Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Fixed vs. Random Proportions Demand Models for the Assortment Planning Problem Under Stockout-Based Substitution

Dorothée Honhon, Sridhar Seshadri,

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 15, No. 3, Summer 2013, pp. 378–386
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)
http://dx.doi.org/10.1287/msom.1120.0425

Fixed vs. Random Proportions Demand Models for the Assortment Planning Problem Under Stockout-Based Substitution

Dorothée Honhon
Operations Planning Accounting and Control, School of Industrial Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, d.b.l.p.honhon@tue.nl

Sridhar Seshadri
Department of Information, Risk, and Operations Management, McCombs School of Business, University of Texas at Austin, Austin, Texas 78712, sridhar.seshadri@mccombs.utexas.edu

We consider the problem of determining the optimal assortment of products to offer in a given product category when each customer is characterized by a type, which is a list of products he is willing to buy in decreasing order of preference. We assume consumer-driven, dynamic, stockout-based substitution and random proportions of each type. No efficient method to obtain the optimal solution for this problem is known to our knowledge. However, if the number of customers of each type is a fixed proportion of demand, there exists an efficient algorithm for solving for the optimal assortment. We show that the fixed proportions model gives an upper bound to the optimal expected profit for the random proportions model. This bound allows us to obtain a measure of the absolute performance of heuristic solutions. We also provide a bound for the component-wise absolute difference in expected sales between the two models, which is asymptotically tight as the inventory vector is made large, while keeping the number of products fixed. This result provides us with a lower bound to the optimal expected profit and a performance guarantee for the fixed proportions solution in the random proportions model.

Key words: assortment planning; inventory management; bounds; stockout-based substitution
History: Received: December 15, 2010; accepted: October 31, 2012. Published online in *Articles in Advance* March 1, 2013.

1. Introduction

An important problem in retailing is that of a firm choosing the assortment of differentiated products to offer when facing customers with heterogeneous tastes. Customers' tastes are characterized by their type, which is a list of products they are willing to buy in decreasing order of preference. The retailer has to decide which products to stock and how much inventory to carry of each of them. The products are offered at different selling prices and have different cost parameters for the retailer. Customers come to the store sequentially, observe the available inventory, and decide which product (if any) to purchase, based on their individual preferences. The customers who come to the store constitute a sample from the entire customer population, so that each customer has a certain probability of being of each type, independently of the other customers. The retailer's objective is to maximize expected profit knowing the distribution of the number of customers who visit the store in one period and the distribution of customer preferences in the population, but not the preferences of the specific customers who come to the store. This problem is known as the one-period assortment planning problem with *consumer-driven*, *dynamic*, *stockout-based* substitution with *random proportions* of customers of each type.

This is a complex and hard problem; see Goyal et al. (2012), who show that it is NP-hard. To our knowledge, no efficient method exists for solving this problem to optimality. Previous work on this problem has focused on providing heuristic methods (e.g., Pentico 1974, van Ryzin and Mahajan 1999, Smith and Agrawal 2000, Mahajan and van Ryzin 2001, Netessine and Rudi 2003). Our paper is most closely related to Honhon et al. (2012), who assume *fixed proportions* instead of *random proportions*, that is, they assume that the proportion of customers of each type who visit the store is fixed and matches the proportion in the entire customer population. They solve the hence modified problem to optimality using a dynamic program approach. For a broader review of the literature, including papers on estimation of demand with dynamic, stockout-based substitution and papers that assume static, assortment-based substitution, we refer to Kok et al. (2009).

Most of the papers do not provide performance bounds for the heuristics they propose (one exception is the work of Gaur and Honhon (2006), who obtain an upper bound based on retailer-controlled substitution for the locational choice model). They compare the *relative* performance of their heuristic to existing ones but do not provide guidance to distance from optimality. The goal of this paper is to provide an upper bound on the optimal expected profit (Proposition 1), which can be computed quickly (at least for up to $n = 16$ products) and used to calculate the distance from optimality of different heuristics, i.e., to obtain a measure of the *absolute* performance of the heuristics. Our upper bound is based on the assumption of *fixed proportions* introduced by Honhon et al. (2012). We show that the optimal expected profit under fixed proportions is always greater or equal to the optimal expected profit under random proportions. This upper bound can be easily computed using the dynamic program approach of Honhon et al. (2012). We also provide a theoretical bound on the component-wise absolute difference in expected sales between the fixed proportions and random proportions models (Lemma 3), which does not depend on the demand distribution or the parameters of the consumer choice model and is asymptotically tight as demand and inventory become larger, keeping the number of products fixed. We use this result to obtain a lower bound on the optimal expected profit in the random proportions model (Proposition 3). The percentage gap between the upper and lower bounds on the optimal expected profit is small when the optimal inventory vector in the fixed proportions model is large and decreases with mean demand and overage costs. Finally we also obtain a performance guarantee for the fixed proportions heuristic (Proposition 2). In our numerical study, we find that the average gap between the upper bound and the profit generated by the fixed proportions heuristic is only 0.72%. This gap is particularly small when the mean and variance of demand are large.

The rest of our paper is organized as follows. In §2, we briefly describe the one-period assortment planning problem with *consumer-driven*, *dynamic*, *stockout-based* substitution under both *random proportions* and *fixed proportions*. Our results are presented in §3. The proofs of these results, which are quite involved, are deferred to the online supplement (available at http://dx.doi.org/10.1287/msom.1120.0425). In §4, we present our numerical study. Finally, we conclude in §5.

2. Model

In this section, we briefly describe the random proportions (RP) and fixed proportions (FP) models and introduce the necessary notation. For a more detailed presentation of these models, we refer to Mahajan and van Ryzin (2001) and Honhon et al. (2012), respectively, for the RP and FP models.

We consider a product category consisting of n potential products, indexed 1 to n. Let $\mathcal{N} = \{1, \ldots, n\}$. The retailer determines the inventory of each of the n products at the beginning of the selling season or period. Let q_j be the inventory of product j at the start of the period. Let $\mathbf{q} = (q_1, \ldots, q_n) \in \mathbb{N}^n$ be the starting inventory vector. During the period, customers come to the store one after the other and progressively deplete the inventory (*dynamic substitution*). We assume that each customer buys at most one unit of the product. The retailer cannot replenish or modify the inventory during the period nor can she decide which product to allocate to each customer (*consumer-driven substitution*). Rather, an arriving customer selects a product from the set of products with positive inventory at the time of their visit to the store or leaves the store without purchasing (*stockout-based substitution*). Any inventory leftover at the end of the period is salvaged.

Customers are heterogeneous in their choice behavior: Each customer belongs to a *consumer type*, independently of the other customers. A consumer type $\boldsymbol{\tau}$ is a vector of products that a customer of that type is willing to purchase, arranged in decreasing order of preference. For example, a customer of type $(1, 2, 4)$ has product 1 as his first choice, product 2 as his second choice, product 4 as his third choice, and he never buys products 3 and 5 to n. Let $\mathcal{T}$ be the set of all possible types.

The retailer does not know the types of the incoming customers. However, she knows how customers make choices and knows the distribution of customer preferences. We assume that the type of a customer is independent of the types of other customers and independent of the total number of customers. Let $\alpha(\boldsymbol{\tau})$ denote the probability that a customer is of type $\boldsymbol{\tau}$, where $\sum_{\boldsymbol{\tau}\in\mathcal{T}} \alpha(\boldsymbol{\tau}) = 1$. Let D denote the number of customer arrivals during the selling season; D is a nonnegative integer-valued finite random variable bounded above by $\bar{D}$.

In the RP model, the number of customers who are of type $\boldsymbol{\tau}$ in a selling season is a random variable with a binomial distribution with D trials and a probability of success of $\alpha(\boldsymbol{\tau})$. In the FP model, the number of customers of type $\boldsymbol{\tau}$ who come to the store is *exactly* $D\alpha(\boldsymbol{\tau})$. Moreover, out of the first k customers who come to the store, there are *exactly* $k\alpha(\boldsymbol{\tau})$ customers of type $\boldsymbol{\tau}$ for all $0 \le k \le D$. In essence, demand is continuous and follows a fluid model: every unit of demand divides exactly into $\alpha(\boldsymbol{\tau})$ customers of type $\boldsymbol{\tau}$ for all possible types.

Let ω denote a generic sample path. In the RP model, a sample path is a sequence of customer arrivals along with their types. In the FP model, there is only one source of uncertainty, which is the demand D; therefore, a sample path is simply a value for D. The expectations are taken over all sample paths in both models accordingly.

The superscripts R and F are appended to the following set of variables to denote random and fixed proportions, respectively. Let $\mathbf{x}(k, \mathbf{q}; \omega)$ denote the leftover inventory vector on sample path ω after k customers have depleted the inventory when the starting inventory is $\mathbf{q}$. Let $\mathbf{y}(\mathbf{q}; \omega)$ denote the sales inventory vector on sample path ω when the starting inventory is $\mathbf{q}$ after all customers on sample path ω have arrived. The one-period profit for starting vector $\mathbf{q}$ on sample path ω is $\Pi(\mathbf{q}; \omega) = \sum_{j=1}^{n}(u_j + o_j) \cdot y_j(\mathbf{q}; \omega) - \sum_{j=1}^{n} o_j q_j$, where u_j and o_j are, respectively, the underage cost and overage cost of product j. The retailer's objective is to solve $\max_{\mathbf{q}\geq 0} \mathbb{E}[\Pi(\mathbf{q})]$. Let $\mathbf{q}^{*R}$ and $\mathbf{q}^{*F}$ denote the optimal inventory vector in the RP and FP models, respectively.

Mahajan and van Ryzin (2001) show that solving for $\mathbf{q}^{*R}$ is very difficult because the objective function is not quasi-concave in inventory levels. To the best of our knowledge, finding the optimal solution $\mathbf{q}^{*R}$ in an efficient way is still an open question. For small values of n and $\bar{D}$, one could resort to an exhaustive search over all integer-value inventory vectors such that $0 \leq q_j \leq \bar{D}$ for $j = 1, \ldots, n$, however this is too time consuming for most problems. Therefore, $\mathbf{q}^{*R}$ is generally unknown.

Honhon et al. (2012) provide an efficient dynamic programming-based algorithm to obtain $\mathbf{q}^{*F}$. This algorithm has a theoretical complexity of $O(8^n)$, but this number is based on a worst-case analysis of the number of breakpoints in the value functions (which is equal to $2^{3n-1} + 2^{2n-1}$). In practice, the number of breakpoints encountered is usually much lower so that the algorithm is very fast: we were able to obtain $\mathbf{q}^{*F}$ in less than 12 minutes for up to 16 products and with an average of approximately 65,000 breakpoints (for $n = 16$ the upper bound on the number of breakpoints is greater than 10^{13}!).

3. Results

In this section, we show that the FP model provides an upper bound on the optimal expected profit $\mathbb{E}[\Pi^R(\mathbf{q}^{*R})]$ in the RP model. This upper bound can be used to estimate optimality gaps when testing the absolute performance of heuristics.

First note that, for a given vector $\mathbf{q}$, it is possible that the expected profit under random proportions is greater than the expected profits under fixed proportions; that is, $\mathbb{E}[\Pi^F(\mathbf{q})] < \mathbb{E}[\Pi^R(\mathbf{q})]$, as the next example illustrates.

Example 1. Let $n = 2$. There are four possible types: $\mathcal{T} = \{(1), (2), (1, 2), (2, 1)\}$. All types are equally likely, that is, $\alpha(\boldsymbol{\tau}) = 1/4$ for all $\tau \in \mathcal{T}$. Let $\mathbf{q} = (2, 1)$. Demand is deterministic and $D = 2$. Let $u_1 = 10$, $u_2 = 1$, $o_1 = 1$, and $o_2 = 3$. We have $\mathbb{E}[\Pi^F(\mathbf{q})] = 10 < \mathbb{E}[\Pi^R(\mathbf{q})] = 10.375$.

However, we have the following result:

Lemma 1. *For every* $\mathbf{q} \geq 0$, *there exists* $\tilde{\mathbf{q}} \geq 0$ *such that* $\mathbb{E}[\Pi^F(\tilde{\mathbf{q}})] \geq \mathbb{E}[\Pi^R(\mathbf{q})]$.

We provide here a sketch of the proof, which is fairly involved. The general idea of the proof is that we are able to match the expected profit from a fixed quantity under random proportions with the expected profit from a random quantity under fixed proportions. Because there is no randomness with regard to the relative number of customers of each type in the fixed proportions model, we introduce randomness in the starting inventory vector in order to match the randomness in the random proportions model. The details are as follows. Given a vector $\mathbf{q}$ for the RP system, for each sample path $\bar{\omega}$ such as $D(\bar{\omega}) = \bar{D}$, we select a starting inventory vector to be used in the FP model in such a way that the two models have the same set of products with positive inventory for all values of demand and the same leftover inventory vector after $\bar{D}$ customers have come. The FP system is started randomly with these vectors with probability $P(\bar{\omega} \mid D(\bar{\omega}) = \bar{D})$, which add up to one. We show that the weighted average of the expected sales in the FP model with these starting inventory vectors matches the expected sales in the RP model and that the same holds for expected profit. As a result, there must exist an inventory vector such that expected profit in the FP model is higher than $\mathbb{E}[\Pi^R(\mathbf{q})]$. In the online supplement, we illustrate the proof technique with an numerical example.

Lemma 1 reveals a startling duality between the random proportions and fixed proportions models, namely, that the fixed proportions model is equivalent to the random proportions model under a properly chosen random starting vector. We conjecture that such an equivalence can be established even when inventory is replenished during a sales campaign.

This leads to our main proposition:

Proposition 1. $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})] \geq \mathbb{E}[\Pi^R(\mathbf{q}^{*R})]$.

Proof. Using Lemma 1 we can find $\tilde{\mathbf{q}}$ such that $\mathbb{E}[\Pi^F(\tilde{\mathbf{q}})] \geq \mathbb{E}[\Pi^R(\mathbf{q}^{*R})]$. The result follows by the optimality of $\mathbf{q}^{*F}$. Q.E.D.

Remember that, while $\mathbb{E}[\Pi^R(\mathbf{q}^{*R})]$ is generally unknown, $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})]$ can be computed as shown in Honhon et al. (2012). It follows that the performance of a heuristic solution $\mathbf{q}$ for the RP model can be evaluated with respect to $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})]$; that is, we can compute $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})] - \mathbb{E}[\Pi^R(\mathbf{q})]$, which is an upper bound on the optimality gap $\mathbb{E}[\Pi^R(\mathbf{q}^{*R})] - \mathbb{E}[\Pi^R(\mathbf{q})]$.

The next two results show that the FP model does not overestimate component-wise expected sales by too much, proving that the difference between $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})]$ and $\mathbb{E}[\Pi^R(\mathbf{q}^{*R})]$ is generally small. Lemma 2 provides a bound on the difference in sales of one product up to the first stockout epoch between the FP and RP models.

Lemma 2. *Given* $\mathbf{q} \geq 0$, *suppose that* $D = k \leq \min_{j=1,\ldots,n} q_j/\rho_j(\mathbf{q})$, *where* $\rho_j(\mathbf{q})$ *is the probability that a customer chooses product* j *given inventory vector* $\mathbf{q}$. *We have* $(\mathbb{E}[y_j^F(\mathbf{q}) \mid D = k] - \mathbb{E}[y_j^R(\mathbf{q}) \mid D = k])^+ \leq (\sqrt{k\rho_j}/\sqrt{2\pi})$ *for* $j = 1, \ldots, n$.

To prove Lemma 2, we consider the extreme case in which customers do not substitute in the event of a stockout in the RP model and maximize the component-wise difference in the expected sales between this case and the FP model. We use Lemma 2 to obtain a bound on the sum of component-wise difference in expected sales between the two models:

Lemma 3. *Given* $\mathbf{q} \geq 0$, $\sum_{j=1}^n (\mathbb{E}[y_j^F(\mathbf{q})] - \mathbb{E}[y_j^R(\mathbf{q})])^+ \leq (\sqrt{2}/\sqrt{\pi}) \sum_{j=1}^n \sqrt{jq_{[j]}}$, *where* $q_{[j]}$ *is the jth greatest value in* $q_1, \ldots, q_n$. *Also, if* $Q = \sum_{j=1}^n q_j$, *we have*

$$\frac{\sum_{j=1}^n (\mathbb{E}[y_j^F(\mathbf{q})] - \mathbb{E}[y_j^R(\mathbf{q})])^+}{Q} \leq \frac{\sqrt{n(n+1)}}{\sqrt{\pi}\sqrt{Q}}. \qquad (1)$$

In the proof of Lemma 3, we compare the FP and RP models for the same starting inventory $\mathbf{q}$. We use Lemma 2 to bound the difference in expected sales up to the first stockout epoch in the FP model. Then, we carefully adjust the leftover inventory vectors at that stockout epoch in the two models so that they match. When adjusting the leftover inventory in the RP model down, we take into account the fact that this decrease may increase the future sales of other products. Then, we use Lemma 2 to bound the difference in expected sales between the first and second stockout epochs in the FP model and repeat the adjustment process. We do so until the very last product runs out in the FP model.

From (1), we see that the bound is asymptotically tight as the inventory vector grows larger and the number of items in the assortment is fixed.

Example 2. For $n = 3$ and $Q = 300$, the right-hand side of (1) is 11.28%. If $Q = 3{,}000$, it is 3.57%.

We use Lemma 3 to obtain a theoretical bound on the performance of the $\mathbf{q}^{*F}$ heuristic in the RP model:

Proposition 2. *Let* $q_{[j]}^{*F}$ *be the jth greatest value in* $q_1^{*F}, \ldots, q_n^{*F}$, *then we have* $\mathbb{E}[\Pi^R(\mathbf{q}^{*F})] \geq \mathbb{E}[\Pi^F(\mathbf{q}^{*F})] - (\max_{j=1,\ldots,n}(u_j + o_j))((\sqrt{2}/\sqrt{\pi}) \sum_{j=1}^n \sqrt{jq_{[j]}^{*F}})$.

Also, we directly obtain a lower bound on the expected profit in the RP model:

Proposition 3. *Let* $q_{[j]}^{*F}$ *be the jth greatest value in* $q_1^{*F}, \ldots, q_n^{*F}$, *then we have* $\mathbb{E}[\Pi^R(\mathbf{q}^{*R})] \geq \mathbb{E}[\Pi^F(\mathbf{q}^{*F})] - (\max_{j=1,\ldots,n}(u_j + o_j))((\sqrt{2}/\sqrt{\pi}) \sum_{j=1}^n \sqrt{jq_{[j]}^{*F}})$.

The gap between this lower bound and the upper bound of Proposition 1, that is, $(\max_{j=1,\ldots,n}(u_j + o_j)) \cdot ((\sqrt{2}/\sqrt{\pi}) \sum_{j=1}^n \sqrt{jq_{[j]}^{*F}})$, increases with $\mathbf{q}^{*F}$ in absolute terms; however, it decreases as a proportion of the upper bound because of the square-root term. An increase in the optimal inventory vector can be caused either by an increase in mean demand or by an increase in the overage cost. In both cases, the gap becomes smaller in relative terms, as illustrated in the next section. In the case of an increase in mean demand, the explanation is that the coefficient of variation of the proportion of customers from each type, which is equal to $\sqrt{\mathbb{E}[1/D](1-\alpha(\tau))/(\alpha(\tau))}$, decreases with the mean μ, which leads to the RP model being closer to the FP model (because $1/D$ is a decreasing convex function, $\mathbb{E}[1/D]$ gets smaller as D becomes stochastically larger, or more generally, smaller in convex order). In the case of an increase in the overage cost, the higher inventory levels lead to fewer stockouts, so that the FP model becomes a better approximation of the RP model.

Note that the bounds we obtain are free of any assumption on the distribution of demand and the preference structure (i.e., on the $\alpha_{\boldsymbol{\tau}}$ for $\boldsymbol{\tau} \in \mathcal{T}$). In such, they apply to a very wide range of problems and settings, because most consumer choice models used in the operations literature are special cases of ours. It may be possible to improve the quality of the bounds by making them model specific and dependent on the demand distribution (e.g., in Lemma 3 one could multiply the upper bound on the expected difference in sales by the probability of the first run out, the second run out, etc.). Our current focus was to demonstrate the use of the fixed proportions model in all generality, so we decided not to do so and leave this extension for our future research plans.

4. Numerical Study

The objective of this section is threefold. First, we study how well the FP model approximates the expected sales obtained with the RP model and compare the FP model with another fluid-model type approximation, namely, the one suggested by Hopp and Xu (2008). Second, we study how the percentage gap between the upper bound from Proposition 1 and the lower bound from Proposition 3 varies with some key parameters of the model. Finally, we compare three heuristics in terms of speed and absolute performance using the upper bound from Proposition 1.

4.1. Study of the Quality of the FP Approximation

Hopp and Xu (2008) propose a static approximation of the assortment planning problem with consumer-driven, dynamic, stockout-based substitution that is based on a fluid network model and a one-to-one mapping of service levels and inventory quantities.

Table 1 Percentage Errors in Expected Sales of the Three Products for the HX and FP Models

	Scenario 1 (%) $\mathbf{v}=(1,1,1,1)$ $\mathbf{q}=(250,250,250)$		Scenario 2 (%) $\mathbf{v}=(1,1,1,1)$ $\mathbf{q}=(150,300,450)$		Scenario 3 (%) $\mathbf{v}=(1,1,2,3)$ $\mathbf{q}=(150,300,450)$		Scenario 4 (%) $\mathbf{v}=(1,1,2,3)$ $\mathbf{q}=(250,250,250)$	
κ	HX	FP	HX	FP	HX	FP	HX	FP
0.25	(1.1, 1.1, 1.1)	(0.9, 0.9, 0.9)	(0, 0.3, −0.1)	(0, 0.3, −0.1)	(1.1, 0.2, −0.3)	(1.1, 0.2, −0.3)	(1.7, 0, 0)	(1.3, 0, 0)
0.5	(1.1, 1, 1.1)	(0.7, 0.7, 0.7)	(0, 0.3, −0.2)	(0, 0.3, −0.2)	(1, 0, −0.2)	(1, 0, −0.2)	(1.7, 0, 0)	(0.9, 0, 0)
0.75	(1.1, 1.2, 1.1)	(0.6, 0.6, 0.6)	(0.1, 0.4, −0.2)	(0, 0.4, −0.2)	(1, 0.2, −0.1)	(1, 0.1, −0.1)	(2, 0.1, 0.1)	(0.8, 0, 0)
1	(1.2, 1.2, 1.2)	(0.5, 0.5, 0.5)	(0.1, 0.5, −0.3)	(0, 0.5, −0.3)	(0.9, 0.2, −0.1)	(0.9, 0.2, −0.1)	(2.2, 0.1, 0.1)	(0.6, 0, 0)
2	(1.8, 1.8, 1.9)	(0.3, 0.3, 0.4)	(0.4, 1, −0.5)	(0, 0.5, −0.2)	(1.4, 0.8, 0.6)	(0.7, 0.2, 0)	(3.7, 0.4, 0.4)	(0.3, 0, 0)
4	(3.5, 3.5, 3.5)	(0.2, 0.3, 0.2)	(1.7, 2.1, −1.2)	(0, 0.4, 0)	(3, 2.7, 2.5)	(0.5, 0.2, 0.1)	(6.9, 2, 1.7)	(0.3, 0, 0)
6	(5.5, 5.5, 5.5)	(0.2, 0.3, 0.3)	(4.2, 3.3, −1.8)	(0.1, 0.4, 0.2)	(5.1, 4.8, 4.7)	(0.5, 0.2, 0.1)	(9.1, 5.8, 4.3)	(0.3, 0.1, 0)
8	(7.4, 7.3, 7.4)	(0, −0.1, 0)	(7.8, 4.2, −2.7)	(0.1, −0.1, −0.4)	(6.9, 6.8, 6.7)	(0.1, −0.1, −0.2)	(9.1, 10.9, 8.8)	(−0.2, 0, 0)

In what follows, we refer to their model as the HX model. Like in the FP model, the HX model ignores the randomness in the number of customers from each type and assumes that the demand is continuous. But the main differences with the FP model are that (a) the HX model only applies to a consumer choice model that is an attraction-type model; and (b) in calculating the service level values corresponding to the inventory levels, the HX model assumes that the probability of switching from product i to product j following a stockout of product i is equal to the probability of choosing product j as a first choice. This simplification, which the authors refer to as the *memoryless flow assumption*, leads to an overestimation of demand because a proportion of unmet demand gets sent back to the original product requested.

Given that an attraction-based model like the multinomial logit (MNL) model is a special case of the ranking-based model we consider in this paper (in the sense that is possible to define $\alpha(\tau)$ for $\tau \in \mathcal{T}$, such that the choice probabilities of any assortment exactly match those obtained with the MNL model), we are able to directly compare the HX, FP, and RP models in terms of expected sales for a number of examples. We use the same parameters as in Example 2 from Hopp and Xu (2008): We assume that the number of customers visiting the store is normally distributed with mean $\mu = 1{,}000$ and standard deviation $\kappa\sqrt{\mu}$ and vary $\kappa \in \{0.25, 0.5, 0.75, 1, 2, 4, 6, 8\}$. We assume $n = 3$ and let $\mathbf{v} = (v_0, v_1, v_2, v_3)$ be the vector of nominal utilities from the MNL model, that is, such that the probability of picking product j from set S is given by $v_j/(v_0 + \sum_{i\in S} v_i)$. We fix the inventory level $\mathbf{q}$ and calculate the expected sales with the HX and FP models exactly. Let $\mathbf{y}^{\mathrm{HX}}$ denote sales in the HX model. Similarly to Hopp and Xu (2008), we use simulation to obtain expected sales under the RP model. For each problem instance we report the percentage error of the two approximations, calculated as $(\mathbb{E}[y_j^F(\mathbf{q})] - \mathbb{E}[y_j^R(\mathbf{q})])/\mathbb{E}[y_j^R(\mathbf{q})]$ and $(\mathbb{E}[y_j^{\mathrm{HX}}(\mathbf{q})] - \mathbb{E}[y_j^R(\mathbf{q})])/\mathbb{E}[y_j^R(\mathbf{q})]$ for $j = 1, 2, 3$, respectively, for the FP and HX models. Like Hopp and Xu (2008), we divide the problem instances into four scenarios; our Table 1 is constructed after Table 1 of Hopp and Xu (2008, p. 636). The percentage errors we obtain for the HX model do not exactly match those from Table 1 in Hopp and Xu (2008) because they are based on simulated data; however, they are always very close.

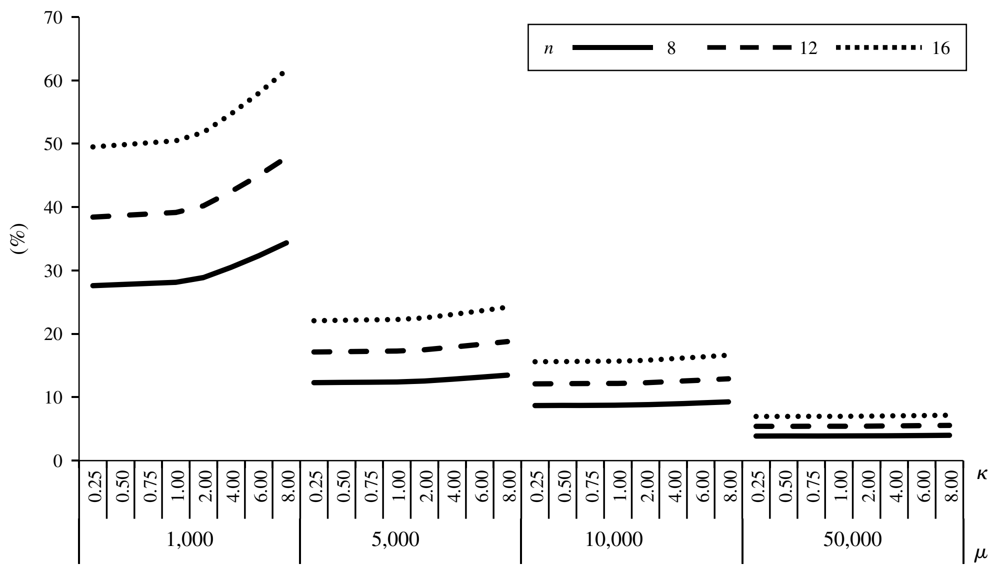
The main observation is that, although both approximations have a tendency to overestimate expected sales, the FP model is always more accurate. Further, the performance of the FP model improves as the variance of demand increases whereas that of the HX model deteriorates. Our interpretation of this seemingly counterintuitive result is as follows: When the number of customers coming to the store is more variable, situations with no product stocking out and situations with every product selling out become more likely. In these two extreme cases, sales under the FP model exactly match sales under the RP model. Differences between the two models occur when only a few products stockout and this is more likely to happen when the variance of demand is low. The results from Table 1 suggests that the FP model is a more accurate approximation of the RP model than the HX model, especially when the variance of demand is large. Our conclusion is that the extra simplification made by the HX model (i.e., the memoryless flow assumption) results in a more tractable formulation for expected sales but also in a deterioration of the quality of the approximation.

4.2. Study of the Gap Between Lower and Upper Bounds

Next we study the percentage gap between the lower bound and the upper bound, measured by

$$\frac{(\max_{j=1,\ldots,n}(u_j + o_j))((\sqrt{2}/\sqrt{\pi})\sum_{j=1}^{n}\sqrt{jq^{*F}_{[j]}})}{\mathbb{E}\,\Pi^F(\mathbf{q}^{*F})} \times 100.$$

Figure 1 **Percentage Gap Between Lower and Upper Bounds as a Function of n, μ, and κ**



We assume preferences are modeled by the MNL model with vector of nominal utilities $\mathbf{v} = (v_0, v_1, \ldots, v_n) = (5, 1, 2, 3, \ldots, n)$. We let $u_j = o_j = 5$ for $j = 1, \ldots, n$. We assume that demand has a normal distribution with mean μ and standard deviation $\kappa\sqrt{\mu}$. We vary the number of products $n \in \{8, 12, 16\}$, $\mu \in \{1{,}000, 5{,}000, 10{,}000, 50{,}000\}$, and $\kappa \in \{0.25, 0.5, 0.75, 1, 2, 4, 6, 8\}$.

In Figure 1, we see that the percentage gap between the lower and upper bound decreases with μ but increases slightly with κ and n. For $\mu = 50{,}000$ and $\kappa = 8$, the average percentage gap is equal to 3.29%. Remember that this percentage gap is an upper bound on the distance between the upper bound and the optimal expected profit and is based on a worst-case analysis that does not include the demand distribution and is based on a rebalancing of the inventory vector when a runout happens. Hence, these numbers are very encouraging and suggest that the upper bound is good especially for large problems.

In our numerical investigations, we also found that the gap between the lower and upper bound increases almost linearly with the maximum overage and underage cost values, which is as expected because the difference between the lower and upper bound is $(\max_{j=1,\ldots,n}(u_j + o_j))((\sqrt{2}/\sqrt{\pi})\sum_{j=1}^{n}\sqrt{jq^{*F}_{[j]}})$.

4.3. Study of the Performance of Heuristics

Next, we study the performance of $\mathbf{q}^{*F}$ as a heuristic for the random proportions model by comparing it to two previously studied heuristics, namely, the assortment-based substitution (ABS) heuristic and the sample path gradient algorithm (SPGA) from van Ryzin and Mahajan (1999). (The solution method suggested by Hopp and Xu (2008) does not apply here because we assume prices are fixed.) Such a comparison was also done by Honhon et al. (2012); however, because the authors were unaware of the results from this paper and because the optimal solution (i.e., $\mathbf{q}^{*R}$) was unknown in all the problem instances they considered, all they could do was compare the relative performance of the three heuristics. Thanks to Proposition 1, we are now able to provide an upper bound on the distance from optimality for each heuristic, that is, measure the absolute performance of the three heuristics.

The methodology used is as follows. We compute $\mathbf{q}^{*F}$ using the dynamic programming algorithm of Honhon et al. (2012), then we round the inventory values to the nearest integer. We refer to Honhon et al. (2012) and van Ryzin and Mahajan (1999) for an explanation of how to compute $\mathbf{q}^{A}$ and $\mathbf{q}^{S}$, which are, respectively, the solutions of the assortment-based substitution model and sample path gradient algorithm. For each heuristic solution $\mathbf{q}$, we estimate $\Pi^R(\mathbf{q}; \omega)$ and $\mathbb{E}[\Pi^F(\mathbf{q}^{*F})]$ by simulation, using 10,000 sample paths of random customer arrivals.

For each problem instance, we plot the percentage gap relative to the bound, computed as $(\mathbb{E}[\Pi^F(\mathbf{q}^{*F})] - \mathbb{E}[\Pi^R(\mathbf{q})])/\mathbb{E}[\Pi^F(\mathbf{q}^{*F})] \times 100$, where $\mathbf{q}$ is equal to $\mathbf{q}^{*F}$, $\mathbf{q}^{A}$, and $\mathbf{q}^{S}$. We present our results in four scenarios (which are inspired by those used in Honhon et al. 2012). We compute the standard error of the paired difference in profit (for each pair of heuristics) over the 10,000 replications. Based on this, we find that the difference in the average profit over the 10,000 replications is significant at the 1% level in 380 of 405 cases. Specifically, FP outperforms ABS 89 times, and ABS outperforms FP 25 times. In

21 experiments, their performance is statistically indistinguishable. FP outperforms SPGA 82 times and is outperformed 50 times by SPGA with three outcomes being indistinguishable. ABS outperforms SPGA 78 times and is outperformed 56 times by SPGA with four ties. Below, we elaborate upon when each algorithm performs the best. Because the average difference in profit is significantly different from zero in 94% of the cases, we report just the average profit over the replications and do not plot the confidence intervals.

In Scenario 1, we study the impact of the mean and variance of demand on the performance of the three heuristics. We set $n = 5$ and use a MNL model with vector of nominal utilities $\mathbf{v} = (v_0, v_1, \ldots, v_n) = (1, \ldots, 1)$. We let $u_j = 4.5$ for $j = 1, \ldots, 5$ and let $\mathbf{o} = (8.5, 7, 5.5, 4, 2.5)$. We assume that demand has a normal distribution with mean μ and standard deviation $\kappa\sqrt{\mu}$ and vary $\mu \in \{1{,}000, 2{,}000, \ldots, 5{,}000\}$ and $\kappa \in \{0.25, 0.5, 0.75, 1, 2, 4, 6, 8\}$. In Figure 2, we see that overall, the FP heuristic performs the best, beating the SPGA heuristic always, and the ABS heuristic for high values of κ. We also see that the FP heuristic performs better as μ and κ increase. The finding with regard to κ is consistent with our finding that the FP model becomes a more accurate approximation to the RP model when the variance of demand increase. The improvement with μ can be explained by the fact that the coefficient of variation of the proportion of customers of a given type, which is equal to $\sqrt{\mathbb{E}[1/D]((1-\alpha(\tau))/\alpha(\tau))}$, is decreasing in μ (given the demand distribution we use), so that the FP model tends to mimic the RP model as mean demand increases. On the other hand, the ABS heuristic performs worse as κ increases. To save on inventory costs, the ABS heuristic stocks fewer products when κ is really high but this results in fewer opportunities for stockout-based substitutions, which affects sales and expected profits. The deterioration of the performance of the SPGA as μ increases can be explained by the fact that the algorithm needs more iterations to converge to a good solution when the problem size is larger.

In Scenario 2, we study the impact of the number of products on the performance of the three heuristics. We use a MNL model with vector of nominal utilities $\mathbf{v} = (v_0, v_1, \ldots, v_n) = (1, \ldots, 1)$. We let $u_j = 2 + 0.5j$ and $o_j = 7 - 0.5j$ for $j = 1, \ldots, n$. We assume that demand has a normal distribution with mean and variance 1,000. We vary $n \in \{5, 6, \ldots, 10\}$. In Figure 3, we see that the FP heuristic performs better as n increases. This can again be explained by the fact that the coefficient of variation of the proportion of customers of a given type decreases as n increases because $\alpha(\tau)$ decreases: the average value of $\alpha(\tau)$ goes from 41/10,000 when $n = 5$ to 92/100,000,000 when $n = 10$. The SPGA tends to perform worse as the number of products increases because it takes longer for the algorithm to converge to a solution.

In Scenario 3, we study the impact of the overage cost (value and asymmetry) on the performance of the three heuristics. We set $n = 5$ and use a MNL model with vector of nominal utilities $\mathbf{v} = (v_0, v_1, \ldots, v_n) = (1, \ldots, 1)$. We assume that demand has a normal distribution with mean and variance 1,000. We let $u_j = 4.5$ for $j = 1, \ldots, 5$ and vary $\mathbf{o} = (o_1, \ldots, o_n)$ in two ways. First, we let $o_j = k$ for $j = 1, \ldots, n$ and vary $k \in \{0.5, 1.5, \ldots, 35.5\}$. This has the effect of varying the sum of overage and underage costs from 5 to 40 and the critical fractiles from 0.11 to 0.9, while keeping the product category homogeneous in terms of costs and popularity. Figure 4(a) shows the average

Figure 2 **Percentage Gaps with Respect to Upper Bound as a Function of μ and κ (Scenario 1)**

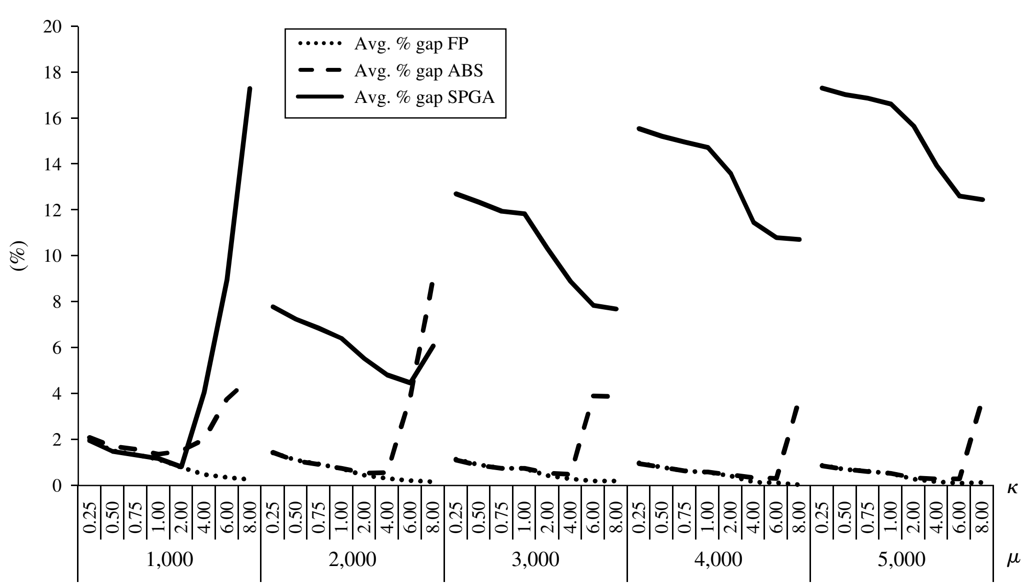


Figure 3 **Percentage Gaps with Respect to Upper Bound as a Function of the Number of Products (Scenario 2)**

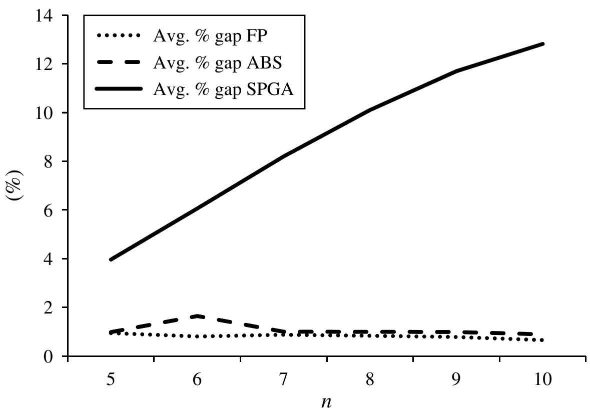
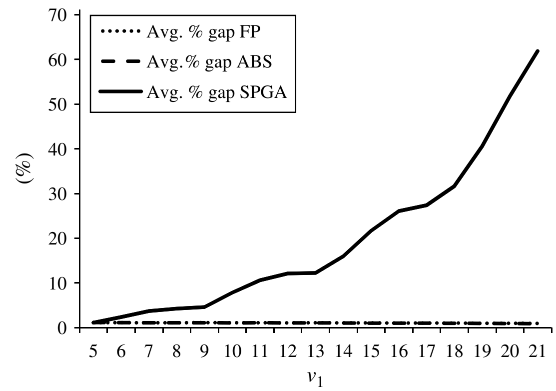


Figure 5 **Percentage Gaps with Respect to Upper Bound as a Function of the Nominal Utilities (Scenario 4)**

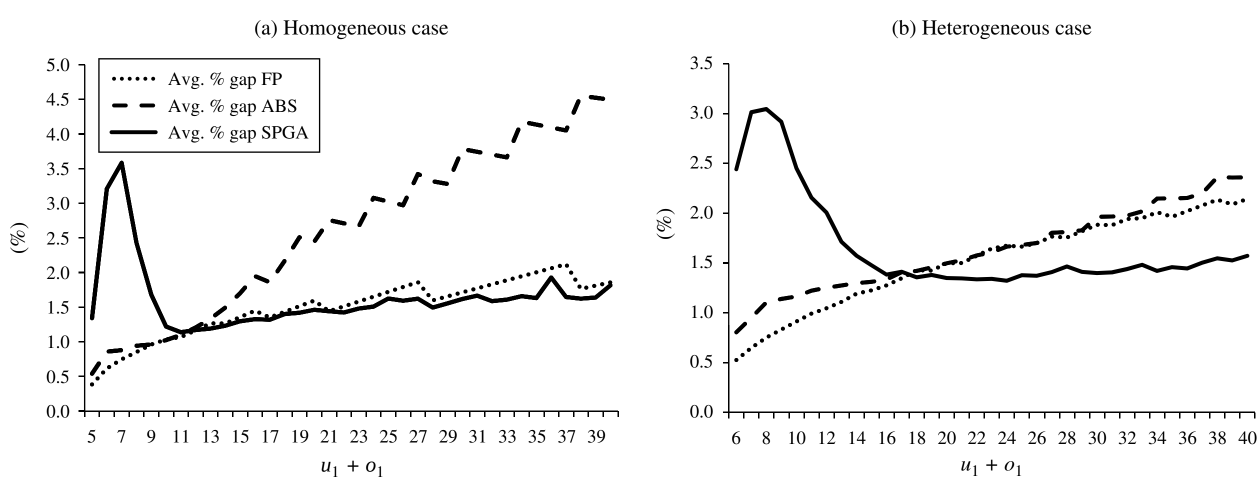


percentage gaps of all three heuristics as a function of $u_1 + o_1$. We see that the performance of the ABS heuristic deteriorates as the overage cost increases. A comparison of the inventory vectors reveals that the ABS tends to understock the products because it ignores the benefits of dynamic substitution. The SPGA initially performs badly for very high critical fractiles, then it becomes the best of all three heuristics. The gap of the FP heuristic is increasing in the overage cost but always remains low.

Second, we let $\mathbf{o} = (0.5+4k, 0.5+3k, 0.5+2k, 0.5+k, 0.5)$ and vary $k \in \{0, 0.25, \ldots, 8.75\}$. This has the effect of varying the maximum sum of overage and underage costs between 5 and 40 and the critical fractiles from 0.11 to 0.9, while progressively making the product cost parameters more and more asymmetric. Figure 4(b) shows the average percentage gaps of all three heuristics as a function of $u_1 + o_1$. We see that performance of the three heuristics is similar as in the previous case: The SPGA initially performs badly, then it becomes the best heuristic, and the ABS and FP heuristics perform worse as the cost parameters become more asymmetric. Note that Mahajan and van Ryzin (2001) also remarked that the ABS heuristic had worse performance when variants have highly different profit margins.

In Scenario 4, we study the impact of the heterogeneity in consumer preferences on the performance of the three heuristics. We set $n = 5$ and assume that demand has a normal distribution with mean and variance 1,000. We let $u_j = 4.5$ for $j = 1, \ldots, 5$ and $\mathbf{o} = (8.5, 7, 5.5, 4, 2.5)$. We use a MNL model with vector of nominal utilities $\mathbf{v}$, which we vary as follows. We start initially from $(v_1, \ldots, v_n) = (5, \ldots, 5)$, then progressively shift the popularity to the first product, creating heterogeneity in customer preferences; the next vectors are $(6, 5, 5, 5, 4)$, followed by $(7, 5, 5, 4, 4)$ etc. until $(21, 1, 1, 1, 1)$. In all cases, we use $v_0 = 5$. Figure 5 shows the percentage gaps of the three heuristics as a function of v_1. We see that the ABS and FP heuristics are very close to one another and they both perform much better than the SPGA heuristic, which cannot handle very heterogenous product categories in terms of popularity. This

Figure 4 **Percentage Gaps with Respect to Upper Bound as a Function of the Overage Cost (Scenario 3)**



is because the SPGA starts with an initial solution where all products are stocked in equal quantities, so it takes longer for the algorithm to converge when the optimal inventory vector is highly asymmetric.

We now discuss the computation (CPU) time required to obtain the solution under the three heuristics. We varied $\mu \in \{1{,}000, 2{,}000, \ldots, 5{,}000\}$ and $n \in \{10, 11, \ldots, 16\}$ (we used a Lenovo Thinkpad X201T series laptop with Intel Core i7 CPU L 620 @ 2.00 GHz and 4.00 GB of RAM). Our results indicate that the ABS heuristic is the fastest, followed by the FP heuristic, followed by the SPGA. The CPU time of the SPGA increases rapidly with mean demand (from about 6 minutes with $\mu = 1{,}000$ to about 30 minutes with $\mu = 5{,}000$), but this is not the case for the other two heuristics, for which n is the main factor determining the speed. The CPU time of the ABS heuristic was under 1 minute in all the cases we considered. The FP heuristic is quick for up to 16 products (less than 12 minutes).

Across all four scenarios, the average percentage gaps relative to the upper bound of each heuristic were equal to 0.72%, 1.79%, and 9.10%, respectively, for the FP, ABS, and SPGA heuristics, which suggests that the FP heuristic performs the best of all three, on average. The ABS heuristic does not perform well when demand has a high variance and overage costs are high or highly asymmetric. The SPGA is not to be used when the product critical fractiles are high and products are heterogenous in terms of popularity. Furthermore, the SPGA becomes impractical with large values of demand because it requires a very large number of iterations to converge to a good solution. For lower values of n (up to 16), when the computational time of the FP heuristic is not significantly larger than that of the ABS heuristic, we recommend the use of the FP heuristic in most cases, except if the overage costs are highly asymmetric and mean demand is low (in which case the SPGA performs better) or if demand has low variance and the cost parameters are low and not too asymmetric (in which case the ABS performs better). Overall, the FP heuristic performs extremely well, especially when n is small and the mean and variance of demand are large, which is the case for high-volume product categories with a relatively small assortment breadth, such as milk or sodas.

5. Conclusion

To our knowledge, our paper is the first to provide a good and computable upper bound on the optimal expected profit for the one-period assortment planning problem with consumer-driven, dynamic, stockout-based substitution with random proportions of customers of each type and a very general model of consumer choice. This upper bound can be used to measure the absolute performance of heuristics in the random proportions model. We show analytically and numerically that the gap between this upper bound and the optimal expected profit is generally small.

Also, we provide a performance guarantee for the fixed proportions heuristic in the random proportions model and further numerical evidence of its good performance.

We hope that our work will be of use to researchers who want to propose new heuristic solutions for this problem and want to test the performance of their method. Also, our proof techniques for establishing the bounds might be of independent interest to researchers who use fluid approximations to dynamic stochastic inventory systems.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.1120.0425.

Acknowledgments

The authors thank the editor-in-chief, the associate editor, and the anonymous reviewers for several comments that helped improve the paper. The authors also thank seminar participants at the McGill Business School, the University of Texas at Dallas, the Department of Industrial Engineering and Operations Research of the University of California at Berkeley, and the Johnson Graduate School of Business of Cornell University for the valuable feedback.

References

Gaur V, Honhon D (2006) Product variety and inventory decisions under a locational consumer choice model. *Management Sci.* 52(10):1528–1543.

Goyal V, Levi R, Segev D (2012) Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. Working paper, Massachusetts Institute of Technology, Cambridge.

Honhon D, Gaur V, Seshadri S (2012) Assortment planning and inventory decisions under stockout-based substitution. *Oper. Res.* 58(5):1364–1379.

Hopp WJ, Xu X (2008) A static approximation for dynamic demand substitution with applications in a competitive market. *Oper. Res.* 56(3):630–645.

Kok AG, Fisher ML, Vaidyanathan R (2009) Assortment planning: Review of literature and industry practice. Agrawal N, Smith SA, eds. *Retail Supply Chain Management: Quantitative Models and Empirical Studies* (Springer, New York), 99–154.

Mahajan S, van Ryzin G (2001) Stock retail assortments under dynamic consumer substitution. *Oper. Res.* 49(3):334–351.

Netessine S, Rudi N (2003) Centralized and competitive inventory models with demand substitution. *Oper. Res.* 51(2):329–335.

Pentico DW (1974) The assortment problem with probabilistic demands. *Management Sci.* 21(3):286–290.

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. *Oper. Res.* 48(1):50–64.

van Ryzin G, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. *Management Sci.* 45(11):1496–1509.