tractable. That is there is no closed-form expression for S_i, S_{ij}, and $\bar{I}_i$ as a function of λ_i, α_{ij}, and Q_i. One way of solving this problem could be using simulation together with an evolutionary search algorithm. However, a simulation-based optimization approach for this problem may suffer from long run lengths due to the difficulty of estimating very low substitution numbers.

In this study, we use analytical approximations to evaluate the performance of the system. In other words, we derive analytical functions that give S_i, S_{ij}, and $\bar{I}_i$ approximately as a function of λ_i, α_{ij}, and Q_i. This approach is computationally very efficient compared to simulation to evaluate the performance of the system. We will show that the approximations are very accurate.

Since the approximations of direct sales, and total sales are non-linear, and approximations of expected inventory levels involve integer variables that determine the depletion order of the products as it will be shown in the following sections, (*OOLPS*) is still a difficult optimization problem. To approximately solve the problem, we resort to a genetic algorithm where the expected inventory levels are estimated with the mean-value approximation (Section 4.1), and expected direct and substitution sales are estimated with the two-moment approximation (Section 4.2).

3. Exact performance evaluation of the two-product case

In order to motivate the approach we used in our approximation method for evaluating the performance of the system, we first consider the performance evaluation of the inventory system for two products. Namely, given the demand rates, substitution structure, review period, and the order-up-to levels, we present analytical methods for the two-product case to determine the expected sales of each product, expected number of substitutions between products, and service levels achieved for each product, and service level achieved by the system.

The exact analysis we present in this section and also the approximation method that is given in the following section are based on determining the expected duration of substitution between the two products. The length of period where Product j is substituted for Product i during one review period is denoted with Γ_{ij}. Then the expected number of substitutions from Product i to Product j is $\lambda_i\alpha_{ij}E[\Gamma_{ij}]$.

Let T_i, $i = 1,2$, be the time when the inventory of Product i is depleted. If $T < \min\{T_1, T_2\}$ then there will be no substitution during the review period and the demands will be satisfied directly from the inventory, i.e., $\Gamma_{12} = 0$. Alternatively, when $T_1 < T < T_2$, Product 1 will be depleted at time T_1 and it will be substituted with Product 2 until the end of the review period, i.e., $\Gamma_{12} = T - T_1$. Finally, when $T_1 < T_2 < T$, the product 1 will be depleted at time T_1 and it will be substituted with Product 2 until T_2. In this case, $\Gamma_{12} = \tau_{12} = T_2 - T_1$ where τ_{12} is the length of the period Product 1 is substituted with Product 2. Then, the substitution duration Γ_{12} can be expressed as

$$\Gamma_{12} = \begin{cases} T - T_1, & T_1 < T < T_2, \\ \tau_{12}, & T_1 < T_2 < T, \\ 0 & T < \min\{T_1, T_2\}. \end{cases} \tag{4}$$

When $T_1 < T_2 < T$, Product 1 is depleted before Product 2 and T_1 is the sum of Q_1 exponentially distributed random variables with rate λ_1. Then, given $T_1 < T_2$, T_1 has an Erlang distribution with Q_1 stages and rate λ_1 for each stage

$$P[T_1 < t \mid T_1 < T_2] = 1 - \sum_{j=0}^{Q_1-1} \frac{(\lambda_1 t)^j}{j!} e^{-\lambda_1 t}. \tag{5}$$

The probability density function of T_1 when $T_1 < T_2$ is denoted with $f_{T_1}(t)$.

When $T_1 < t < T_2 < T$, the inventory of Product 1 is depleted at time T_1 and Product 2 is still available. In this period, the demand rate of Product 2 increases to $\lambda_2 + \alpha_{12}\lambda_1$. Then the distribution of τ_{12} is also Erlang with rate $\lambda_2 + \alpha_{12}\lambda_1$ and $I_2(T_1)$ stages where $I_2(T_1)$ is the inventory level of Product 2 at the beginning of the substitution period. Since both I_2 and T_1 are random variables that depend on random demands of Products 1 and 2, the distribution of τ_{12} is a compound probability distribution.

During the period $t<T_1$, both of the products are available and the demands for these products are satisfied directly from their own stock. During the period $[0,T_1]$, the number of units of Product 2 sold has a Poisson distribution with rate λ_2. Therefore, the distribution of the inventory level is given as

$$P[I_2(t)=n|T_1=t]=\frac{(\lambda_2 t)^{Q_2-n}}{(Q_2-n)!}e^{-\lambda_2 t}. \tag{6}$$

Since T_1 has an Erlang distribution with Q_1 stages and rate λ_1 for each stage as given in Eq. (5), the steady-state distribution of the inventory level of Product 2 at the beginning of the substitution period, i_2 can be given as

$$P[i_2=n]=\int_0^T P[I_2(t)=n|T_1=t]f_{T_1}(t)\,dt. \tag{7}$$

By using Eqs. (5) and (6), the probability distribution of i_2 can be written as

$$P[i_2=n]=\frac{(Q_1+Q_2-n-1)!}{(Q_1-1)!(Q_2-n)!}\frac{\lambda_1^{Q_1}\lambda_2^{Q_2-n}}{(\lambda_1+\lambda_2)^{Q_1+Q_2-n}}\left(1-\sum_{j=0}^{Q_1+Q_2-n-1}\frac{(\lambda_1+\lambda_2)^j T^j}{j!}e^{-(\lambda_1+\lambda_2)T}\right). \tag{8}$$

Therefore, the probability distribution of τ_{12} can be written as

$$P[\tau_{12}<t|T_1<T_2<T] = \sum_{n=0}^{Q_2}\left(1-\sum_{j=0}^{n-1}\frac{(\lambda_2+\alpha_{12}\lambda_1)t^j}{j!}e^{-(\lambda_2+\alpha_{12}\lambda_1)t}\right)P[i_2=n]. \tag{9}$$

Since $T_2=T_1+\tau_{12}$, once the distributions of T_1 and τ_{12} are determined, T_2 is obtained from the convolution of T_1 and τ_{12}. Finally, $E[\Gamma_{12}]$ is evaluated from Eq. (4) with the distributions of T_1, T_2, and τ_{12}. The case $T_2<T_1$ is similar and yields $E[\Gamma_{21}]$. The expected number of substitutions are directly computed as $S_{12}=\lambda_1\alpha_{12}E[\Gamma_{12}]$, $S_{21}=\lambda_2\alpha_{21}E[\Gamma_{21}]$, $S_{11}=\lambda_1 E[\min\{T_1,T\}]$ and $S_{22}=\lambda_2 E[\min\{T_2,T\}]$.

Although this approach yields the expected direct and substituted sales numbers exactly, extending this method to more than two products is not practical due to the large number of cases that need to be considered.

The exact analysis of the system with more than two products can be done by modeling the system as a Markov Chain. However, this method will suffer from exponential growth of the number of states with the number of products. Therefore this method is not a feasible alternative for exact performance evaluation.

Given the limitations of the exact analytical methods for systems with more than two products, in the next section, we present an approximation method to evaluate the performance of the system.

4. Approximate performance evaluation of the multi-product case

In this section, first we present a mean-value approximation to determine the average inventory levels $\bar{I}_i$ and then we present a two-moment approximation to determine the expected number of substitutions S_i, S_{ij}. Both of these methods use the approach we used in the preceding section to determine the substitution times between different products.

4.1. Mean-value approximation of the multiple-product case

Our simulation studies show that the dynamics of the evolution of inventory levels is mainly determined by the average values of the system parameters. In other words, the average inventory levels can be accurately evaluated by modeling the system only with the mean-values of the customer arrival and choice processes. However, when the mean-values are used, the substitutions quantities that are affected by the stochastic nature of these parameters are poorly estimated. In the next section, we focus on a two-moment approximation to determine the expected substitution numbers.

In a multiple-product setting with given order-up-to levels and simultaneous replenishments, the inventory system does not experience any stock-out instances in the period that starts with the replenishments and ends with the depletion of one of the product inventories or completion of the review period. Let l_1 be the index of the product which has the smallest (Q_i/λ_i) ratio: $l_1=\arg\min_i (Q_i/\lambda_i)$. If $(Q_{l_1}/\lambda_{l_1}>T)$, then, under the assumption that the customer arrival rates are constant, the inventory system does not experience any stock-outs, and, therefore, substitutions. Let T_{l_1} be equal to $\min\{T,Q_{l_1}/\lambda_{l_1}\}$. If $T_{l_1}<T$, then, in the period that follows T_{l_1}, the effective arrival rates of other products increase due to the substitutions the customers of Product l_1 may make. If we assume that the customers make substitutions with fixed proportions that are equal to the substitution probabilities, in the period that follows T_{l_1}, the effective arrival rate of Product i, $i\neq l_1$, increases by $\lambda_{l_1}\alpha_{l_1,i}$. Let l_2 be the product index with

$$l_2=\arg\min_{i;i\neq l_1}\frac{Q_i-\lambda_i T_{l_1}}{\lambda_i+\lambda_{l_1}\alpha_{l_1,i}}.$$

We note that $Q_i-\lambda_i T_{l_1}$ in the above relationship corresponds to the inventory level of Product $i,i\neq l_1$, at time T_{l_1}. Similarly, let T_{l_2} be equal to $\min\{T,T_{l_1}+Q_{l_2}-\lambda_{l_2}T_{l_1}/\lambda_{l_2}+\lambda_{l_1}\alpha_{l_1,l_2}\}$. If $T_{l_2}<T$, then, in the period that follows T_{l_2}, the effective arrival rate of Product $i,i\neq l_1,l_2$ becomes $\lambda_i+\sum_{k=1}^{2}\lambda_{l_k}\alpha_{l_k,i}$.

In Fig. 1, we present a graphical representation of a system with three products, and with $l_1=1$ and $l_2=3$. In the example presented in Fig. 1, Product 2 is not depleted within the review period, and completes the period with positive stock.

The above outlined approximation scheme can be generalized by defining l_n and T_{l_n}, $n=1,\ldots,N$ as follows:

$$l_n=\arg\min_{i;i\neq l_1,\ldots,l_{n-1}}\frac{Q_i-\sum_{j=1}^{n-1}(T_{l_j}-T_{l_{j-1}})(\lambda_i+\sum_{k=1}^{j-1}\lambda_{l_k}\alpha_{l_k i})}{\lambda_i+\sum_{j=1}^{n-1}\lambda_{l_j}\alpha_{l_j i}},$$

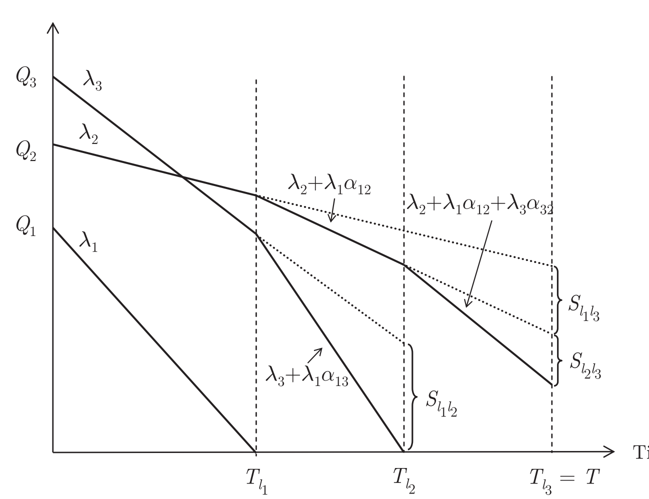

Fig. 1. Graphical representation of the mean-value approximation.

and

$$T_{l_n}=\min\left\{T, T_{l_{n-1}}+\frac{Q_{l_n}-\sum_{j=1}^{n-1}(T_{l_j}-T_{l_{j-1}})(\lambda_{l_n}+\sum_{k=1}^{j-1}\lambda_{l_k}\alpha_{l_k l_n})}{\lambda_{l_n}+\sum_{j=1}^{n-1}\lambda_{l_j}\alpha_{l_j l_n}}\right\}.$$

Once the l_n and T_{l_n}, $n=1,\ldots,N$ values are computed, the average inventory levels can be computed in a straightforward manner by considering beginning and ending inventories of each product in periods defined by the $T_{l_n}, n=1,\ldots,N$ values. As illustrated in Fig. 1, the substitutions between the products can be computed comparing the direct and effective demand rates of each product in periods formed by the $T_{l_n}, n=1,\ldots,N$ values. However, as we have noted earlier, we use the deterministic approximation only to estimate the average inventory levels.

4.2. Two-moment approximation of the multiple-product case

In this section, we will use the approach used in the exact analysis of the two-product case, presented in Section 3 to develop an approximation method for the multiple-product case.

In order to develop the approximation method, we make a number of simplifying assumptions to make the problem more tracktable. The validity of this approach is tested numerically by comparing the accuracy of the approximation with simulation.

We first assume that the retailer sets the order-up-to levels in such a way that the probability of observing stock out during the review period $P[D_i \geq Q_i]$ is small. This assumption implies that the holding cost is much lower than the cost of losing sales which is the case for majority of the product categories. With this assumption, the probability that three or more products will be depleted before the review period is very small and can be neglected. As a result, we only consider the cases where at most two products are depleted before the review period. This allows us to use pairwise substitutions to determine the expected substitution numbers.

Let $T_i < T_j$ and $T_k > T$, $k \neq i,j$. The distribution of T_i is Erlang with rate λ_i and Q_i stages (see Section 3). This random variable can be approximated with a normal random variable with mean

$$\mu_i = E[T_i] = \frac{Q_i}{\lambda_i}, \tag{10}$$

and variance

$$\sigma_i^2 = \mathrm{Var}[T_i] = \frac{Q_i}{\lambda_i^2}. \tag{11}$$

Note that this approximation follows the central limit theorem and is quite accurate for large values of Q_i. Similarly, when $T_i < T_j$, assuming constant demand and substitution rates, T_j can be approximated with a normal distribution with mean

$$\mu_j = E[T_j] = \frac{Q_j-(Q_i/\lambda_i)\lambda_j}{\lambda_j+\alpha_{ij}\lambda_i}+\frac{Q_i}{\lambda_i}, \tag{12}$$

and variance

$$\sigma_j^2 = \mathrm{Var}[T_j] = \frac{Q_j-(Q_i/\lambda_i)\lambda_j}{(\lambda_j+\alpha_{ij}\lambda_i)^2}+\frac{Q_i}{\lambda_i}\frac{1}{\lambda_j}. \tag{13}$$

In our approximation method, we compare expected depletion times of products without substitution to set their approximate distributions. For Products i and j, if $(Q_i/\lambda_i < Q_j/\lambda_j)$, we use Eqs. (10) and (11) for the approximate distribution of T_i and Eqs. (12) and (13) for the approximate distribution of T_j. Similarly, if $(Q_i/\lambda_i > Q_j/\lambda_j)$, we use Eqs. (12) and (13) for the approximate distribution of T_i and Eqs. (10) and (11) for the approximate distribution of T_j.

We also assume that T_i and T_j are independent. Substitution from Product i to j causes depletion of product j at an earlier time compared to the no-substitution case. However, when $T_i < T < T_j$, this change will have a minor effect on $E[(T-T_i)^+]$. Similarly, when $T_i < T_j < T$, the effect on $E[(T_j-T_i)^+]$ will not be significant.

With this assumption T_j-T_i is approximately normal with mean

$$\mu_{ji} = E[T_j-T_i] = E[T_j]-E[T_i], \tag{14}$$

and variance

$$\sigma_{ji}^2 = \mathrm{Var}[T_j-T_i] = \mathrm{Var}[T_j]+\mathrm{Var}[T_i]. \tag{15}$$

Let Γ_{ij} be the length of the time Product i is substituted with Product j.

$$\Gamma_{ij} = \begin{cases} T-T_i, & T_i < T < T_j, \\ T_j-T_i, & T_i < T_j < T, \\ 0, & T < \min\{T_i, T_j\}. \end{cases} \tag{16}$$

Following Eq. (16),

$$E[\Gamma_{ij}] = E[(T-T_i)^+ \mid T_i < T < T_j]P[T_i < T < T_j] \tag{17}$$

$$+E[(T_j-T_i)^+ \mid T_i < T_j < T]P[T_i < T_j < T]. \tag{18}$$

Since it is assumed that T_i and T_j are independent

$$E[\Gamma_{ij}] \cong E[(T-T_i)^+ \mid T_i < T < T_j]P[T < T_j] \tag{19}$$

$$+E[(T_j-T_i)^+ \mid T_i < T_j < T]P[T_i < T]P[T_j < T]. \tag{20}$$

With the normal approximation of T_i and T_j, we can determine $E[(T-T_i)^+ \mid T_i < T < T_j]$, $E[(T_j-T_i)^+ \mid T_i < T_j < T]$ and $P[T_i < T]$ directly.

For normally distributed random variable T_i

$$E[(T-T_i)^+ \mid T_i < T < T_j] = T-\mu_i+\sigma_i\eta\left(\frac{T-\mu_i}{\sigma_i}\right), \tag{21}$$

where $\eta(z)$ is the expected number of units short of a standard normal random variable. $\eta(z)$ is determined as

$$\eta(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}}(x-z)e^{-1/2x^2}\,dx = \phi(z)-z\Phi(z), \tag{22}$$

where $\phi(z)$ and $\Phi(z)$ are the density function and the complementary cumulative distribution function of the standard normal given as $\phi(z)=e^{1/2z^2}$ and $\Phi(z)=\int_z^\infty \phi(z)\,dz$.

Eq. (22) now yields

$$E[(T-T_i)^+ \mid T_i < T < T_j] = T-\mu_i+\sigma_i\left(\phi\left(\frac{T-\mu_i}{\sigma_i}\right)-\left(\frac{T-\mu_i}{\sigma_i}\right)\Phi\left(\frac{T-\mu_i}{\sigma_i}\right)\right). \tag{23}$$

Similarly, since T_j-T_i is approximated with a normal random variable with mean and standard deviation given in Eqs. (14) and (15), $E[(T_i-T_j)^+ \mid T_i < T_j < T]$ can be evaluated as:

$$E[(T_i-T_j)^+ \mid T_i < T_j < T] = \sigma_{ji}\left(\phi\left(\frac{\mu_{ji}}{\sigma_{ji}}\right)-\frac{\mu_{ji}}{\sigma_{ji}}\Phi\left(\frac{\mu_{ji}}{\sigma_{ji}}\right)+\phi\left(\frac{T+\mu_{ji}}{\sigma_{ji}}\right)+\frac{T+\mu_{ji}}{\sigma_{ji}}\Phi\left(\frac{T+\mu_{ji}}{\sigma_{ji}}\right)\right)-T\Phi\left(\frac{T+\mu_{ji}}{\sigma_{ji}}\right). \tag{24}$$

$E[(T-T_j)^+ \mid T_j < T < T_i]$ and $E[(T_j-T_i)^+ \mid T_j < T_i < T]$ are determined by using the above equations after interchanging indices i and j.

As a result, the expected time Product i is substituted with Product j can be written in closed form as

$$E[\Gamma_{ij}] = E[(T_j-T_i)^+ \mid T_i < T_j < T]\left(1-\Phi\left(\frac{T-\mu_i}{\sigma_i}\right)\right)\left(1-\Phi\left(\frac{T-\mu_j}{\sigma_j}\right)\right)$$
$$+E[(T-T_i)^+ \mid T_i < T < T_j]\Phi\left(\frac{T-\mu_j}{\sigma_j}\right). \tag{25}$$

Once the expected substitution times are determined, we can determine S_{ij} $i \neq j$ as

$$S_{ij} = E[\Gamma_{ij}]\lambda_i\alpha_{ij}, \tag{26}$$

and S_{ii} as

$$S_{ii} = E[\min\{Q_i, \lambda_i T\}]. \tag{27}$$

Following the normal approximation, S_{ii} can be approximated as

$$S_{ii} = \lambda_i T - \sqrt{\lambda_i T}\left[\phi\left(\frac{Q_i - \lambda_i T}{\sqrt{\lambda_i T}}\right) - \left(\frac{Q_i - \lambda_i T}{\sqrt{\lambda_i T}}\right)\left(1 - \Phi\left(\frac{Q_i - \lambda_i T}{\sqrt{\lambda_i T}}\right)\right)\right]. \tag{28}$$

Finally, if this approximation yields S_{ij} values such that $\sum_j S_{ij} \geq Q_i$, we normalize the values such that

$$S'_{ij} = S_{ij}\frac{Q_i}{\sum_j S_{ij}}.$$

Note that the approximation method presented in this section yields the performance measures of interest in closed form as a function of system parameters and decision variables. Therefore this method is computationally superior to a simulation approach. Moreover, it is better suited to be used in optimization. Before focusing on optimization, we first evaluate the accuracy of the proposed method in the next section.

5. Computational study: accuracy of the approximation approaches

In this section, we present a computational analysis of the approximation quality of the approaches developed in Section 4. In the computational analysis, we consider a set of randomly created four-product problems. As noted in Karabatı et al. (2009), when the observed number of substitutions is not statically significant, estimation of substitution probabilities is a very challenging task. To deal with this issue, a certain number of products with similar characteristics can be lumped together for analysis purposes. For example, modeling the first three products with the largest market shares explicitly, and lumping the other products with smaller market shares into a single product yields a model with four products. The performance of a model of this size can be evaluated quite accurately by using the approximation method presented in this study and also can be optimized effectively.

The products' order-up-to levels are set to satisfy a randomly selected fill rate without taking the substitution effect into account. Three different service level ranges, [60%, 99%], [70%, 99%], [80%, 99%], are used in problem generation, and the target service level of each product is randomly selected using a uniform distribution between the lower and upper limits of the ranges. The demand rates of the Products 1 and 2 (3 and 4) are generated using a uniform distribution in the [15,25] ([5,15]) range. The customer choice model is assumed to be *Market-Share Based* (see Smith and Agrawal, 2000; Netessine and Rudi, 2003; Kök and Fisher, 2007) where the substitute product is chosen according to the substitution probability matrix $\alpha_{ij} = \theta(\lambda_j / \sum_{l \in N\setminus\{i\}} \lambda_l)$, $i,j = 1,2,\ldots,N$, and $i \neq j$. The review time is taken as 20 time units, and the substitution probability, θ, is taken as 60%. For each service level range, 120 problems are randomly generated and simulated for 10 independent replications with 50 review periods in each replication. Ten replications of the simulation are run to generate the average value for a particular set of data.

In Table 1, we report the mean-value approximation's performance for products' average inventory levels, and stochastic approximation's performance for products' total direct sales and total sales. The average and maximum approximation errors are reported, over 120 problems in each row of Table 1, relative to the average performance observed over 10 replications of the simulation model in each problem instance.

We note that, in order to capture the probabilistic nature of substitutions, all performance measures, with the exception of

Table 1
Approximation errors.

Service level distribution	Estimation errors					
	Inventory (%)		Total sales (%)		Direct sales (%)	
	Average	Maximum	Average	Maximum	Average	Maximum
[60%, 99%]	0.587	2.287	0.005	0.894	0.386	2.972
[70%, 99%]	0.510	1.798	0.010	0.905	0.461	2.215
[80%, 99%]	0.422	1.295	0.071	1.897	0.529	2.584

inventory levels, are estimated with the probabilistic approximation. The figures presented in Table 1 indicate that the inventory levels, products' total direct and total sales can be closely estimated with approximation approaches described earlier.

In Fig. 2, we report the average performance of the probabilistic approach in estimating the number of substitutions between products. The accuracy of the probabilistic approach is assessed by comparing the substitution numbers obtained by using the probabilistic approximation with the substitution numbers obtained with the simulation. We first note that the performance of the approximation approach is dependent on the rate of the realized substitutions: when the number of substitutions per review period is low, the observations are not statistically significant, and approximations are relatively poor. In light of this observation, we report the approximation performance for three different minimum levels of substituted demand. For example, when the minimum level is taken as 1%, and when the estimated number of substitutions from Product i to Product j, i.e., S_{ij}, is less than $\lambda_i \times T \times 0.01$, substitutions from Product i to Product j are considered to be statistically insignificant, and approximation error with S_{ij} is not included in the reported average approximation errors.

In Fig. 2, we observe that the approximation quality increases when the minimum substitution level threshold for approximation error calculations and the variability in service levels increase. As we discuss in the next section, where a model to find the optimal order-up-to levels under substitution is presented, we need better substitution approximations in cases where the service level of a product is deliberately set to low channel some of its demand to other products through substitutions. In these instances, because the number of substitutions is high, the approximation quality of the probabilistic approach will be high too.

6. Optimization of the order-up-to levels under stock-out based dynamic substitution

The preceding sections show that we can evaluate the performance of the system efficiently and accurately by using the approximation method we presented. Now, we focus on the optimization problem we outlined in Section 2. The basic question we want to answer is the following: *if a retailer can estimate the substitution structure among different products, how much can the retailer improve its profits by determining the order-up-to levels accordingly?* Our previous work on estimation the substitution structure (Karabatı et al., 2009) gives a practical method to estimate the demand characteristics and stock-out based substitution probabilities by only using the Point-of-Sales data. As a result, in this section, we focus on the improvements in expected profits by taking the substitution structure into account and show that a retailer can significantly improve its profits by managing its inventory in a way that captures substitution

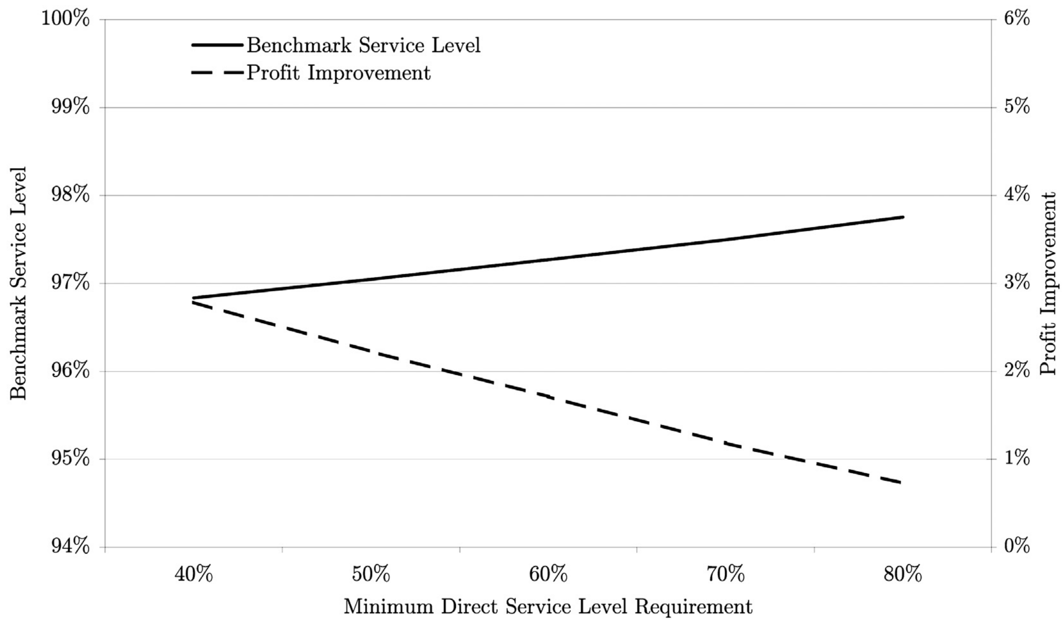

Fig. 2. Substitution estimation errors.

dynamics especially when the service level is low and profit margins of the products are different.

The approximation and optimization method presented in this study are implemented in software as a Visual Basic program that runs as an Excel Add-in. The program works with the point-of-sales data, first estimates the substitution probabilities, then optimizes the system and compares with simulation. The system parameters are entered from an Excel spreadsheet. For the computational results reported in this section, the parameters of the genetic algorithm are set as the following: population size: 100, generations: 250, crossover probability: 0.85, and mutation probability: 0.05.

We first discuss a specific case and then analyze the results over a larger set of problems.

6.1. A specific case

We consider a four-product problem with $T=20$, $h=1.64\%$, and the following parameters:

Product	p_i	π_i	s_i	λ_i
1	6.00	0.60	0.06	12
2	6.00	0.60	0.06	12
3	8.00	1.20	0.12	8
4	10.00	2.00	0.20	6

The substitution probabilities are as follows:

α_{ij}	1	2	3	4
1	–	0.25	0.1	0.1
2	0.25	–	0.1	0.1
3	0.1	0.1	–	0.3
4	0.1	0.1	0.1	–

When we do not take the effects substitution into account, for a fill-rate service level of 99%, the order-up-to levels are set as $Q=(251,251,170,130)$. The total profit of the system, when simulated with substitutions, is computed as $\Pi=670.98$.

The optimization model yields, under a 40% minimum service level constraint, a solution with $Q=(236,243,171,136)$, and a total profit of $\Pi=672.90$. The realized total substitutions from Products 1, 2, 3, and 4 are 3.25, 1.41, 0.52, and 0.11, respectively. Although the cost of substitution is low relative to the differences between profit margins of the products, there is no forced substitution in the optimal solution. This is mainly due to the substitution probabilities of the problem, because when we decrease the order-up-to levels of Products 1 and/or 2 to increase substitution from these products to more profitable products, i.e., Products 3 and 4, a substantial portion of their demand is lost.

When we increase $\alpha_{1,3}$ and $\alpha_{2,3}$ to 0.3, the optimal solution becomes $Q=(97,276,207,139)$, with a total profit of $\Pi=680.00$. Because of its low order-up-to level, the direct sales of Product 1 realizes as 40%, the minimum level required by the direct service level constraint. The realized total substitutions from Products 1, 2, 3, and 4 are now 89.35, 1.55, 0.82, and 0.29, respectively. We note that, in the optimal solution, forced substitution is observed in only one of the less profitable products. The substitution probabilities between Products 1 and 2 ($\alpha_{1,2}=\alpha_{2,1}=0.25$) are significant, and lowering the order-up-to levels of Products 1 and 2 simultaneously results in lost sales when customers of Products 1 and 2 attempt to substitute their preferred product with another one.

When we increase $\alpha_{1,3}$ and $\alpha_{2,3}$ to 0.5, a major portion of customers of both Product 1 and Product 2 are forced to substitute, because, in the optimal solution, the order-up-to levels are set as $Q=(98,99,302,149)$. The total profit of the system is now $\Pi=715.60$. Although we lost 25% of demands of Products 1 and 2, we recover lost sales with the increased profits when 28% of customers of Products 1 and 2 substitute with Product 3.

6.2. Randomly generated problems

In this section we consider a larger set of randomly created problems to study the impact of substitutions on system's profit performance.

We consider four-product problems with identical costs, four demand scenarios ((10,10,10,10), (15,15,5,5), (12,12,8,8), and

(20,10,5,5)), four levels of substitution costs (0%, 5%, 10%, and 25% of products' profit margins), and nine market share dependent profit margin scenarios where $\pi_i = (A - B\lambda_i / \sum_j \lambda_j)$ with $(A,B) \in \{(0.2,0.3),$ $(0.2,0.2),(0.2,0.1)$,$(0.1,0.15),(0.1,0.1),(0.1,0.05),(0.05,0.075),(0.05,$ $0.05),(0.05,0.025)\}$. We note that, according to above expression, profit margins are negatively correlated with market shares.

The problem generation scheme results in 144 test problems with profit margins that are negatively correlated with market shares, and identical product costs. The customer choice model is again assumed to be *Market-Share Based* (see Section 5).

For comparison purposes, we create a benchmark solution for every problem instance. We first solve a given problem instance optimally and then compute the value of the initial inventory by multiplying the optimal order-up-to levels by the corresponding product costs. We then find the fill-rate service level that would result in the same inventory value when the order-up-to level of each product is determined, independently and by ignoring the effects of substitution, with this particular service level. We then report the profit performance of the optimization approach relative to the profit obtained in the benchmark solution.

In Fig. 3, we analyze the relationship between the profit improvement and substitution probability θ. For every value of θ in set (60%, 70%, 80%, 90%, 100%), we solve the 144 test problems with $T=20$, $h=1.37\%$, and minimum direct service level of 40%, and report the average profit improvement over the benchmark solutions. The half-width confidence intervals for profit improvements are (0.68%, 0.71%, 0.75%, 0.81%, 0.9%) for $\theta \in (60\%,70\%,80\%,90\%,100\%)$. The results presented in Fig. 3 indicate that, by accounting for substitutions, the performance of the inventory system can be substantially improved. As expected, the higher the substitution probability, the larger the profit improvement.

In Fig. 4, we analyze the relationship between the profit improvement and minimum direct service level requirement. For every value of in set (40%, 50%, 60%, 70%, 80%), we solve the 144 test problems with $T=20$, $h=1.37\%$, and $\theta=60\%$, and report the average profit improvement over the benchmark solutions. The half-width confidence intervals for profit improvements are (0.68%, 0.54%, 0.41%, 0.29%, 0.18%) for $\theta \in (40\%,50\%,60\%,70\%,80\%)$. The results presented in Fig. 4 show that minimum direct service level requirement can have a significant impact on the profit improvement, and it is difficult to achieve substantial improvements when the inventory system operates under a high level of minimum direct service level requirement.

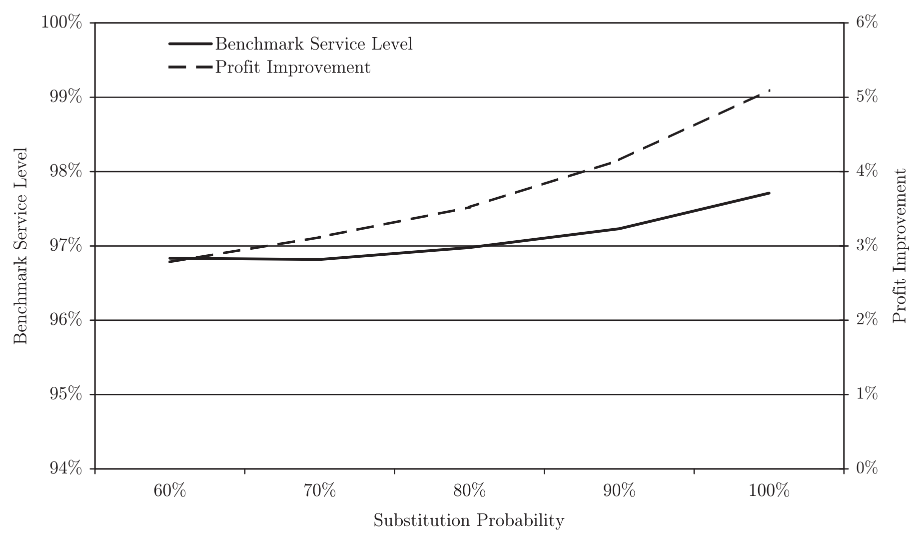

Fig. 3. Profit improvement and substitution probability: negatively correlated profit margin and market share.

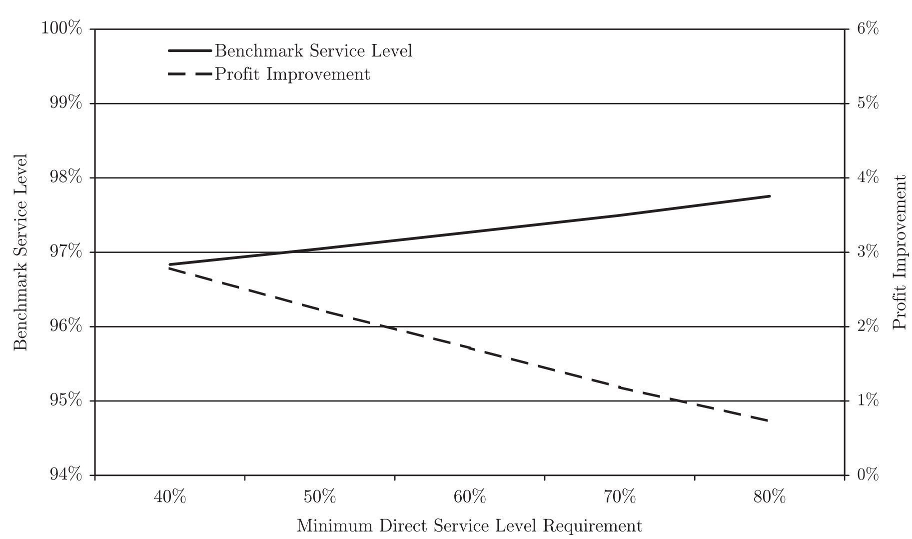

Fig. 4. Profit improvement and minimum direct sales requirement: negatively correlated profit margin and market share.

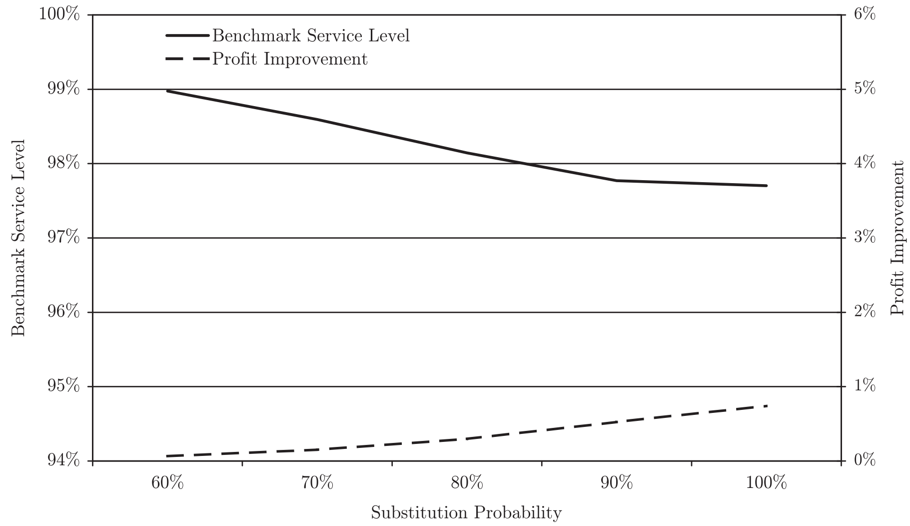

Fig. 5. Profit improvement and substitution probability: positively correlated profit margin and market share.

Finally in Fig. 5, we present the analysis of Fig. 3 when profit margins and market shares are positively correlated, i.e., $\pi_i = (A + B(\lambda_i / \sum_j \lambda_j)$. The half-width confidence intervals for profit improvements are (0.01%, 0.03%, 0.05%, 0.08%, 0.10%) for $\theta \in (60\%, 70\%, 80\%, 90\%, 100\%)$. The results presented in Fig. 5 indicate that when profit margins and market shares are positively correlated, it is difficult to substantially improve the performance of the inventory system by accounting for substitutions.

We assume that the retailer sets the order-up-to levels in such a way that the probability of observing stock out during the review period is small. This assumption implies that the holding cost is much lower than the cost of losing sales which is the case for majority of the product categories. With this assumption, the probability that three or more products will be depleted before the review period is very small and can be neglected. With this assumption, at most two products are depleted before the review period. Our method allows us to use pairwise substitutions to determine the expected substitution numbers in this case. Under these conditions, the accuracy of the approximation method does not increase with the number of items or with shortages over the review interval increases.

If the review interval is much longer than the expected depletion times of the products, there will be no shortages and therefore there will be no substitutions. On the contrary, if the review period is longer than the expected depletion times but the probability of observing stock out during the review period is small, the approximation method will be accurate. As the review period increases, the probability of observing stock out increases and this will decrease the accuracy of the approximation method. However, note that the substitution numbers will also decrease after a certain time period due to unavailability of products.

7. Conclusions

In this paper, we consider the inventory management problem of a product category with stock-out based dynamic demand substitutions and lost sales. Although the retailer can only indirectly affect customers' decisions through his inventory management decisions, as discussed in the literature, ignoring product substitutions in managing the inventories may result in sub-optimal performance. We present an approximate approach to find the order-up-to levels in a profit maximization setting with profit margins, inventory holding and substitution costs, and service level constraints. Through a computational study, we show that, by explicitly accounting for substitutions, the performance of the inventory system can be improved. The amount of improvement depends on the minimum direct service level requirement as well as the correlation between the market share and the profit margin of the products.

By combining the method we presented in an earlier study to estimate the demand and customer choice parameters, the method we presented in this study can be used to manage inventory in a better way in retailing.

Acknowledgments

This research has been sponsored by TÜBİTAK Grant No. 106K175.

References

Anupindi, R., Dada, M., Gupta, S., 1998. Estimation of consumer demand with stock-out based substitution: an application to vending machine products. Marketing Science 17, 406–423.

Avsar, Z.M., Baykal-Gursoy, M., 2002. Inventory control under substitutable demand: a stochastic game application. Naval Research Logistics 49, 359–375.

Ernst, R., Kamrad, B., 2006. Estimating demand by using sales information: inaccuracies encountered. European Journal of Operational Research 174, 675–688.

Helvacıoğlu, F., 2009. Multi-product Inventory Control Under Stock-out Based Substitution. Master's Thesis. Koç University Department of Industrial Engineering.

Hopp, W., Xu, X., 2008. A static approximation for dynamic demand substitution with applications in a competitive market. Operations Research 56, 630–645.

Hsu, V.N., Li, C., Xiao, W., 2005. Dynamic lot size problems with one-way product substitution. IIE Transactions 37, 201–215.

Karabatı, S., Tan, B., Öztürk, C., 2009. A method for estimating stock-out based substitution rates by using point-of-sale data. IIE Transactions 41, 408–420.

Kök, A., Fisher, M., Vaidyanathan, R., 2008. Assortment planning: review of literature and industry practice. In: Agrawal, N., Smith, S.A. (Eds.), Retail Supply Chain Management: Quantitative Models for Supply Chain Management. Springer-Verlag, New York, pp. 99–154.

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

Mahajan, S., van Ryzin, G., 1998. Retail inventories and consumer choice In: Sridhar, R.T., Magazine, M.J., Ganeshan, R. (Eds.), Quantitative Models for Supply Chain Management. Springer-Verlag, New York, pp. 491–554.

Mahajan, S., van Ryzin, G., 2001. Stocking retail assortments under dynamic consumer substitution. Operations Research 49 (3), 334–351.

McGillivray, A., Silver, E., 1978. Some concepts for inventory control under substitutable demands. INFOR 16 (1), 47–63.

Nagarajan, M., Rajagopalan, S., 2008. Inventory models for substitutable products: optimal policies and heuristics. Management Science 54 (8), 453–1466.
Netessine, S., Rudi, N., 2003. Centralized and competitive inventory models with demand substitution. Operations Research 51 (2), 329–335.
Parlar, M., 1985. Optimal ordering policies for a perishable and substitutable product: a Markov decision model. INFOR 23, 182–195.
Rajaram, K., Tang, C., 2001. The impact of product substitution on retail merchandising. European Journal of Operational Research 135, 582–601.
Rao, U.S., Swaminathan, J.M., Zhang, J., 2004. Multi-product inventory planning with downward substitution, stochastic demand and setup costs. IIE Transactions 36, 59–71.
Smith, S., Agrawal, N., 2000. Management of multi-item retail inventory systems with demand substitution. Operations Research 48 (1), 50–64.
Yücel, E., Karaesmen, F., Salman, F.S., Türkay, M., 2009. Optimizing product assortment under customer-driven demand substitution. European Journal of Operational Research, 759–768.