

Assortment Planning and Inventory Decisions Under a Locational Choice Model*

Vishal Gaur[†] Dorothée Honhon[†]

December 12, 2005

Abstract

We consider a single-period assortment planning and inventory management problem for a retailer, using a locational choice model to represent consumer demand. We first determine the optimal variety, product location, and inventory decisions under static substitution, and show that the optimal assortment consists of products equally spaced out such that there is no substitution among them regardless of the distribution of consumer preferences. The optimal solution can be such that some customers prefer not to buy any product in the assortment, and such that the most popular product is not offered.

We then obtain bounds on profit when customers dynamically substitute, using the static substitution for the lower bound, and a retailer-controlled substitution for the upper bound. We thus define two heuristics to solve the problem under dynamic substitution, and numerically evaluate their performance. This analysis shows the value of modeling dynamic substitution and identifies conditions in which the static substitution solution serves as a good approximation.

*The authors are thankful to Professors Ravi Anupindi, Marshall Fisher, and Ananth Raman for their suggestions on previous versions of this paper. They are also thankful to the department editor, the associate editor and three anonymous referees for numerous helpful comments that significantly improved this paper.

[†]Department of Information, Operations and Management Science, Leonard N. Stern School of Business, New York University, 8-160, 44 West 4th St., New York, NY 10012. E-mail: vgaur@stern.nyu.edu, dhonhon@stern.nyu.edu.

1 Introduction

The multinomial logit (MNL) model has been commonly used by researchers in recent years to represent consumer choice behavior for planning assortments in a product category. Thus, our understanding of the tradeoffs involved in assortment decisions is largely based on this model. In contrast, Lancaster (1966, 1975), extending the work of Hotelling (1929), proposed a locational model of consumer choice behavior. In this model, products are perceived as bundles of attributes, and each product is defined by its location in an attribute space. Individual preferences are defined by associating a point in the attribute space with each consumer. Thus, preferences are defined on attributes rather than on the products themselves as in the MNL model. Each consumer chooses the product that is located closest to him in the attribute space and substitutes to other products in increasing order of their distances from him.

In this paper we study the optimal product assortment and inventory decisions of a retailer under a locational choice model, and contrast the operational insights thus obtained with those available from previous research using mainly the MNL model. The decisions include the number of products to offer, their locations in the attribute space and their inventory levels. These decisions encompass both product selection and product design. Thus, the decision-maker can be a manufacturer, or a retailer choosing from a potentially infinite set of available products. We take the perspective of a retailer throughout this paper. For simplicity, we consider a single-period problem of maximization of expected profit.

Lancaster developed his model for deterministic demand and uniform density of consumers in the attribute space. We use a generalized locational choice model with stochastic demand and nonuniform distribution of consumers in the attribute space. Like Lancaster, we consider horizontal product differentiation, i.e., prices and quality levels are equal across all products in the category. This assumption fits several product categories, such as yogurt with different amounts of fat-content, shirts of different colors, etc.

Our paper exploits many key differences between the locational choice model and the MNL model. First, substitution between products in our model is localized to products with specifications that are close to each other on the attribute space¹ so that a retailer can control the rates of

¹We further note that, due to localized substitution, our model does not impose the assumption of independence from irrelevant alternatives or IIA, which is often presented as the main drawback of the MNL model; see Anderson et al. (1992: p.23).

substitution between products by suitably selecting their locations to be far apart or close to each other. In contrast, substitution in the MNL model can always happen between any two products. Second, the locational model enables us to independently specify the degree of heterogeneity of the consumer population and the degree of substitutability between products. A heterogeneous population is spread out on the entire attribute space, whereas a homogeneous population is concentrated in a small subset of the attribute space. Likewise, the degree of substitutability of a product category can be measured by the maximum distance between the most preferred good of a customer and a product that gives him positive utility. Thus, the locational model provides us more parameters to control than the MNL model in understanding assortment decisions. We show that these differences have significant implications on the optimal assortment.

We consider both static and dynamic substitution. Under static substitution, a consumer chooses a first choice product (if any) by looking at the assortment but without observing inventory levels, and does not make a second choice if the first choice is unavailable. We obtain the optimal assortment under static substitution and analyze its properties. Under dynamic substitution, the consumer observes inventory levels when he arrives in the store and then chooses a product (if any) among the available ones. We derive lower and upper bounds on the optimal expected profit under dynamic substitution and propose two heuristics based on these bounds. In a numerical study using 3150 problem instances over a range of parameter values, we find that the average optimality gaps of the heuristics are 1.44% and 1.24%. The optimality gap of each heuristic decreases as mean demand increases, as consumer preferences become more homogeneous, as the profit margin increases, and as the degree of substitutability of products increases.

Our paper yields the following main insights.

- Under static substitution, products in the optimal assortment are spaced out such that there is no substitution between them regardless of the volume of demand or the distribution of customer preferences on the attribute space.
- The retailer provides higher variety under dynamic substitution than under static substitution, and locates products closer to each other on the attribute space so that consumers can derive positive utility from more than one product and the retailer is able to benefit from substitution. The number of products that give positive utility to a consumer varies with the location of that consumer on the attribute space, being generally higher in regions of greater consumer density. These differences between the static and dynamic substitution models are

significant because they show that the dynamic substitution model better fits real life observations. No comparable insights into the dynamic substitution problem are available under the MNL model.

- The optimal assortment need not include the most popular product, i.e., the product that would yield the highest expected demand if it were the only product carried in the assortment. This result occurs when stocking the most popular product creates a *fragmentation of demand* such that the economies of scale enjoyed by the most popular product are insufficient to overcome the diseconomies of scale suffered by the remaining products in the assortment. This property contrasts with results obtained with the MNL model by van Ryzin and Mahajan (1999).
- Dynamic substitution has the greatest impact on the profits of the retailer for low demand items with high customer heterogeneity and high degree of substitutability. Conversely, the static substitution solution serves as a good approximation for the dynamic substitution problem when consumers are more homogeneous in preferences, or when they are willing to substitute in a narrow region around their most preferred goods, or when the coefficient of variation of demand is low.
- The retailer may choose not to cover the entire market, i.e., leave some segments of the attribute space uncovered by any product. While an analogous result was obtained under the MNL model as well (van Ryzin and Mahajan 1999), we find that this result in our model is driven by fixed costs associated with including a product in the assortment. Thus, in our model, it is optimal to cover the entire market when fixed costs are zero.

The rest of this paper is organized as follows: §2 reviews the relevant literature; §3 presents our model; §4 and §5 analyze the model under static substitution and dynamic substitution, respectively; finally, §6 concludes with a discussion of the weaknesses of our model and directions for future research. All proofs are provided in the Appendix unless otherwise stated.

2 Literature Review

Research on assortment planning and inventory management has advanced rapidly in the recent years, particularly on modeling substitution between products using individual-level consumer

choice theory from the marketing literature. We first summarize the literature on assortment planning and inventory management under the MNL model (McFadden 1974, Guadagni and Little 1983) and its generalizations. We then relate our paper to the existing research based on locational choice models (Hotelling 1929, Lancaster 1966).

van Ryzin and Mahajan (1999) were the first to study assortment planning and inventory decisions under the MNL model for the case of static substitution with exogenous prices. They determine many properties of the optimal solution, the main being that the optimal assortment consists of the most popular products from the finite set of potential products to offer. Aydin and Ryan (2000) apply the MNL model to study the joint assortment planning and pricing problem under static substitution. They find that the optimal solution is such that all products have equal margins (i.e., the difference between price and cost). Cachon et al. (2005) generalize the consumer choice process to incorporate search costs. They show that ignoring consumer search in demand estimation can result in an assortment with less variety and significantly lower expected profits compared to the optimal solution. They further show that search costs can induce a retailer to carry an unprofitable product in its assortment to reduce consumer search. Chong, Ho and Tang (2001) consider brand-level assortment decisions using a nested MNL model including brand-width measures associated with each brand in a product category. They use a local improvement heuristic based on pairwise interchanges of product variants to modify the assortment and achieve a substantial increase in profit. Rajaram (2001) applies an assortment planning model to a catalog retailer to determine optimal inventory levels under static substitution.

The assortment planning and inventory management problem under dynamic substitution is far more complex than that under static substitution. Smith and Agrawal (2000) consider this problem using a general customer demand model described by first choice probabilities and a substitution matrix. They present a solution approach by showing that static substitution yields bounds on the demand for each product under dynamic substitution. Mahajan and van Ryzin (2001) consider this problem for a choice model requiring only a preference relation defined over all possible products and satisfying the axioms of completeness, reflexivity and transitivity (see Anderson et al. 1992:p.17). They develop a sample path gradient algorithm to determine the optimal assortment and inventory levels in order to maximize expected profits. Kok and Fisher (2004) consider this problem in the context of a supermarket chain. They show how to estimate assortment based substitution in an MNL model by leveraging data from stores with varying assortments. They present an algorithm

to solve the assortment planning and inventory problem with one-level stock-out based substitution in the presence of shelf-space constraints.

While the above papers deal with designing algorithms for assortment planning, relatively little is known about the impact of dynamic substitution on the optimal solution. In this context, Rajaram and Tang (2001) consider the problem of determining optimal inventory levels under dynamic substitution for a fixed assortment of substitutable products. They show that increase in degree of substitution significantly increases expected profits, and thus designing products to be substitutable has a large impact on profitability.

In contrast to the above literature, we use a locational model of consumer choice to determine demand. In the seminal work on such models, Hotelling (1929) studied competition on price and location between two firms in a linear city. Lancaster (1966, 1975) translated the locational model to the context of product differentiation and studied optimal variety from a welfare perspective. Lancaster (1979) analyzes different market structures under this model, such as monopolistic competition and multi-product monopoly. For a monopoly, Lancaster shows that the firm covers the entire market and offers products spaced equally from each other under the assumptions of deterministic demand with uniform customer preferences and economies of scale in the cost function of each product.

de Groote (1994) and Alptekinoglu (2004) integrate product differentiation and inventory costs in the context of locational choice models with deterministic demand and uniform customer preferences. de Groote considers a monopoly firm and analyzes the coordination between the marketing decision of product line breadth and the operations decision of production flexibility. Similar to Lancaster, de Groote shows that the firm chooses an optimal number of products such that the market is covered and product locations are equal spaced. Further, optimal variety increases with demand, length of the market and unit travel costs, and decreases with changeover costs and inventory holding costs. Alptekinoglu analyzes competitive positioning and pricing for two firms, one offering infinite variety through mass customization and the other offering a finite set of different products. He shows the counterintuitive result that the mass producer needs to reduce variety to soften price competition with the mass customizer firm.

Ansari et al. (1994, 1998) study positioning and pricing of products for one-product firms competing at the marketplace. They use two and three-dimensional generalizations of Hotelling's model with non-uniform consumer preferences modeled by uni-modal, bi-modal and generalized

beta distributions, and show that the equilibrium product positions can be asymmetric. They too assume deterministic demand, and focus on the demand perspective rather than inventory management or dynamic substitution.

Chen et al. (1998) study optimal product positioning and pricing, extending Lancaster’s model to incorporate varying prices and quality levels in the attribute space, as well as varying reservation prices of customers. They show that the optimal solution for this model under stochastic demand and static substitution can be constructed using dynamic programming by utilizing a ‘cross-point property’ to determine the demands for individual products. The cross-point property differs from the property of equally spaced product locations in that the cross-points may not be equally spaced.

Our paper contributes to the research literature in several ways. It generalizes locational choice models by considering stochastic demand, nonuniform consumer preferences, and inventory costs. It shows that the property of equally spaced product locations is obtained under this generalization when substitution is static but not when substitution is dynamic. It reports many new structural properties of the optimal assortment that contrast with those obtained using the MNL model in previous research. For further research on empirical and analytical models on consumer choice, we refer the reader to Anderson et al. (1992), Ho and Tang (1998) and Lancaster (1990).

3 Model Formulation

We consider a retailer serving a market with heterogeneous consumers over a single time period. The retailer seeks to determine the optimal number of products to stock in a given product category, their design characteristics and inventory levels in order to maximize expected profit. The products in the category are *horizontally differentiated*, i.e., they differ by characteristics that do not affect quality or price, for example, shirts of different colors or yogurt with different amounts of fat-content. Since products have homogeneous quality, we assume that they are sold at identical exogenous prices and have identical costs. Let r be the revenue per unit, c the procurement cost per unit, and K the fixed cost associated with ordering and stocking a product. Excess demand at the end of the period is lost and excess inventory is salvaged at value v per unit. We assume that $r > c > v$. Demand is generated by an individual-level locational consumer choice model based on Hotelling (1929) and Lancaster (1966, 1975, 1979) as described below. The limitations of the choice model are discussed in §6.

Locational Demand Model

The following terminology and assumptions are identical to those of Lancaster (1966, 1975, 1979). Products are perceived as bundles of characteristics and individual preferences are defined on these characteristics rather than on the products themselves. We assume that each characteristic, or *attribute*, is quantifiable. We define the *preference spectrum* or the *attribute space* to be the space of all possible combinations of levels of attributes, where each point corresponds to a potential product location in the category. The preference spectrum is denoted as \mathcal{L} , and is considered to be the real line, i.e., one attribute is sufficient to discriminate between the products.

Let the assortment carried by the retailer be represented by a *vector of product specifications* $\mathbf{b} = (b_1, \dots, b_n)$ where $n \equiv \text{length}(\mathbf{b})$ is the number of products in the assortment, and $b_j \in \mathcal{L}$ denotes the location of product j on the preference spectrum. We assume that every product specification desired by the retailer is offered by the manufacturer. We say that \mathbf{b} is a valid vector of product specifications if

$$b_j > b_{j-1} \quad \text{for } j = 2, \dots, n.$$

Since we seek to optimize the assortment with respect to n and \mathbf{b} , we define \mathcal{B}_n as the set of valid vectors \mathbf{b} of size n . Also let $\mathcal{B} = \bigcup_{n=1}^{\infty} \mathcal{B}_n$.

Each consumer is characterized by the specification of his *most preferred good* in the preference spectrum, defined as the good that represents the optimal transfer of characteristics to him. A consumer i with most preferred good $x_i \in \mathcal{L}$ associates a utility U_{ij} with a product j in the assortment as given by:

$$U_{ij} = Z - r - g(|x_i - b_j|).$$

Here, Z is a positive constant representing the surplus associated with the product category, r denotes the price of the product, and $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a strictly increasing function representing the disutility associated with the distance between the consumer's most preferred good and the product's specification. Without loss of generality, we assume that the utility of not buying anything in the store is equal to zero for every consumer, i.e., $U_{i0} = 0$. The consumer selects the product that maximizes his utility, provided that it is nonnegative, otherwise he does not make a choice.

We define the *coverage distance* of any product with specification b as the maximum distance between the product's specification and the most preferred good of a consumer for whom this product gives a nonnegative utility. The coverage distance is identical for all products and denoted

as L . We have

$$L = \max_i \{|x_i - b| : Z - r - g(|x_i - b|) \geq 0\} \quad (1)$$

Thus, we define the *coverage interval* of product j as the interval $[b_j - L, b_j + L]$. This interval contains the most preferred goods of all customers who obtain a nonnegative utility from product j .

Product j is said to be the *first choice* of a given consumer if it gives a nonnegative utility to that consumer and its distance from the most preferred good of that consumer is the minimum among all products in the assortment. Thus, the *first choice interval* of product j in assortment \mathbf{b} is defined as the subinterval of \mathcal{L} that contains the most preferred goods of all consumers who choose product j as a first choice. The first choice interval is denoted as $[b_j^-, b_j^+]$, and is given by

$$\begin{aligned} b_j^- &= \max \left\{ b_j - L, \frac{b_j + b_{j-1}}{2} \right\}, \\ b_j^+ &= \min \left\{ b_j + L, \frac{b_j + b_{j+1}}{2} \right\} \end{aligned} \quad (2)$$

for $j = 1, \dots, n$ and with the convention that $b_0 = -\infty$ and $b_{n+1} = +\infty$. Further, let $l_j = b_j^+ - b_j^-$ denote the length of the first choice interval of product j .

We note that while there can be overlap between the coverage intervals of products, there is no overlap between the first choice intervals. We also note that L measures the degree of *substitutability* of the product category for any given assortment. A small value of L implies that consumers are willing to substitute in only a small region around their most preferred goods, while a large value of L implies that consumers are more tolerant of differences between products and their most preferred goods.

Additional Assumptions

We make the following additional assumptions to extend Lancaster's model to stochastic demand and random distribution of consumers on the preference spectrum. We assume that the specifications of the most preferred goods of all consumers are independent and identically distributed with continuous probability distribution F and density function f on the finite support $[0, 1] \subset \mathcal{L}$. The *first choice probability* of product j , denoted p_j , is defined as:

$$p_j = \int_{b_j^-}^{b_j^+} f(x) dx = F(b_j^+) - F(b_j^-). \quad (3)$$

Let \mathbf{p} denote the vector of first choice probabilities. We refer to $[0, 1]$ as the *market* and $\sum_{j=1}^n p_j$ as the *market coverage* of the retailer.

We assume that f is either the uniform density function or unimodal. f is said to be unimodal if the mode $x^* \equiv \arg \max\{f(x) : x \in [0, 1]\}$ exists and is unique, and $f(x)$ is non-decreasing for all $x < x^*$ and non-increasing otherwise. This assumption enables us to obtain insights into the effect of heterogeneity of consumer preferences on the structure of the optimal solution. It is also practically useful since it fits the demand distributions for many product categories.

If f is unimodal, this implies that there exists a unique *most popular product* on the spectrum, and the more distant a product is from the most popular product, the less popular it is. We define the most popular product as the product with specification x such that $F(x + L) - F(x - L)$ is maximized. This corresponds to the product with that would have the highest expected demand if it was the only product carried in the assortment. Note that the specification of the most popular product may not be equal to x^* .

Finally we assume that customers arrive to the store over a finite period of time according to a Poisson process with rate λ . Let N denote the total number of customers visiting the retailer, and N_j denote the number of customers who choose product j as their first choice. N is Poisson with rate λ and N_j is Poisson with rate λp_j .

4 Static Substitution

Under static substitution, customers do not substitute if the store has run out of inventory of their first choice product by the time of their arrival. Therefore, the demand for product j , denoted D_j , is simply the number of customers that pick product j as their first choice, that is N_j , and is independent of inventory levels.

The retailer's decision problem is to determine the number of products in the assortment, n , the vector of product specifications, \mathbf{b} , and the vector of order quantities, $\mathbf{q} = (q_1, \dots, q_n)$, in order to maximize the total expected profit. Let $\pi^s(p_j, q_j)$ denote the profit under static substitution for product j . The expected profit under static substitution for product j is written as:

$$E\pi^s(p_j, q_j) = rE[\min\{q_j, D_j(p_j)\}] - cq_j + vE[q_j - D_j(p_j)]^+, \quad (4)$$

and the total expected profit is given by

$$E\Pi^s(\mathbf{b}, \mathbf{q}) = \sum_{j=1}^n E\pi^s(p_j(\mathbf{b}), q_j) - nK.$$

Given \mathbf{b} , this problem is separable into n subproblems, one for each product. Let $E\pi^s(p_j) =$

$\max_{q_j} E\pi^s(p_j, q_j)$ denote the expected profit for product j as a function of p_j when q_j is chosen optimally. Let $E\Pi^s(\mathbf{b}) = \sum_j E\pi^s(p_j)$ be the corresponding total expected profit.

In this section we use the Normal approximation for the Poisson distribution to represent the demand for each product, i.e., we assume that the demand for product j is normally distributed with mean λp_j and standard deviation $\sqrt{\lambda p_j}$. The advantage of the normal distribution is that it gives a closed form expression for $E\pi^s(p_j)$ (see Porteus (2002: p.13) or Zipkin (2000: p.216)):

$$E\pi^s(p_j) = (r - c)p_j\lambda - (r - v)\sqrt{p_j\lambda}\phi(z) \quad (5)$$

where $z = \Phi^{-1}(\theta)$ is the service level, $\theta = \frac{r-c}{r-v}$ is the critical fractile, and $\phi(\cdot)$ and $\Phi(\cdot)$, respectively, denote the density function and the cumulative distribution function of the standard normal distribution.

Any product j is profitable if its first choice probability is such that $E\pi^s(p_j) > K$. Let

$$\underline{p} = \max\{p : E\pi^s(p) \leq K\}. \quad (6)$$

The value of \underline{p} gives the minimum first choice probability at which a product is profitable. It is to be noted that under the normal approximation, $E\pi^s(p)$ is negative for p close to zero. Thus, \underline{p} is strictly positive under this approximation even if $K = 0$. Also let \bar{p} denote the maximum first choice probability for any product. We have

$$\bar{p} = \max_{x \in [0,1]} \{F(x+L) - F(x-L)\}. \quad (7)$$

We assume that $\bar{p} \geq \underline{p}$. Otherwise, the problem has a trivial solution wherein it is optimal not to stock any item.

The following lemma is useful in the sequel.

Lemma 1. *$E\pi^s(p)$ is a convex increasing function of p for all $p \geq \underline{p}$.*

Proof. Omitted.

In other words, the retailer can benefit from economies of scale in each product because her expected profit increases more than proportionately with mean demand. We note that if demand is represented by a Poisson distribution then the resulting expression for $E\pi^s(p)$ is not convex in p since the inventory levels are drawn from a discrete set. Thus, the normal approximation is useful because it convexifies the expected profit function. Further, we evaluate expected profit in our

numerical analysis using both the Poisson distribution and the normal approximation and find the normal approximation to be quite accurate.²

The optimization problem under static substitution is thus formulated as follows:

$$\begin{aligned} \mathbb{P}^s : \quad & \max_{\substack{\mathbf{b} \in \mathcal{B} \\ n = \text{length}(\mathbf{b}) \geq 0}} & E\Pi^s(\mathbf{b}) = \sum_{j=1}^n \left[(r-c)p_j\lambda - (r-v)\sqrt{p_j}\lambda\phi(z) \right] - nK \\ & \text{s.t.} \quad \begin{cases} p_j = F(b_j^+) - F(b_j^-) & j = 1, \dots, n \\ b_j^- = \max \left\{ b_j - L, \frac{b_j + b_{j-1}}{2} \right\} & j = 1, \dots, n \\ b_j^+ = \min \left\{ b_j + L, \frac{b_j + b_{j+1}}{2} \right\} & j = 1, \dots, n \end{cases} \end{aligned}$$

4.1 Structure of the optimal policy

Problem \mathbb{P}^s is difficult to solve since the objective function as well as the first constraint are non-linear. If we attempt to maximize the objective function with respect to \mathbf{p} and then compute the corresponding value of \mathbf{b} , then the objective function becomes separable across products, but the vector \mathbf{p} need not correspond to any feasible value of \mathbf{b} . On the other hand, if we optimize directly with respect to \mathbf{b} , then the expected profit is not separable into the sum of the expected profit functions for each product as a function of their specification. Instead, changing the specification, b_j , for any one product j affects not only p_j , but possibly also p_{j-1} and p_{j+1} .

We simplify problem \mathbb{P}^s by establishing a set of properties that the optimal assortment(s) should satisfy. Thus, we are able to reduce the problem into the maximization of expected profit with respect to one variable over a bounded region.

Let

$$\begin{aligned} \alpha &= \min \left\{ x \in \mathcal{L} : F(x+L) - F(x-L) \geq \underline{p} \right\}, \\ \beta &= \max \left\{ x \in \mathcal{L} : F(x+L) - F(x-L) \geq \underline{p} \right\}. \end{aligned} \tag{8}$$

When $F(\cdot)$ is uniform or unimodal, $[\alpha, \beta]$ defines the line segment where it is profitable to locate a product. It follows that the interval $[\max(0, \alpha - L), \min(1, \beta + L)]$ gives the part of the preference spectrum containing the most preferred goods of all customers that can be served profitably. We call this interval as the *profitable region*. Note that it is possible to have $\alpha < 0$ or $\beta > 1$, i.e., there can be products that are outside the $[0, 1]$ finite support of the preference spectrum in the retailer's profitable region.

We find that the optimal solution satisfies the following properties.

²The results of §4 hold not only for normal distribution but also for any other distribution of demand such that $E\pi^s(p)$ is a convex increasing function of p . We use the normal distribution for numerical analysis.

Proposition 1. *If \mathbf{b} is an optimal solution to \mathbb{P}^s , then*

(i) $b_j \in [\alpha, \beta]$ for $j = 1, \dots, n$.

(ii) $b_j^+ = b_{j+1}^-$ for $j = 1, \dots, n-1$.

(iii) If x^* is uniquely defined, then $b_1^- \leq x^* \leq b_n^+$.

(iv) $b_1 < \alpha + 2L$ and $b_n > \beta - 2L$.

(v) $b_j^+ - b_j^- = 2L$ for $j = 1, \dots, n$.

Property (i) follows directly from the definitions of α and β . Property (ii) states that the first choice intervals of all the products must be adjacent so that the portion of the preference spectrum that is served by the retailer constitutes a compact interval. By property (iii), this interval must contain the mode of the distribution if F is unimodal. By property (iv), the first product should be distant from α by strictly less than $2L$, since otherwise the expected profit of the assortment is less than or equal to that of a similar assortment that contains one additional product placed at $b_1 - 2L$. Similarly, the last product should be distant from β by strictly less than $2L$. Finally, property (v) states that products should have first choice intervals of length equal to $2L$. Thus, products are spaced exactly by $2L$, their coverage intervals do not overlap, and each customer receives positive utility from at most one product.

Property (v) is the surprising result of Proposition 1 because it shows that the lengths of the first choice intervals of products are independent of the distribution of consumer preferences on the preference spectrum. The property of equally spaced product specifications has been obtained previously by Lancaster (1979) and de Groote (1994) for deterministic demand with uniform customer preferences. Proposition 1 shows that this property extends to the case when demand is random, customer preferences have a non-uniform density and prices are exogenous.

Using Proposition 1, with a slight abuse of notation, we define the interval containing b_1 as $[\alpha, \min(\alpha + 2L, \beta))$. Now, if we fix b_1 in this interval, Proposition 1 determines the values of b_2, \dots, b_n as functions of b_1 : $b_j = b_1 + 2L(j-1)$ for $j = 2, \dots, n$. Further, the optimal number of products in the assortment, n , is given by the following corollary.

Corollary 1. *Given b_1 , the optimal number of products in the assortment is $n(b_1) = \left\lfloor \frac{\beta - b_1}{2L} \right\rfloor + 1$. Further, when b_1 is chosen optimally, the optimal number of products takes one of two values, $\left\lfloor \frac{\beta - \alpha}{2L} \right\rfloor$ or $\left\lfloor \frac{\beta - \alpha}{2L} \right\rfloor + 1$.*

Proof. Omitted. □

Thus, we have reduced \mathbb{P}^s into a maximization problem \mathbb{P}_l^s in one variable, b_1 , over a bounded region:

$$\mathbb{P}_l^s : \quad \max_{b_1 \in [\alpha, \min(\alpha + 2L, \beta))} \quad E\Pi^s(\mathbf{b}) = \sum_{j=1}^{n(b_1)} \left[(r-c)p_j\lambda - (r-v)\sqrt{p_j\lambda}\phi(z) \right] - nK$$

$$\text{s.t.} \quad \begin{cases} p_j & = & F(b_j^+) - F(b_j^-) & \text{for } j = 1, \dots, n(b_1) \\ b_j^- & = & b_1 - L + 2L(j-1) & \text{for } j = 1, \dots, n(b_1) \\ b_j^+ & = & b_1 - L + 2Lj & \text{for } j = 1, \dots, n(b_1) \\ n(b_1) & = & \left\lfloor \frac{\beta - b_1}{2L} \right\rfloor + 1 \end{cases}$$

The optimal solution to \mathbb{P}^s can now be obtained by solving \mathbb{P}_l^s numerically by doing a line search in the range $[\alpha, \min(\alpha + 2L, \beta))$. This range is subdivided into two subintervals $\left[\alpha, \alpha + h\left(\frac{\beta - \alpha}{2L}\right)2L\right]$ and $\left[h\left(\frac{\beta - \alpha}{2L}\right)2L, \beta\right]$ where $h(\cdot)$ denotes the fractional part of any real number, the difference being that the number of products induced by b_1 in the first interval is one more than in the second interval. We note that the objective function of \mathbb{P}_l^s may admit more than one local maximum over each subinterval. Figure 1 illustrates this with an example. Thus, even though the assortment problem has been considerably simplified, we require a line search in both subintervals to obtain the optimal solution. In the special case when the distribution of preferences is symmetric around 0.5, it is sufficient to search for the maximum in one-half of each of the two subintervals.

When preferences are distributed according to a uniform distribution on $[0, 1]$, then it can further be shown that $b_1 = L$ is optimal. The optimal values of n and b_2, \dots, b_n are then obtained from Proposition 1 and Corollary 1.

4.2 Numerical examples

We identify the tradeoffs incorporated in the optimal assortment using a numerical example. The distribution of consumer preferences is represented by a Beta distribution with parameters γ_1 and γ_2 on $[0, 1]$.³ We let $\gamma_1 = \gamma_2 = \gamma$ so that the distribution is symmetric around its mode 0.5. The values of γ measure the degree of heterogeneity of the consumer population.⁴ The larger

³The density function for a Beta distribution with parameters γ_1 and γ_2 is $\frac{1}{B(\gamma_1, \gamma_2)}x^{\gamma_1-1}(1-x)^{\gamma_2-1}$ where $B(\gamma_1, \gamma_2) = \int_0^1 x^{\gamma_1-1}(1-x)^{\gamma_2-1}dx$.

⁴More generally, the variance of the Beta distribution, $\frac{\gamma_1\gamma_2}{(\gamma_1+\gamma_2)^2(\gamma_1+\gamma_2+1)}$, can be used to measure the degree of heterogeneity of the consumer population. When $\gamma_1 = \gamma_2$, this formula reduces to $\frac{1}{4(2\gamma+1)}$. Therefore, for symmetric Beta distributions, we use γ to measure the degree of heterogeneity of the consumer population. All our results also

	$\mathbf{b} = (0.4, 0.6)$	$\mathbf{b} = (0.3, 0.5, 0.7)$
$\gamma = 5$	$\mathbf{p} = (0.401, 0.401)$ $(E\Pi_j) = (89.64, 89.64)$ Total expected profit = 139.28	$\mathbf{p} = (0.247, 0.467, 0.247)$ $(E\Pi_j) = (53.38, 105.22, 53.38)$ Total expected profit = 151.98
$\gamma = 10$	$\mathbf{p} = (0.467, 0.467)$ $(E\Pi_j) = (105.35, 105.35)$ Total expected profit = 170.71	$\mathbf{p} = (0.1845, 0.6278, 0.1845)$ $(E\Pi_j) = (38.90, 143.62, 38.90)$ Total expected profit = 161.42

Table 1: Results for Example 1, Case 3

the value of γ , the more concentrated the distribution of consumer preferences is around 0.5, and thus, the more homogeneous the consumer population is. We use different values of K, L and γ to illustrate the tradeoffs represented in the optimal assortment. Recall that L measures the degree of substitutability of the product category.

Example 1. Suppose $\lambda = 50, r = 10, c = 5$ and $v = 3$.

Case 1: Let $K = 50, L = 0.1$ and $\gamma = 2$. We find that the minimum profitable first choice probability \underline{p} is 0.23 which gives $\alpha = 0.27$ and $\beta = 0.73$. Thus, the profitable region, given by $[\max(0, \alpha - L), \min(1, \beta + L)]$, is equal to $[0.17, 0.83]$ and corresponds to 84.6% of the market. The optimal assortment consists of two products with specifications $\mathbf{b} = (0.4, 0.6)$, first choice probabilities $\mathbf{p} = (0.284, 0.284)$, and market coverage 56.8%. The maximum expected profit is 24.07.

Case 2: Let $K = 0, L = 0.2$ and $\gamma = 2$. The profitable region is $[0, 1]$ and optimal assortments are $(0.21, 0.61, 1.01)$ and by symmetry its mirror image $(-0.01, 0.39, 0.79)$ for a total market coverage of 99.98%.

Case 3: Let $K = 20$ and $L = 0.1$. Consider the solutions for $\gamma = 5$ and $\gamma = 10$ as shown in Table 1. The optimal assortment is $\mathbf{b} = (0.3, 0.5, 0.7)$ for $\gamma = 5$, and $\mathbf{b} = (0.4, 0.6)$ for $\gamma = 10$. The table compares the expected profits from these assortments under both values of γ . \square

All cases illustrate property (v) in Proposition 1, i.e., the products offered in the optimal assortment are separated by $2L$ regardless of the scale of demand or the distribution of consumers on the preference spectrum. It does not happen that the retailer provides more variety in parts of the spectrum with higher density of consumers than the other parts. The $2L$ separation also apply to the case when $\gamma_1 \neq \gamma_2$.

implies that the coverage intervals of products do not overlap. It follows that consumers get a positive utility from at most one product in the assortment.

Case 1 further illustrates that even though the distribution of preferences has finite support, it may not be profitable to satisfy all customers; the profitable region covers only 84.6% of the market in Case 1. This result in our model comes strictly from positive fixed costs. When fixed costs are zero, it is possible to capture the entire market by stocking a finite number of products whose demands are independent of one another. In general, the size of the profitable region $[\max(0, \alpha - L), \min(1, \beta + L)]$ is a decreasing function of K and γ , and an increasing function of L and λ .

The optimal assortment in Case 1 covers only 56.8% of the market, indicating that the retailer chooses not to satisfy all the profitable customers in the market. This outcome is due to the tradeoff of market coverage with fixed costs and economies of scale in the demand for each product offered. If the retailer were to add a new product to increase market coverage while respecting the spacing of product specifications by $2L$, the retailer would have to change the locations of existing products (in order that all product specifications are in the range $[\alpha, \beta]$). This would result in a different allocation of demand that would potentially reduce economies of scale. This effect, along with the increase in fixed cost would then cause a drop in expected profit.

The size distribution of clothing in apparel stores illustrates the above results in Case 1. Many retailers do not stock clothing articles in very large or very small sizes because there is not enough demand in a fringe size to justify the fixed cost of offering that size. Thus, the preferences of customers at the two extremes of the size distribution may not be covered by a retailer. We note that the MNL model also yields the result that the optimal assortment does not cover the entire market (van Ryzin and Mahajan 1999). However, the MNL model uses a different substitution structure, and this result in the MNL model comes from economies of scale regardless of fixed costs because adding a product always decreases the demand for each preexisting product in the assortment.

Case 2 shows that when $K = 0$ the optimal market coverage is virtually 100%.⁵ It also illustrates the fact that even though the Beta distribution is symmetric around 0.5, the optimal assortment is not necessarily symmetric around 0.5. The symmetric assortment of three products is given by $[0.1, 0.5, 0.9]$, and yields lower expected profit than the optimal assortment $[0.21, 0.61, 1.01]$. The intuition for this result is as follows. The symmetric assortment has a higher fraction of demand

⁵The reason why the market coverage is not exactly equal to 100% is that the Normal approximation of demand forces $\underline{p} > 0$ when $K = 0$. However, \underline{p} is very close to zero so that $\alpha \approx -L$ and $\beta \approx 1 + L$.

served by the second and third products, but a lower fraction of demand served by the first product. Thus, the higher economies of scale for the second and third products in the symmetric assortment are more than offset by the diseconomies of scale for the first product, so that the optimal assortment is not symmetric around 0.5. The optimality of an asymmetric assortment also implies it is not possible to simplify the optimization problem by restricting the set of possible product locations to a (discrete) subset of $[0, 1]$. Finally, the optimal assortment in Case 2 contains a product that is located outside the interval $[0, 1]$. This is so because locating the product in the interval $[0, 1]$ would reduce the demand for its neighboring product, thus imposing a higher inventory cost on the retailer.

Case 3 shows that it can be optimal not to stock the most popular product possible, i.e., the product with specification b_j such that $F(b_j + L) - F(b_j - L)$ is maximized. This product has specification $b = 0.5$ in this example. We find that for $\gamma = 5$, the assortment $[0.4, 0.6]$ is optimal, while for $\gamma = 10$, the assortment $[0.3, 0.5, 0.7]$ is optimal. Stocking the most popular product is not optimal for $\gamma = 5$ because it fragments the demand for the remaining products in the assortment in such a way that the economies of scale enjoyed by the most popular product are insufficient to overcome the diseconomies of scale suffered by the remaining products.⁶ This finding is consistent with the decision of some food retailers to stock flavors of ice cream containing vanilla (such as vanilla chocolate chips, vanilla with cookies or cherry vanilla) but not to stock the plain vanilla flavor, which is often the most popular. Another example is that of blueberry muffins. Retailers sometimes choose not to stock this most popular flavor or to carry less inventory of this flavor because its absence might result in a more profitable split of demand among other types of muffins. Under the MNL model, the most popular product is always included in the optimal assortment.

We note that some of the above insights change when we consider dynamic substitution. In particular, it may be optimal to allow overlap between products by spacing them by less than $2L$. It may also be optimal to provide different levels of variety in different parts of the market depending on the distribution of customers. Further, the products in the optimal assortment will always be located on the interval $[0, 1]$. We discuss these results in §5.

⁶As suggested by a referee, stocking the most popular product is optimal in our model if preferences are monotonically increasing or decreasing over $[0, 1]$. The most popular product is located at $b = L$ for decreasing preferences and at $b = 1 - L$ for increasing preferences.

5 Dynamic Substitution

Under dynamic substitution, a customer chooses the product that gives him the highest positive utility among those that are available in stock at the time when he visits the store. If all available products give him a negative utility, then he leaves the store empty-handed. It follows that the customer may not be able to buy his first choice product, i.e., the product in the original assortment that is the closest to his most preferred good. Instead, all products located within a distance L of the customer's most preferred good are possible purchases.

Under dynamic substitution, the total demand for any product j from customers who by first choice or substitution want to buy product j cannot be expressed analytically. The demand function depends on the stocking quantities of all products as well as the sequence of arrivals of customers and requires some accounting assumptions (for example, whether or not to double-count a substitution purchase as a demand for the first choice and the final choice). To circumvent this problem, we define S_j as a random variable denoting the sales of product j . Thus, the expected profit under dynamic substitution for a given assortment \mathbf{b} and inventory vector \mathbf{q} can be written as:

$$\begin{aligned} \mathbb{P}^d : \quad \max_{\substack{\mathbf{b} \in \mathcal{B} \\ \mathbf{q} \geq 0 \\ n = \text{length}(\mathbf{b}) \geq 0}} E\Pi^d(\mathbf{b}, \mathbf{q}) &= \sum_{j=1}^n rE[S_j(\mathbf{b}, \mathbf{q})] - cq_j + vE[q_j - S_j(\mathbf{b}, \mathbf{q})] - nK \\ &= \sum_{j=1}^n (r - v)E[S_j(\mathbf{b}, \mathbf{q})] + (c - v)q_j - nK. \end{aligned}$$

The value of $E[S_j(\mathbf{b}, \mathbf{q})]$ also cannot be written in closed form and is difficult to estimate since it requires simulating various possible sequences of customer arrivals. However, we define a lower bound and an upper bound on $E[S_j(\mathbf{b}, \mathbf{q})]$ for given \mathbf{q} and \mathbf{b} . Since $E\Pi^d(\mathbf{b}, \mathbf{q})$ is an increasing function of total expected sales, $\sum_{j=1}^S E[S_j(\mathbf{q}, \mathbf{b})]$, we thus obtain bounds on $E\Pi^d(\mathbf{b}, \mathbf{q})$. These bounds are presented in §5.1 and §5.2. In §5.3, we compare the bounds, use them to obtain two heuristics for solving \mathbb{P}^d , and conduct numerical analysis to derive insights into the structure of the optimal solution.

5.1 Lower bound on profit

Let ω denote a sample path of customer arrivals, that is, a realization of the sequence $X(\omega) = \{x_1, \dots, x_N\}$ where x_i is the most preferred good of the i th arriving consumer and N is the total number of customers visiting the store during the period. Also let N_j be the number of customers whose first choice is product j .

To construct a lower bound, we restrict the consumers to purchase only their first choice products, if available. This yields the static substitution problem of §4. Note that if customers are not permitted to substitute in the event of a stockout of their first choice product, it will result in fewer sales, more inventory to salvage, and a loss of $r - v$ on each unsold unit. Therefore, we obtain a lower bound on the sales as well as the profit under dynamic substitution for all values of \mathbf{b} and \mathbf{q} and *for every sample path of customer arrivals*.

Proposition 2. $\Pi^s(\mathbf{b}, \mathbf{q}, \omega) \leq \Pi^d(\mathbf{b}, \mathbf{q}, \omega)$ for all $\omega, \mathbf{b}, \mathbf{q}$ where

$$\Pi^s(\mathbf{b}, \mathbf{q}, \omega) = \sum_{j=1}^n \{r \min\{q_j, N_j\} - cq_j + v[q_j - N_j]^+ - nK\}$$

5.2 Upper bound on profit

Consider the relaxation of the dynamic substitution problem wherein the retailer first observes the locations of the most preferred goods of all arriving customers $X(\omega) = \{x_1, \dots, x_N\}$, and then allocates inventory to all customers in order to maximize profit. Thus, the retailer may assign to a customer a product that is not his first choice but gives him positive utility. This relaxed problem gives an upper bound on $\Pi^d(\mathbf{b}, \mathbf{q}, \omega)$. Let \mathbb{P}^u denote the relaxed problem, and $\Pi^u(\mathbf{b}, \mathbf{q}, \omega)$ denote the profit under the relaxed problem given \mathbf{b}, \mathbf{q} and sequence X of customer arrivals.

Problem \mathbb{P}^u merits comparison with demand substitution models studied previously in the literature. Lippman and McCardle (1997), Netessine and Rudi (2003), and Bassok, Anupindi and Akella (1999) consider models wherein the retailer observes the entire demand before allocating any inventory. Therefore the customers do not directly choose a product. Rather, they are assigned a product either according to exogenous rules or by a decision of the retailer. Agrawal et al. (2002) use similar assumptions to analyze dynamic pricing of capacity. The paper of Bassok, Anupindi and Akella is closely related to our paper since they consider a setting where the allocation of inventory is mediated by the retailer. While they consider a downward substitution rule, we consider a neighborhood substitution rule implied by the locational model of consumer choice. Further, in our model, the choice of \mathbf{b} enables the retailer to control the rate of substitution. For these reasons, we call problem \mathbb{P}^u as the retailer-controlled substitution model. In contrast to the above-referenced papers, Smith and Agrawal (2000), Mahajan and van Ryzin (2001) and Kok and Fisher (2004) consider dynamic substitution where the incidence of substitution depends on the sample path of customer arrivals.

We first show how to obtain the optimal allocation of inventory to the customers who visit the store for a given sample path ω and given vectors \mathbf{b} and \mathbf{q} . For this, we partition customers based on the number and indices of products that give them a positive utility. We say that a customer is of *type* (l, i) if there are l products located within distance L of his most preferred good and i is the product with the lowest index among these. Thus, l measures the customer's level of flexibility with $l = 1$ denoting the *inflexible* customers and $l = n$ denoting the *perfectly flexible* customers. A customer of type (l, i) can only be served using inventory of products $i, i + 1, \dots, i + l - 1$.

Let $\eta_{(l,i)}$ denote the number of customers of type (l, i) in a given sequence X and let $x_{j,(l,i)}$ be the amount of inventory of product j allocated to customers of type (l, i) . Given η, \mathbf{b} and \mathbf{q} , the problem to find the optimal allocation can be formulated as

$$\Pi^u(\mathbf{b}, \mathbf{q}, \omega) = \max r \sum_{j=1}^n \sum_{l=1}^n \sum_{i=1}^{n-l+1} x_{j,(l,i)} - c \sum_{j=1}^n q_j + v \sum_{j=1}^n \left[q_j - \sum_{l=1}^n \sum_{i=1}^{n-l+1} x_{j,(l,i)} \right] - nK \quad (9)$$

such that

$$\sum_{j=1}^n x_{j,(l,i)} \leq \eta_{(l,i)} \quad \text{for } l = 1, \dots, n; \quad i = 1, \dots, n - l + 1 \quad (10)$$

$$\sum_{l=1}^n \sum_{i=1}^{n-l+1} x_{j,(l,i)} \leq q_j \quad \text{for } j = 1, \dots, n \quad (11)$$

$$x_{j,(l,i)} \geq 0 \quad \text{for } j = 1, \dots, n; \quad l = 1, \dots, n; \quad i = 1, \dots, n - l + 1 \quad (12)$$

$$x_{j,(l,i)} = 0 \quad \text{for } j < i \text{ or } j > i + l - 1. \quad (13)$$

Here, the objective function is equivalent to maximizing the sum of allocations

$\sum_{j=1}^n \sum_{l=1}^n \sum_{i=1}^{n-l+1} x_{j,(l,i)}$. Therefore, this problem is equivalent to a restricted transportation problem. We find that the problem admits a greedy optimal solution as proved by the following lemma.

Lemma 2. *An optimal allocation policy can be obtained using the following algorithm:*

Algorithm

For $(j = 1 : n; \quad l = 1 : (n - j + 1); \quad k = 0 : (j - 1))$ do

$$x_{j,(l+k,j-k)} = \min \{ \eta_{(l+k,j-k)}, q_j \};$$

$$\eta_{(l+k,j-k)} = \eta_{(l+k,j-k)} - x_{j,(l+k,j-k)};$$

$$q_j = q_j - x_{j,(l+k,j-k)};$$

Endfor.

Here, j indexes products and l and k correspond to levels of customer flexibility. The intuition is that products should be allocated to customers in increasing order of index j and for a given value of j , in increasing levels of *remaining flexibility*. We define remaining flexibility as the number of products not yet allocated that can satisfy the customers. For example, customers of type $(3, 1)$ have a remaining flexibility of 2 after product 1 has been allocated, then of 1 after product 2 has also been allocated. Thus, customers of type $(3, 1)$ will be served from the inventory of product 3 before customers of type $(2, 3)$ even though the latter are more flexible *a priori*. This is because after allocating products 1 and 2, customers of type $(3, 1)$ can be served only by product 3, while customers of type $(2, 3)$ can be served from the inventory of product 4 as well. In the algorithm, l indexes remaining flexibility for a given j , and k indexes all customer types with a given amount of remaining flexibility.

Lemma 2 enables us to obtain the following upper bound on the dynamic substitution problem.

Proposition 3. $\Pi^u(\mathbf{b}, \mathbf{q}, \omega) \geq \Pi^d(\mathbf{b}, \mathbf{q}, \omega)$ for all $\mathbf{b}, \mathbf{q}, \omega$ where

$$\Pi^u(\mathbf{b}, \mathbf{q}, \omega) = r \sum_{l=1}^n \sum_{i=1}^{n-l+1} \sum_{j=1}^n x_{j,(l,i)}^* - c \sum_{j=1}^n q_j + v \sum_{j=1}^n \left[q_j - \sum_{l=1}^n \sum_{i=1}^{n-l+1} x_{j,(l,i)}^* \right] - nK,$$

and $x_{j,(l,i)}^*$ are the optimal allocations given by Lemma 2.

Proof. Omitted. □

Taking the expectation of $\Pi^u(\mathbf{b}, \mathbf{q}, \omega)$, the optimization problem \mathbb{P}^u is formulated as:

$$\mathbb{P}^u : \max \{ E\Pi^u(\mathbf{b}, \mathbf{q}) : \mathbf{b} \in \mathcal{B}, \mathbf{q} \geq 0, n = \text{length}(\mathbf{b}) \geq 0 \}.$$

For a given assortment \mathbf{b} , the maximization of $E\Pi^u(\mathbf{b}, \mathbf{q}, \omega)$ with respect to \mathbf{q} can now be represented by a two-stage stochastic linear program where the choice of the inventory vector is the first stage and the allocation of inventory to customers is the second stage. The variables $\eta_{(l,i)}$ are stochastic with probability distributions determined by \mathbf{b} . We have the following theorem.

Theorem 1. $E\Pi^u(\mathbf{b}, \mathbf{q})$ is concave in \mathbf{q} .

Thus, the first order conditions with respect to \mathbf{q} are both necessary and sufficient to obtain the optimal inventory levels \mathbf{q}^* . Since the first order conditions involve convolutions of probability distributions, Proposition 4 below facilitates the computation of the optimal solution by obtaining bounds on the optimal values of q_j .

Proposition 4. Let $\theta = \frac{r-c}{r-v}$ be the critical fractile. Then,

$$q_j^* \geq \min \{q_j : P[\eta_{(1,j)} \leq q_j] \geq \theta\} \quad j = 1, \dots, n,$$

$$q_j^* \leq \min \left\{ q_j : P \left[\eta_{(1,j)} + \sum_{l=2}^n \sum_{i=(j-l+1) \wedge 1}^j \eta_{(l,i)} \leq q_j \right] \geq \theta \right\} \quad j = 1, \dots, n.$$

Proof. Omitted. □

We now turn to the problem of finding the optimal value of \mathbf{b} . We find that this value satisfies the following properties, to be compared with Proposition 1 in the case of static substitution.

Proposition 5. If \mathbf{b} is an optimal solution to \mathbb{P}^u , then

- (i) $b_j \in [\alpha, \beta]$ for $j = 1, \dots, n$.
- (ii) $b_j^+ = b_{j+1}^-$ for $j = 1, \dots, n-1$.
- (iii) If x^* is uniquely defined, then $b_1^- \leq x^* \leq b_n^+$.
- (iv) $b_1 < \alpha + 2L$ and $b_n > \beta - 2L$.
- (v) $b_1 \geq L$ and $b_n \leq 1 - L$.

Properties (i)-(iv) also hold under static substitution (see Proposition 1), while (v) may not. Property (v) states that the specification as well as the entire coverage interval of each product should be included within the $[0, 1]$ interval. Note that this result differs from that obtained for static substitution. Further, note that we do not have an equivalent of Proposition 1(v) for \mathbb{P}^u , i.e., the coverage intervals of products at the optimal solution need not be of length $2L$. Properties (i)-(v) enable us to restrict the set of possible values of \mathbf{b} to be considered for finding the optimal assortment. Thus, the optimal assortment can be obtained by numerical optimization.

5.3 Comparison of the bounds and insights

Let $(\mathbf{b}^{*s}, \mathbf{q}^{*s})$ denote the optimal solution to \mathbb{P}^s , $(\mathbf{b}^{*d}, \mathbf{q}^{*d})$ denote the optimal solution to \mathbb{P}^d , and $(\mathbf{b}^{*u}, \mathbf{q}^{*u})$ denote the optimal solution to \mathbb{P}^u . The following proposition gives bounds on the expected profit that a retailer can achieve under dynamic substitution.

Proposition 6. (i) $E\Pi^s(\mathbf{b}, \mathbf{q}) \leq E\Pi^d(\mathbf{b}, \mathbf{q}) \leq E\Pi^u(\mathbf{b}, \mathbf{q})$ for all \mathbf{b}, \mathbf{q} .

(ii) $E\Pi^s(\mathbf{b}^{*s}, \mathbf{q}^{*s}) \leq E\Pi^d(\mathbf{b}^{*d}, \mathbf{q}^{*d}) \leq E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})$.

Proof. Omitted. □

Thus, our analysis yields two heuristics for solving \mathbb{P}^d . The first heuristic, denoted H1, consists in offering the optimal solution for static substitution, $(\mathbf{b}^{*s}, \mathbf{q}^{*s})$, to customers who substitute dynamically. The associated expected profit, $E\Pi^d(\mathbf{b}^{*s}, \mathbf{q}^{*s})$, is identical to $E\Pi^s(\mathbf{b}^{*s}, \mathbf{q}^{*s})$ because the optimal assortment under static substitution provides no overlap between the coverage intervals of products. The optimality gap of this heuristic can be estimated by comparing $E\Pi^s(\mathbf{b}^{*s}, \mathbf{q}^{*s})$ with the upper bound $E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})$.

The second heuristic, denoted H2, consists in offering the optimal solution for the retailer-controlled substitution, $(\mathbf{b}^{*u}, \mathbf{q}^{*u})$, to customers who substitute dynamically. The expected profit given by this heuristic is $E\Pi^d(\mathbf{b}^{*u}, \mathbf{q}^{*u})$, which can again be compared with the upper bound $E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})$.

We conducted a numerical study to benchmark the performances of the heuristics using 3150 problem instances. In each case, we set $r = 10$ and $v = 3$, and use a Beta distribution symmetric around 0.5 (with parameter γ) to represent the probability distribution of consumers' most preferred goods on the attribute space. The remaining parameters are varied as follows:

$$\begin{aligned} \lambda &: \{20, 50, 100\}, \\ \gamma &: \{1, \dots, 10\}, \\ L &: \{0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}, \\ K &: \{0.1, 1, 5, 10, 20\} \\ c &: \{4, 5, 6\} \end{aligned}$$

To obtain the solutions in each problem instance, we first solve \mathbb{P}^s and \mathbb{P}^u as described in §4.2 and §5.2, respectively. Since $E\Pi^d(\mathbf{b}^{*u}, \mathbf{q}^{*u})$ cannot be computed analytically, we then estimated all three expected profit values, $E\Pi^s(\mathbf{b}^{*s}, \mathbf{q}^{*s})$, $E\Pi^d(\mathbf{b}^{*u}, \mathbf{q}^{*u})$ and $E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})$, by simulation using identical random number streams to facilitate comparison between the expected profits. The optimality gaps of the two heuristics are computed as

$$\begin{aligned} \frac{UB - H1}{UB} &= \frac{E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u}) - E\Pi^s(\mathbf{b}^{*s}, \mathbf{q}^{*s})}{E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})}, \\ \frac{UB - H2}{UB} &= \frac{E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u}) - E\Pi^d(\mathbf{b}^{*u}, \mathbf{q}^{*u})}{E\Pi^u(\mathbf{b}^{*u}, \mathbf{q}^{*u})}. \end{aligned}$$

Note that the solutions given by H2 allow for dynamic substitution while those from H1 do not because there is no overlap between the coverage regions of products in H1. Thus, a comparison of

these solutions shows the benefit from incorporating dynamic substitution in assortment planning.

We find that the average optimality gaps of the two heuristics are 1.44% and 1.24% across all problem instances, respectively for H1 and H2. The two heuristics give identical solutions in 1330 (42.2%) instances, showing that static substitution yields the optimal solution in these instances. H2 performs better than H1 in 1227 (39.0%) instances, and H1 performs better than H2 in the remaining 593 (18.8%) instances; these differences are statistically significant at $p=0.05$. The instances where H1 performs better than H2 occur when L is small. It is possible that the solution from H2 in these instances could be improved by reducing the search interval used in numerical optimization. H2 performs progressively better than H1 as λ and γ decrease or as L increases. This shows that the impact of dynamic substitution on the profits of the retailer is the highest for low demand items (λ) with greater customer heterogeneity (γ) and with greater degree of substitutability (L).

Figures 2 and 3 show the optimality gaps for H1 and H2, respectively, as a function of γ , λ and L when K is fixed at 0.1 and c at 5. We observe that the optimality gaps of both heuristics decrease as γ increases, i.e., as the customer population becomes more homogeneous and concentrated in a smaller region of the preference spectrum. This is so because the need for variety decreases as the population becomes more homogeneous. In the extreme case, if all consumers had identical preferences, then a single product would be sufficient to maximize expected profit. The optimality gaps also decrease as L increases, i.e., as products become more substitutable. The explanation for this result is similar to that for γ since an increase in L also decreases the need for variety. Also the optimality gaps decrease as the mean demand λ increases. This is so because the incremental profit obtained under retailer-controlled substitution or dynamic substitution becomes a progressively smaller fraction of total sales as λ increases and the coefficient of variation of demand decreases.

Table 2 compares the assortments, inventory levels and expected profits given by H1 and H2 for a few representative problem instances. The columns in the table correspond to different values of γ and K for three pairs of values of L and λ , namely (0.25, 20), (0.25, 50) and (0.15, 50). The value of c is fixed at 5.

We observe that the amount of variety offered in H2 is always greater than or equal to the amount of variety offered in H1. Correspondingly, in H2, products are often located at distances smaller than $2L$. This is so because, by allowing overlap between the coverage intervals of products, the retailer can allocate to customers a product that is not their first choice product, resulting in

more sales. This benefit increases with the degree of overlap so that the optimal assortment may be such that a customer can get positive utility from several products (in Table 2, up to 7 products). Further, the number of products whose coverage regions overlap varies along the attribute space, being generally higher in regions of higher customer density. Thus, consumers located in regions of higher density generally have a greater variety of products to choose from than consumers located in regions of lower density. This is a significant difference between the solutions obtained from H1 and H2, since it shows that the solution obtained from H2 fits real life observations better than that obtained from H1. Note that the benefit from overlap comes at the expense of an increase in fixed cost and possibly a decrease in total market coverage.

Table 2 also shows how the amount of variety changes with the parameters of the problem instances. We observe that the amount of variety offered by the retailer under both H1 and H2 increases as λ increases, and as K, L or γ decrease. Further, the difference in variety between dynamic and static substitution increases as λ and L increase or as K and γ decrease. The difference in expected profit behaves the same way as the difference in variety.

Finally, Table 3 presents a few problem instances illustrating the effect of c on the optimality gaps of both heuristics and the size of the optimal assortment. We find that the optimality gaps decline as c decreases, however, the size of the optimal assortment is not monotonic in c for both H1 and H2. First, a decrease in c expands the profitable region, $[\alpha, \beta]$. Second, a decrease in c increases the expected profit of the more popular products proportionately more than other products. Third, a decrease in c increases the service levels of products, and thus, decreases the expected gain from substitution between products. While the first effect is a drive towards more variety, the other two lead to less variety. As a result, we find that the size of the optimal assortment can increase or decrease in c . The salvage value, v , has the same effect as c .

In §4.2, we contrasted the outcomes under MNL and locational choice models for static substitution. Under retailer-controlled substitution, the same insights continue to hold. In particular, the most popular product need not be included in the optimal assortment. Further, the partial market coverage is driven by fixed costs rather than by the substitution structure as in the MNL model. In addition, Mahajan and van Ryzin (2001) present a numerical example for the MNL model showing that the optimal assortment under dynamic substitution is larger than that under static substitution. This result is consistent with our results for the locational choice model. Unfortunately, there are no known results on the effects of mean demand or substitution structure on

the optimal assortment for the MNL model under dynamic substitution. There is need for future research on this topic.

6 Conclusions

We have analyzed the optimal assortment planning and inventory decisions for a retailer when its demand is represented by a locational model of consumer choice. By comparing our results with those under the MNL model, we have obtained insights into the effects of substitution structure, degree of heterogeneity of consumers, degree of substitutability of products, and fixed costs on the optimal solution. When consumers substitute dynamically, our heuristic can be used effectively to solve the assortment planning and inventory problem. The heuristic is particularly useful when mean demand is large, and algorithms based on the enumeration of sample paths become more complex.

The locational choice model is useful to analyze markets when customer preferences are heterogeneous or consumers substitute locally. However, this model has some important limitations. First, the model considers only continuous-valued attributes. The qualitative insights obtained by us apply more generally to the case when attributes are discrete-valued or where the set of available products is finite. However, it would be computationally more intensive to obtain the optimal assortment in this case. Another limitation of the locational choice model is with respect to the nature of randomness in customer choice. In this model, given the most preferred good of a consumer, the sequence of product selections that the consumer makes is known precisely. It would be natural to generalize the model to allow randomness both in the locations of most preferred goods of consumers and in customers' sequence of product selections. A third limitation of this model is that the total demand for an assortment that covers the entire attribute space is the same regardless of the number of products in the assortment. In contrast, demand always increases with variety in the MNL model.

Our paper yields some key managerial insights that may apply more generally in spite of the limitations of the locational choice model. We find that under static substitution, products in the optimal assortment are equally spaced out at intervals such that there is no substitution between them regardless of the distribution of customer preferences. In contrast, under dynamic substitution, products are located closer to each other resulting in more variety especially in regions of higher consumer density, thus allowing the retailer to benefit from substitution. We also find that

the optimal assortment may not include the most popular product because stocking this product may fragment the demand for remaining products such that they suffer diseconomies of scale. Even though in practice, other considerations such as competition and the impact on store loyalty may force the retailer to stock the most popular product, she may want to offer a relatively low inventory of that product in order to take advantage of substitution. Finally, we find that the static substitution solution serves as a good approximation for the dynamic substitution problem when consumers are more homogeneous in preferences, or when they are willing to substitute in a narrow region around their most preferred goods, or when the coefficient of variation of demand is low. In other cases, dynamic substitution can have a large impact on profits.

Our paper points to some important directions for future research. First, the locational choice model may be generalized to incorporate changes in price and quality levels along the attribute space. Price impacts our model in two ways: it affects the sales revenue, and it affects the coverage distance L of each product. If prices are allowed to vary across products, then it is possible that products located in different parts of the market are offered at differing prices even when they are only horizontally differentiated. Second, while the locational choice model yields results that differ from the MNL model, it has some similarities with the nested MNL model. Both models imply a localized substitution structure and do not impose the assumption of independence of irrelevant alternatives. To our knowledge, the structure of the optimal assortment under the nested MNL model has not been studied so far. It is possible that its results may be similar to those from the locational choice model, however, this is a subject for future research.

References

- Agrawal, V., G. Lopomo, S. Seshadri. 2002. Web Based Capacity Allocation Strategies for Customers with Heterogeneous Preferences. *Electronic Commerce Research* **2** 359-384.
- Alptekinoglu, A. 2004. Mass Customization vs. Mass Production: Variety and Price Competition. *MESOM* **6**(1) 98-103.
- Anderson, S., A. De Palma, J.F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.
- Ansari, A., N. Economides, A. Ghosh. 1994. Competitive Positioning in Markets with Non-Uniform Preferences. *Marketing Science* , **13**(3) 248-273.

- Ansari, A., N. Economides, J. Steckel. 1998. The Max-Min-Min Principle of Product Differentiation. *Journal of Regional Science* , **38**(2) 207-230.
- Aydin, G., J.K. Ryan. 2000. Product Line Selection and Pricing under the Multinomial Logit Choice model. Working paper. Stanford University.
- Bassok, Y., R. Anupindi, R. Akella. 1999. Single-Period Multiproduct Inventory Models with Substitution. *Operations Research* **47** 632-642.
- Birge, J.R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.
- Cachon, G., C. Terwiesch, Y. Xu. 2005. Retail Assortment Planning in the Presence of Consumer Search. *M&SOM* **7**(4) 330-346.
- Chen, F., J. Eliashberg, P. Zipkin. 1998. Customer Preferences, Supply-Chain Costs, and Product-Line Design. In *Product Variety Management: Research Advances*, T.-H. Ho and C. S. Tang (Eds.), Kluwer Academic Publishers.
- Chong, J-K., T-H. Ho, C. S. Tang. 2001. A Modeling Framework for Category Assortment Planning. *M&SOM*, **3**(3) 191-210.
- de Groote, X. 1994. Flexibility and Marketing/Manufacturing Coordination. *International Journal of Production Economics* **36** 153-167.
- Guadagni, P., J. Little. 1983. A Logit Model of Brand Choice Calibrated on Scanner Data. *Marketing Science* **2**(3) 203-238.
- Ho, T-H., C. S. Tang. 1998. *Research Advances in Product Variety Management*. Kluwer Academic Publishers.
- Hotelling, H. 1929. Stability in Competition. *Economic Journal* **39** 41-57.
- Kok, A. G., M. L. Fisher. 2004. Demand Estimation and Assortment Optimization under Substitution: Methodology and Application. Working Paper. Fuqua School of Business, Duke University.
- Lancaster, K. 1966. A New Approach to Consumer Theory. *Journal of Political Economy* **74** 132-157.

- Lancaster, K. 1975. Socially Optimal Product Differentiation. *American Economic Review* **65** 567-585.
- Lancaster, K. 1979. *Variety, Equity and Efficiency*. Columbia University Press, New York.
- Lancaster, K. 1990. The Economics of Product Variety: A survey. *Marketing Science* **9** 189-206.
- Lippman, S. A., K. F. McCardle. 1997. The Competitive Newsboy. *Operations Research* **45** 54-65.
- Mahajan, S., G. van Ryzin. 2001. Stock Retail Assortment under Dynamic Consumer Substitution. *Operations Research* **49** 334-351.
- Marshall, A.W., I. Olkin. 1979. *Inequalities: Theory of Majorization and its Application*. Academic Press, New York.
- McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. in *Frontiers in Econometrics* P. Zarembka, ed. Academic Press, New York. 105-142.
- Netessine, S., N. Rudi. 2003. Centralized and competitive inventory models with demand substitution. *Operations Research* **51**(2) 329-335.
- Porteus E.L. 2002. *Foundations of Stochastic Inventory Theory*. Stanford Univ Press.
- Rajaram, K. 2001. Assortment Planning in Fashion Retailing: Methodology, Application and Analysis. *European Journal of Operational Research* **129** 186-208.
- Rajaram, K., C. S. Tang. 2001. The Impact of Product Substitution on Retail Merchandising. *European Journal of Operational Research* **135** 582-620.
- Shaked, M., J. G. Shanthikumar. 1994. *Stochastic Orders and Their Applications*. Academic Press, San Diego, CA.
- Shamir, R., B. Dietrich. 1990. Characterization and Algorithms for Greedily Solvable Transportation Problems. *Proceedings of the First ACM/SIAM Symposium on Discrete Algorithms*. SIAM, Philadelphia, PA.
- Smith, S.A, N. Agrawal. 2000. Management of Multi-Item Retail Inventory Systems with Demand Substitution. *Operations Research* **48** 50-64.

van Ryzin, G., S. Mahajan. 1999. On the Relationship between Inventory Costs and Variety Benefits in Retail Assortments. *Management Science* **45** 1496-1509.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill/Irwin, New York.

Appendix: Proofs

The following definition and lemma are useful for the proof of Proposition 1:

Definition 1. (Marshall and Olkin 1979) A vector $\bar{x} = (x_1, \dots, x_n)$ is said to majorize the vector $\bar{y} = (y_1, \dots, y_n)$, denoted $\bar{x} \succ \bar{y}$, if

$$\sum_{j=1}^k x_{[j]} \geq \sum_{j=1}^k y_{[j]}, \quad \text{for } k = 1, \dots, n-1, \quad \text{and} \quad \sum_{j=1}^n x_j = \sum_{j=1}^n y_j$$

where $x_{[j]}$ denotes the j th greatest element of vector \mathbf{x} .

Lemma 3. (Marshall and Olkin 1979) *If $f : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function and $\bar{x} \succ \bar{y}$ then $\sum_{i=1}^n f(x_i) \geq \sum_{i=1}^n f(y_i)$.*

Proof of Proposition 1. In the following proof, $\tilde{\mathbf{b}}$ and $\hat{\mathbf{b}}$ denote modifications of the vector \mathbf{b} , $\tilde{\mathbf{p}}$ and $\hat{\mathbf{p}}$, respectively denote the corresponding first choice probability vectors, and $\tilde{\mathbf{I}}$ and $\hat{\mathbf{I}}$, respectively denote the corresponding vectors for lengths of coverage intervals.

(i) Follows directly from the definitions of α and β .

(ii) Suppose to the contrary that the optimal value of \mathbf{b} is such that $b_j^+ < b_{j+1}^-$ for some j . Suppose that the distribution of preferences is unimodal. The proof in the case of uniform distribution is simpler and, thus, omitted. There are three cases depending on the location of the mode x^* . In each case, create a new solution $\tilde{\mathbf{b}}$ as follows:

- If $x^* \geq b_{j+1}^-$, then set $\tilde{b}_i = b_i + (b_{j+1}^- - b_j^+)$ for all $i \leq j$, and $\tilde{b}_i = b_i$ for all $i \geq j + 1$.
- If $x^* \leq b_j^+$, then set $\tilde{b}_i = b_i$ for all $i \leq j$, and $\tilde{b}_i = b_i - (b_{j+1}^- - b_j^+)$ for all $i \geq j + 1$.
- If $b_j^+ < x^* < b_{j+1}^-$, then set $\tilde{b}_i = b_i + (x^* - b_j^+)$ for all $i \leq j$, and $\tilde{b}_i = b_i - (b_{j+1}^- - x^*)$ for all $i \geq j + 1$.

$\tilde{\mathbf{b}}$ satisfies the equation $\tilde{b}_j^+ = \tilde{b}_{j+1}^-$. Further, the first choice probabilities, $\tilde{\mathbf{p}}$, corresponding to $\tilde{\mathbf{b}}$ are at least as large as the first choice probabilities, \mathbf{p} , corresponding to \mathbf{b} because $\tilde{\mathbf{b}}$ is obtained

by shifting the products towards the mode of the distribution. The result follows by contradiction since, by Lemma 1, expected profit is increasing in p_j for all j .

(iii) This proof is similar to that for (ii). If $x^* < b_1^-$, then setting $\tilde{b}_1 = x^* + L$ and $\tilde{b}_j = b_j - (b_1^- - x^*)$ for $j = 2, \dots, n$ increases the first choice probability of each product and, therefore, by Lemma 1, increases expected profit. Likewise, if $x^* > b_n^+$.

(iv) If \mathbf{b} is such that $b_1 \geq \alpha + 2L$, then add a product with specification $b_1 - 2L$ to the assortment. From the definition of α , this product yields positive expected profit. Further, it does not affect the first choice probability of any other product. Thus, total expected profit increases. Likewise, if $b_n \leq \beta - 2L$.

(v) We show the proof for the case of a unimodal distribution. The proof for a uniform distribution is simpler and therefore omitted. Let $\mathbf{b}^- = (b_1^-, \dots, b_n^-)$ and $\mathbf{b}^+ = (b_1^+, \dots, b_n^+)$ and consider the following relaxation of problem \mathbb{P}^s :

$$\begin{aligned} \mathbb{P}_r^s : \quad & \max_{(\mathbf{b}^-, \mathbf{b}^+) \in \mathcal{B}^\pm} \quad E\Pi^s(\mathbf{b}^-, \mathbf{b}^+) = \sum_{j=1}^n \left[(r-c)p_j\lambda - (r-v)\sqrt{p_j}\lambda\phi(z) \right] - Kn \\ & \text{s.t.} \quad \begin{cases} p_j = F(b_j^+) - F(b_j^-) & \text{for } j = 1, \dots, n \\ b_j^+ \leq b_j^- + 2L & \text{for } j = 1, \dots, n \\ b_j^- \leq b_j^+ & \text{for } j = 1, \dots, n. \end{cases} \end{aligned}$$

Here, \mathcal{B}^\pm is defined using Proposition 1(ii) and (iii) as

$$\mathcal{B}^\pm \equiv \left\{ (\mathbf{b}^-, \mathbf{b}^+) \in \mathcal{L} \times \mathcal{L} : b_j^+ = b_{j+1}^- \text{ for } j = 1, \dots, n-1 \text{ and } b_1^- \leq x^* \leq b_n^+ \right\}.$$

Note that a solution to \mathbb{P}_r^s may not be feasible for \mathbb{P}^s because there may not exist a vector \mathbf{b} corresponding to $(\mathbf{b}^-, \mathbf{b}^+)$ which satisfies the requirements:

$$b_j^- = \max \left\{ b_j - L, \frac{b_j + b_{j-1}}{2} \right\} \quad \text{and} \quad b_j^+ = \min \left\{ b_j + L, \frac{b_j + b_{j+1}}{2} \right\} \quad \text{for } j = 1, \dots, n.$$

We show that the optimal solution for \mathbb{P}_r^s satisfies $b_j^+ - b_j^- = 2L$ for $j = 1, \dots, n$. Then, a feasible vector \mathbf{b} is obtained by letting $b_j = b_j^- + L = b_j^+ - L$ for $j = 1, \dots, n$. This shows that the optimal solution for \mathbb{P}_r^s is feasible for \mathbb{P}^s , and thus, optimal for \mathbb{P}^s . This proves property (v) for \mathbb{P}^s .

To prove (v) for \mathbb{P}_r^s , we show that the first choice probability vector corresponding to any initial solution is majorized by the first choice probability vector associated with a solution that satisfies property (v). The proof is by construction, that is, we give a sequence of modifications to the vector $(\mathbf{b}^-, \mathbf{b}^+)$ such that at each step we obtain a first choice probability vector that majorizes

all the previous ones. By Lemma 1 and Lemma 3, $\tilde{\mathbf{p}} \succ \mathbf{p}$ implies $\sum_{j=1}^S E\Pi_j^s(\tilde{p}_j) \geq \sum_{j=1}^n E\Pi_j^s(p_j)$, and therefore, the expected profit increases at each step.

Consider an initial feasible solution $(\mathbf{b}^-, \mathbf{b}^+)$ for \mathbb{P}_r^s , and let $l_j = b_j^+ - b_j^-$. Also let j^* be the product such that $b_{j^*}^- < x^* < b_{j^*}^+$. If $x^* = b_j^+ = b_{j+1}^-$ for some j , then no such j^* exists and the proof is slightly simplified. We ignore this case.

Step 1. Consider the following swaps of first choice intervals to create new solutions $(\tilde{\mathbf{b}}^-, \tilde{\mathbf{b}}^+)$ and $(\hat{\mathbf{b}}^-, \hat{\mathbf{b}}^+)$ with first choice probabilities $\tilde{\mathbf{p}}$ and $\hat{\mathbf{p}}$:

- (i) If $\exists j < j^*$ such that $l_{j-1} > l_j$, then $\tilde{b}_{j-1}^+ = \tilde{b}_j^- := b_{j-1}^- + l_j$.
- (ii) If $\exists j > j^*$ such that $l_{j+1} > l_j$, then $\tilde{b}_j^+ = \tilde{b}_{j+1}^- := b_j^- + l_{j+1}$.
- (iii) If $l_{j^*} < \min\{l_{j^*-1}, l_{j^*+1}\}$, then $\tilde{b}_{j^*-1}^+ = \tilde{b}_{j^*}^- := b_{j^*-1}^- + l_{j^*}$ and $\hat{b}_{j^*}^+ = \hat{b}_{j^*+1}^- := b_{j^*}^- + l_{j^*+1}$.

In each case, the remaining values in $(\tilde{\mathbf{b}}^-, \tilde{\mathbf{b}}^+)$ and $(\hat{\mathbf{b}}^-, \hat{\mathbf{b}}^+)$ are kept unchanged. Each swap satisfies all the constraints of problem \mathbb{P}_r^s and preserves the definition of j^* . It is straightforward to show in cases (i) and (ii) that $\tilde{\mathbf{p}} \succ \mathbf{p}$.

Case (iii) is somewhat tricky. We show that either $\tilde{\mathbf{p}} \succ \mathbf{p}$ or $\hat{\mathbf{p}} \succ \mathbf{p}$ (or both). First note that $\tilde{p}_k = p_k$ for $k \neq j^* - 1, j^*$ and $\hat{p}_k = p_k$ for $k \neq j^*, j^* + 1$. We have,

$$\begin{aligned}\tilde{p}_{j^*} &= p_{j^*} + \left[F(b_{j^*-1}^+) - F(\tilde{b}_{j^*-1}^+) \right] \geq p_{j^*}, \\ \hat{p}_{j^*} &= p_{j^*} + \left[F(\hat{b}_{j^*}^+) - F(b_{j^*}^+) \right] \geq p_{j^*}.\end{aligned}$$

We show that it cannot be true that $\tilde{p}_{j^*} < p_{j^*-1}$ as well as $\hat{p}_{j^*} < p_{j^*+1}$. This is because the first one implies $f(b_{j^*}^+) < f(b_{j^*}^-)$ and the second one implies $f(b_{j^*}^+) > f(b_{j^*}^-)$. Therefore, either $\tilde{p}_{j^*} > \max\{p_{j^*}, p_{j^*-1}\}$ or $\hat{p}_{j^*} > \max\{p_{j^*}, p_{j^*+1}\}$, implying that either $\tilde{\mathbf{p}} \succ \mathbf{p}$ or $\hat{\mathbf{p}} \succ \mathbf{p}$. In either case, Step 1 improves the solution.

We apply Step 1 repeatedly until none of the three conditions is satisfied. In each iteration, we obtain a new first choice probability vector that majorizes all the previous ones. Thus, the final vector $(\mathbf{b}^-, \mathbf{b}^+)$ satisfies one of the following chains of inequalities:

$$l_1 \leq l_2 \leq \dots \leq l_{j^*-1} \leq l_{j^*} \geq l_{j^*+1} \geq \dots \geq l_{n-1} \geq l_n \quad (14)$$

$$l_1 \leq l_2 \leq \dots \leq l_{j^*-1} \geq l_{j^*} \geq l_{j^*+1} \geq \dots \geq l_{n-1} \geq l_n \quad (15)$$

$$l_1 \leq l_2 \leq \dots \leq l_{j^*-1} \leq l_{j^*} \leq l_{j^*+1} \geq \dots \geq l_{n-1} \geq l_n. \quad (16)$$

Step 2. Suppose that vector $(\mathbf{b}^-, \mathbf{b}^+)$ satisfies inequalities (14). Consider the following sets of modifications to create a new solution $(\hat{\mathbf{b}}^-, \hat{\mathbf{b}}^+)$:

- (i) Let $j_a = \max\{k \leq j^* : l_k < 2L\}$, $j_b = \min\{k : l_k \geq 0\}$ and $\delta = \min\{2L - l_{j_a}, l_{j_b}\}$. Set $\widehat{b}_j^- = \widehat{b}_{j-1}^+ = b_j^- - \delta$ for $j = j_b + 1, \dots, j_a$, and $\widehat{b}_j^- = b_j^-$, $\widehat{b}_j^+ = b_j^+$ for all remaining values.
- (ii) Let $j_a = \min\{k > j^* : l_k < 2L\}$, $j_b = \max\{k : l_k \geq 0\}$ and $\delta = \min\{2L - l_{j_a}, l_{j_b}\}$. Set $\widehat{b}_j^+ = \widehat{b}_{j+1}^- = b_j^+ + \delta$ for $j = j_a, \dots, j_b - 1$, and $\widehat{b}_j^- = b_j^-$, $\widehat{b}_j^+ = b_j^+$ for all remaining values.

In each case, $(\widehat{\mathbf{b}}^-, \widehat{\mathbf{b}}^+)$ satisfies the constraints of \mathbb{P}_r^s and preserves inequalities (14). Further, we show that $\widehat{\mathbf{p}} \succ \mathbf{p}$. We prove the result for case (i); case (ii) is similar. Also, the proofs for the remaining inequalities obtained after Step 1, (15) and (16), are similar and are omitted.

In case (i), let $x_i = F(b_{j_a+1-i}^-) - F(b_{j_a+1-i}^- - \delta) > 0$ for $i = 1, \dots, j_a - j_b$. Since $j_a \leq j^*$, we have $x_1 \geq x_2 \geq \dots \geq x_{j_a-j_b}$. The elements of $\widehat{\mathbf{p}}$ are now given by

$$\begin{aligned} \widehat{p}_{j_a} &= p_{j_a} + x_1, \\ \widehat{p}_{j_a-i} &= p_{j_a-i} - x_i + x_{i+1} \quad i = 1, \dots, j_a - j_b - 1, \\ \widehat{p}_{j_b} &= p_{j_b} - x_{j_a-j_b}. \end{aligned}$$

From (14), we have $\widehat{l}_{j_b+1} \leq \dots \leq \widehat{l}_{j_a}$. Thus, $\widehat{p}_{j_b+1} \leq \dots \leq \widehat{p}_{j_a}$. Finally,

$$\sum_{i=1}^k \widehat{p}_{j_a-i+1} = \sum_{i=1}^k p_{j_a-i+1} + x_k \geq \sum_{i=1}^k p_{j_a-i+1} \quad \text{for all } k = 1, \dots, j_a - j_b,$$

and

$$\sum_{i=1}^{j_a-j_b+1} \widehat{p}_{j_a-i+1} = \sum_{i=1}^{j_a-j_b+1} p_{j_a-i+1}.$$

Therefore, $\widehat{\mathbf{p}} \succ \mathbf{p}$. This shows that Step 2 improves the solution.

We apply Step 2 repeatedly until none of the two conditions is satisfied. After that we renumber the products so that only the products with $b_j^+ > b_j^-$ are included in the assortment. The final vector $(\mathbf{b}^-, \mathbf{b}^+)$ is such that $l_j = 2L$ for all products with the possible exception of the two located at the extremities of the market. In this case, we set $\widetilde{b}_1^- = b_1^+ - 2L$ and $\widetilde{b}_n^+ = b_n^- + 2L$. By Lemma 1, this increases expected profit. \square

Proof of Proposition 2. For a given assortment \mathbf{b} and inventory vector \mathbf{q} , consider a sample path, ω , of the customer arrival and choice process. Let $R_j^s(i)(\omega)$ and $R_j^d(i)(\omega)$ denote the remaining inventory of product j after the arrival of customer i under static and dynamic substitution, respectively. The reference to the sample path is hereafter omitted for convenience.

We show that if $R_j^s(i) \geq R_j^d(i)$ for $j = 1, \dots, n$, then $R_j^s(i+1) \geq R_j^d(i+1)$ for $j = 1, \dots, n$. Without loss of generality, let product k be the first choice of customer $i+1$. There are three cases:

- (1) If $R_k^s(i) \geq R_k^d(i) > 0$, then customer $i + 1$ gets his first choice under both static and dynamic substitution. Thus,

$$\begin{aligned} R_k^s(i+1) &= R_k^s(i) - 1 \geq R_k^d(i) - 1 = R_k^d(i+1) \\ R_j^s(i+1) &= R_j^s(i) \geq R_j^d(i) = R_j^d(i+1), \quad j \neq k. \end{aligned}$$

- (2) If $R_k^s(i) > 0$ and $R_k^d(i) = 0$, then customer $i + 1$ gets his first choice under static substitution and may get an alternative choice under dynamic substitution. Thus,

$$\begin{aligned} R_k^s(i+1) &= R_k^s(i) - 1 \geq R_k^d(i) = R_k^d(i+1) = 0 \\ R_j^s(i+1) &= R_j^s(i) \geq R_j^d(i) \geq R_j^d(i+1), \quad j \neq k. \end{aligned}$$

- (3) If $R_k^s(i) = R_k^d(i) = 0$, then customer $i + 1$ leaves unsatisfied under static substitution but may get an alternative choice under dynamic substitution. Thus,

$$\begin{aligned} R_k^s(i+1) &= R_k^s(i) = R_k^d(i) = R_k^d(i+1) = 0 \\ R_j^s(i+1) &= R_j^s(i) \geq R_j^d(i) \geq R_j^d(i+1), \quad j \neq k. \end{aligned}$$

Let $i = 0$ correspond to the initial assortment in the store before the visit of the first customer. The condition $R_j^s(i) \geq R_j^d(i)$ is satisfied for $i = 0$ for all j because $R_j^s(0) = R_j^d(0) = q_j$ for all j . Hence, by induction, we obtain $R_j^s(N) \geq R_j^d(N)$ for all $j = 1, \dots, n$. Further, $S_j^s = q_j - R_j^s \leq q_j - R_j^d = S_j^d$ for all $j = 1, \dots, n$. This completes the proof since profit is an increasing function of sales. \square

Proof of Lemma 2. Problem (9) can be formulated as a restricted transportation problem in which the source nodes are j for $j = 1, \dots, n$ with inventory q_j , and the destination nodes are (l, i) for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$ with corresponding demand $\eta_{(l,i)}$. We add one dummy source, indexed 0, with inventory $q_0 = \sum_{l=1}^n \sum_{i=1}^{n-l+1} \eta_{j,(l,i)}$, and one dummy destination, indexed $(0,0)$, with demand $\eta_{(0,0)} = \sum_{j=1}^n q_j$ in order to balance the flow. Let $c_{j,(l,i)}$ be the cost associated to arc $(j, (l, i))$. Arcs to each destination (l, i) from sources $i, \dots, i + l - 1$ have a cost of minus one. Arcs $(0, (l, i))$ from the dummy source to all destinations, and arcs $(j, (0, 0))$ from all sources to the dummy destination have zero costs. All other arcs, i.e., those corresponding to constraint (13), are forbidden and have infinite cost. Thus, the objective function of (9) is equivalent to:

$$\min \sum_{j=1}^n \sum_{l=1}^n \sum_{i=1}^{n-l+1} c_{j,(l,i)} x_{j,(l,i)}.$$

The algorithm in Lemma 2 rank-orders the non-dummy arcs connecting inventory and demand in a specific way. Now add the dummy arcs at the end of that sequence (in any order). It can be verified that the sequence thus obtained satisfies the following property: whenever $(j, (l, i))$ precedes $(j, (l', i'))$ and $(j', (l, i))$ then:

$$c_{(j,(l,i))} + c_{(j',(l',i'))} \leq c_{(j,(l',i'))} + c_{(j',(l,i))}$$

where $c_{(j,(l,i))}$ is the cost associated with arc $(j, (l, i))$. Such a sequence is called a Monge sequence. Shamir and Dietrich (1990) establish that a greedy solution to a transportation problem is optimal if and only if it rank-orders the arcs in a Monge sequence. \square

Proof of Theorem 1. We write the maximization of $E\Pi^u(\mathbf{b}, \mathbf{q})$ with respect to \mathbf{q} as a two-stage stochastic linear program with fixed recourse:

$$\begin{aligned} E\Pi^u(\mathbf{b}) &= \max_{\mathbf{q} \geq 0} E\Pi^u(\mathbf{b}, \mathbf{q}) \\ &= \max_{\mathbf{q} \geq 0} \{E_\eta [\Pi^u(\mathbf{b}, \mathbf{q}, \omega)]\}. \end{aligned}$$

Here, $\Pi^u(\mathbf{b}, \mathbf{q}, \omega)$ is as defined in problem (9)-(13), and η denotes the vector of stochastic variables $\eta_{(l,i)}$ for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$. The first stage decision variable is $\mathbf{q} = (q_1, \dots, q_n)$, and the second stage decisions variables are $x_{j,(l,i)}$ for $l = 1, \dots, n$, $i = 1, \dots, n - l + 1$ and $j = i, \dots, i + l - 1$.

For each realization of $\eta_{(l,i)}$ there exists a solution $x_{j,(l,i)}$ given by Lemma 2. Therefore, $E\Pi^u(\mathbf{b}, \mathbf{q})$ is defined for every value of $\eta_{(l,i)}$. Since the variables $\eta_{(l,i)}$ have finite second moments, we use Theorem 6(a) of Birge and Louveaux (1997: p.89) applied to a maximization problem to conclude that $E\Pi^u(\mathbf{b}, \mathbf{q})$ is concave in \mathbf{q} . \square

Proof of Proposition 5. (i) Suppose that we have an assortment, \mathbf{b} , with $b_1 < \alpha$. Decompose this assortment into two parts, $\mathbf{b}_1 = (b_1)$, consisting of a single product located at b_1 , and $\mathbf{b}_{\bar{1}} = (b_2, \dots, b_n)$ consisting of the remaining $n - 1$ products in \mathbf{b} .

For a given vector \mathbf{q} and sample path ω , let $\mathbf{q}_1 = (q_1)$ and $\mathbf{q}_{\bar{1}} = (q_2, \dots, q_n)$. We have:

$$\Pi^u(\mathbf{b}, \mathbf{q}, \omega) \leq \Pi^u(\mathbf{b}_1, \mathbf{q}_1, \omega) + \Pi^u(\mathbf{b}_{\bar{1}}, \mathbf{q}_{\bar{1}}, \omega).$$

since the profit values on the RHS double-count the $\sum_{k=2}^n \eta_{(k,1)}$ customers who get a positive utility from consuming product 1 as well as some other product in \mathbf{b} .

Taking expectations we get:

$$\begin{aligned} E\Pi^u(\mathbf{b}, \mathbf{q}) &\leq E\Pi^u(\mathbf{b}_1, \mathbf{q}_1) + E\Pi^u(\mathbf{b}_{\bar{1}}, \mathbf{q}_{\bar{1}}) \\ &< E\Pi^u(\mathbf{b}_{\bar{1}}, \mathbf{q}_{\bar{1}}). \end{aligned}$$

Here the last inequality follows because $E\Pi^u(\mathbf{b}_1, \mathbf{q}_1) = E\Pi^s(\mathbf{b}_1, \mathbf{q}_1) < 0$ from Proposition 1(i) since $b_1 < \alpha$. Thus, \mathbf{b} cannot be an optimal solution.

(ii) & (iii) Let $p_{(l,i)}$ be the probability that a customer visiting the store is of type (l, i) . The number of such customers, $\eta_{(l,i)}$, is Poisson with rate $\lambda p_{(l,i)}$. Note that $\eta_{(l,i)}$ is stochastically increasing in $p_{(l,i)}$ for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$ because a Poisson random variable is stochastically increasing in its mean. Also $\eta_{(l,i)}$ are the RHS of the first set of constraints in (9).

Suppose we have two assortments $\tilde{\mathbf{b}}$ and \mathbf{b} such that $\tilde{\eta}_{(l,i)} \geq_{s.t.} \eta_{(l,i)}$ for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$ where $\geq_{s.t.}$ denotes larger under first order stochastic dominance. By theorem Theorem 1.A.1. of Shaked and Shanthikumar (1994), there exist random variables $\tilde{\eta}'_{(l,i)}$ and $\eta'_{(l,i)}$ such that $\tilde{\eta}'_{(l,i)} \stackrel{D}{=} \tilde{\eta}_{(l,i)}$, $\eta'_{(l,i)} \stackrel{D}{=} \eta_{(l,i)}$ and $\tilde{\eta}'_{(l,i)} \geq \eta'_{(l,i)}$ with probability one for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$. This along with the fact that the value of the objective function is increasing in the RHS of the constraints, gives

$$\Pi^u(\tilde{\mathbf{b}}, \mathbf{q}) \geq_{s.t.} \Pi^u(\mathbf{b}, \mathbf{q}) \quad \forall \mathbf{q}.$$

This implies that $E\Pi^u(\mathbf{b}, \mathbf{q})$ is increasing in $p_{(l,i)}$ for $l = 1, \dots, n$ and $i = 1, \dots, n - l + 1$. The rest of the proof is similar to Proposition 1 (ii) or (iii).

(iv) Proof is identical to that of Proposition 1 (iv).

(v) We show the proof for the case $n = 2$ but the reasoning holds for any value of n . Suppose that we have an optimal assortment \mathbf{b} with $b_1 < L$.

Now consider the alternative assortment $\tilde{\mathbf{b}}$ such that $\tilde{b}_1 = L$ and $\tilde{b}_2 = b_2$.

We have $\tilde{p}_{(l,i)} = p_{(l,i)}$, for $(l, i) = (1, 1), (1, 2)$ and $\tilde{p}_{(2,1)} > p_{(2,1)}$. Since $E\Pi^u(\mathbf{b}, \mathbf{q})$ is increasing in $p_{(l,i)}$ for all fixed \mathbf{q} , we get a contradiction to the fact that \mathbf{b} is optimal. \square

Table 2: Effects of λ , γ , K and L on optimal assortments, inventory levels, and expected profits under heuristics H1 and H2

γ	1					3					5					
K	20	10	5	1	0.1	20	10	5	1	0.1	20	10	5	1	0.1	
L=0.25, $\lambda=20$																
Retailer-controlled substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	0.25	12	0.25	12	0.25	12	0.25	9	0.25	8	0.5	18	0.31	11	0.31	11
2	0.75	12	0.75	12	0.75	12	0.5	6	0.41	4			0.69	11	0.69	11
3							0.75	8	0.59	4					0.74	6
4									0.75	7						0.74
EII	44.86	64.86	74.86	85.74	88.74						48.99	67.04	77.04	85.15	87.85	59.75
Static substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	0.25	12	0.25	12	0.25	12	0.25	12	0.25	12	0.5	18	0.26	12	0.26	12
2	0.75	12	0.75	12	0.75	12	0.75	12	0.75	12			0.76	11	0.76	11
EII	44.86	64.86	74.86	82.86	84.66						48.99	63.63	73.63	81.63	83.43	59.75
L=0.25, $\lambda=50$																
Retailer-controlled substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	0.25	28	0.25	28	0.25	20	0.25	20	0.25	18	0.3	27	0.3	27	0.3	27
2	0.75	28	0.75	28	0.5	14	0.5	14	0.41	9	0.7	27	0.7	27	0.5	24
3					0.75	20	0.75	20	0.59	9					0.75	15
4									0.75	18						0.75
EII	186.25	206.25	217.70	229.70	232.54						190.37	210.37	220.37	228.40	231.10	191.87
Static substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	0.25	28	0.25	28	0.25	28	0.25	28	0.25	28	0.21	24	0.21	24	0.21	24
2	0.75	28	0.75	28	0.75	28	0.75	28	0.75	28	0.71	31	0.71	31	0.71	31
EII	186.25	206.25	216.25	224.25	226.05						184.86	204.86	214.86	222.86	224.66	188.75
L=0.15, $\lambda=50$																
Retailer-controlled substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	0.15	15	0.15	15	0.15	15	0.15	12	0.15	12	0.25	14	0.25	14	0.2	9
2	0.38	13	0.38	13	0.38	13	0.30	8	0.28	6	0.5	24	0.5	24	0.4	18
3	0.62	13	0.62	13	0.62	13	0.44	7	0.39	6	0.75	15	0.75	15	0.6	18
4	0.85	15	0.85	15	0.85	15	0.56	7	0.5	6				0.8	9	
5									0.7	8					0.66	13
6									0.85	12					0.83	6
7										0.85	12					0.7
EII	146.63	186.63	206.63	224.97	230.77						163.09	193.09	229.66	225.98	230.65	191.87
Static substitution																
items	b	q	b	q	b	q	b	q	b	q	b	q	b	q	b	q
1	-0.04	28	-0.04	7	-0.04	7	-0.04	7	-0.04	7	0.21	14	0.21	14	0.21	14
2	0.26		0.26	17	0.26	17	0.26	17	0.26	17	0.51	29	0.51	29	0.51	29
3	0.56		0.56	17	0.56	17	0.56	17	0.56	17	0.81	13	0.81	13	0.81	13
4	0.86	28	0.86	17	0.86	17	0.86	17	0.86	17						
EII	137.41	176.81	196.81	212.81	216.41						158.93	188.93	203.93	215.93	218.63	175.76

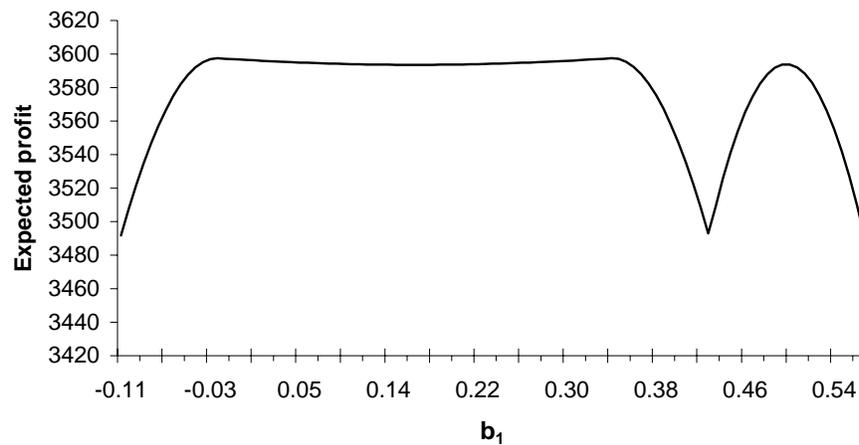
Notes : The table shows the locations (b) and inventory levels (q) of all items, and total expected profit (denoted EII) for each set of parameter values under both heuristics. The cost parameters are as shown in §5.3.

Table 3: Effect of c on optimal assortments, inventory levels, and expected profits under heuristics H1 and H2

	c=4		c=5		c=6	
	Retailer-controlled substitution					
Items	b	q	b	q	b	q
1	0.25	6	0.2	3	0.25	6
2	0.4	22	0.35	13	0.4	20
3	0.6	22	0.5	21	0.6	20
4	0.75	7	0.65	13	0.75	5
5			0.8	3		
E Π	287.67		232.59		180.47	
Optimality gap	2.26%		3.00%		4.62%	
	Static substitution					
Items	b	q	b	q	b	q
1	0.15	6	0.05	1	0.1	2
2	0.45	40	0.35	27	0.4	31
3	0.75	16	0.65	27	0.7	19
4			0.95	1		
E Π	280.74		222.36		168.92	
Optimality gap	2.41%		4.45%		6.40%	

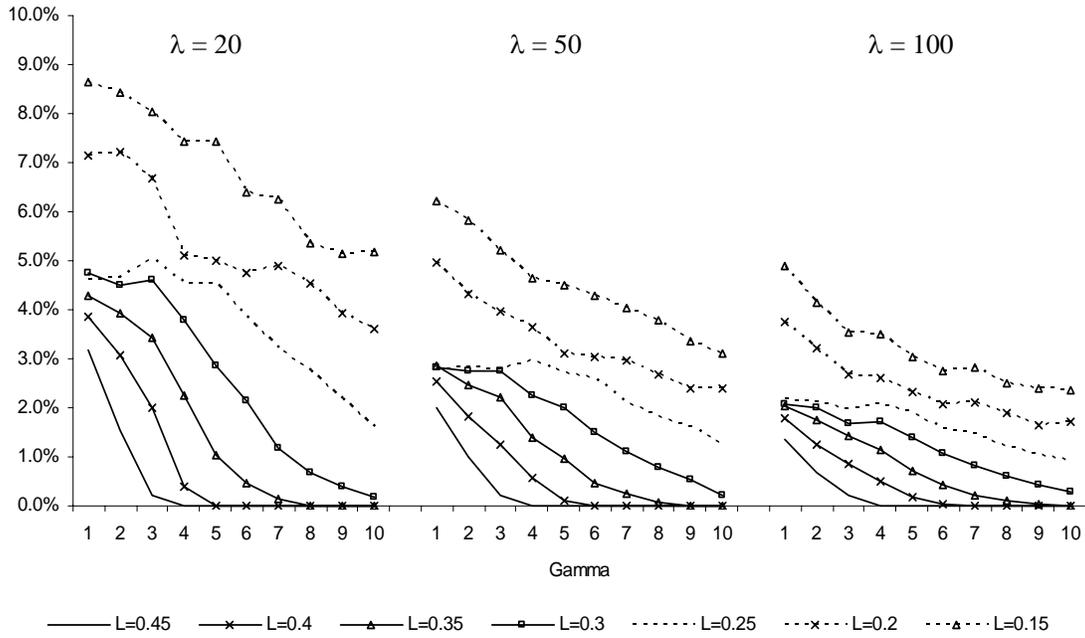
Note : Parameters other than c are set at $\lambda=50$, $\gamma=6$, $L=0.15$, $K=0.1$, $p=10$, $v=3$.

Figure 1: Expected profit under static substitution as a function of b_1 in $[\alpha, \alpha+2L]$



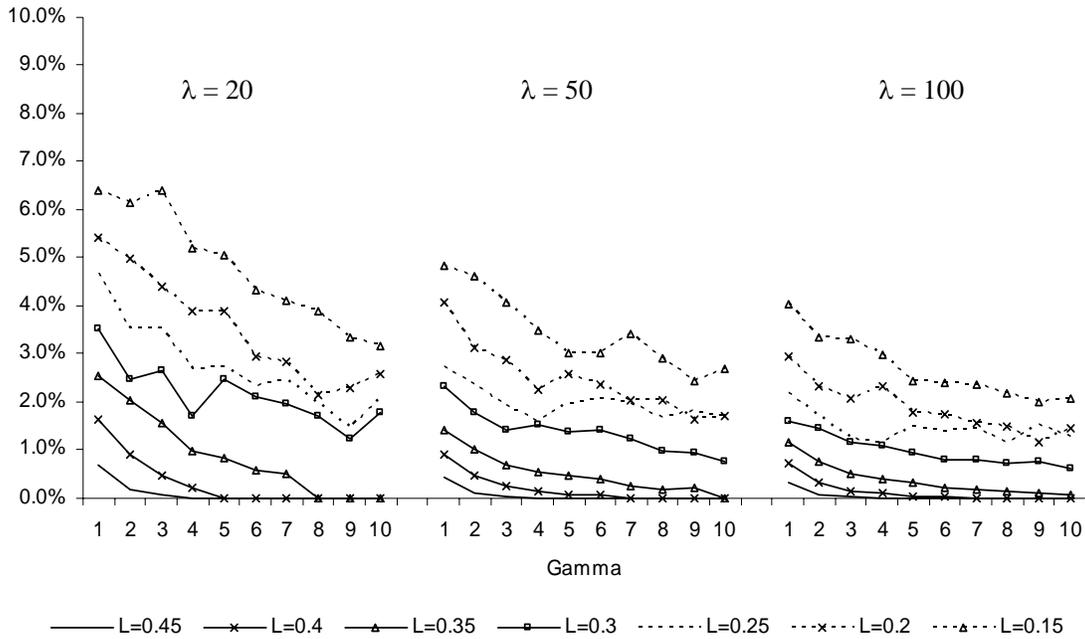
Notes: The parameter values are: $p=10$, $c=5$, $s=3$, $K=650$, $\lambda=1000$, $\gamma=2$, $L=0.34$. The value of α computed using these values is -0.11 .

Figure 2: Optimality gap for heuristic H1 for the dynamic substitution problem as a function of L , γ and λ



Notes: The y-axis shows the optimality gap defined as $[UB - H1]/UB$ in percentages. The parameters are set as $K=0.1$, $c=5$, and the remaining as specified in §5.3.

Figure 3: Optimality gap for heuristic H2 for the dynamic substitution problem as a function of L , γ and λ



Notes: The y-axis shows the optimality gap defined as $[UB - H2]/UB$ in percentages. The parameters are set as $K=0.1$, $c=5$, and the remaining as specified in §5.3.