M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 7, No. 4, Fall 2005, pp. 330–346
ISSN 1523-4614 | EISSN 1526-5498 | 05 | 0704 | 0330

informs®

DOI 10.1287/msom.1050.0088
© 2005 INFORMS

Retail Assortment Planning in the Presence of Consumer Search

Gérard P. Cachon, Christian Terwiesch
Operations and Information Management, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104 {cachon@wharton.upenn.edu, terwiesch@wharton.upenn.edu}

Yi Xu
School of Business Administration, University of Miami, Coral Gables, Florida 33124-6524, yxu@miami.edu

Consumers often know what kind of product they wish to purchase, but do not know which specific variant best fits their needs. As a result, a consumer may find an acceptable product in one retailer but nevertheless purchase nothing, opting to search other retailers for an even better product. We study several models of retail assortment planning, some of which explicitly account for consumer search and one that does not, which we call the "no-search" model. Even though the no-search model never includes an unprofitable variant in the assortment, in the presence of consumer search, it may indeed be optimal to include an unprofitable variant. Furthermore, we find that the no-search model can lead to an assortment with an expected total profit that is significantly less than optimal. In the extreme, the no-search model may recommend closing down a category (i.e., carry no variants) even if a profitable assortment exists (a 100% profit loss). We conclude that failing to incorporate consumer search into an assortment planning process may cause a retailer to underestimate the substantial value a broad assortment has in preventing consumer search. We discuss how the insights from our stylized models may apply to actual assortment planning processes.

Key words: product proliferation; multinomial logit model; assortment optimization
History: Received: August 25, 2004; accepted: June 8, 2005. This paper was with the authors 3 months for 2 revisions.

1. Introduction

Consumers often know what kind of product they wish to purchase, but do not know which specific variant best fits their needs. Consider, for example, a consumer shopping for a new digital camera. Upon inspecting the cameras in the assortment of a retail store, the consumer might be able to assess the utility associated with each of the present cameras, yet would face uncertainty about those outside the store's assortment. Hence, even if she finds an acceptable camera in the current store, exceeding her utility associated with not buying a camera at all, she may nevertheless continue her search at other retailers. This paper studies whether it is important for an assortment planning process to explicitly account for consumer search.

It is well documented that in most retail categories, there is a considerable growth in the number of products available in the market (e.g., Pashigian 1988, Bayus and Putsis 1999), which raises the issue of whether the desire to please all consumers may have led to excessive product proliferation, i.e., assortments with more variants than optimal (Quelch and Kenny 1994). Although assortment planning in practice requires a blend of art (i.e., intuition) and science (i.e., analytical models), some argue that pushing the balance further in the direction of science could help to control the problem of unproductive variety (Fisher et al. 2000). Not surprising, the academic literature includes a growing number of models that balance the trade-off between the expanded revenue of a deeper assortment and its higher operational costs, e.g., Aydin and Hausman (2002), Chong et al. (2001), Kok and Fisher (2004), Mahajan and van Ryzin (2001), Smith (2002), Smith and Agrawal (2000), and van Ryzin and Mahajan (1999). All of those papers begin with the presumption that a retailer has a consumer choice model with known parameters and then,

assuming the consumer choice model accurately represents actual consumer choices, these papers develop methods for choosing an assortment that optimizes expected profit.

Our work is similar to the current literature in that we also develop consumer choice models and methods for finding assortments that optimally balance revenue expansion with operational costs. We differ in two respects. The current literature models search implicitly, but we model it explicitly. An implicit model of search includes an option for the consumer to purchase nothing from the assortment, which can be either because the consumer chooses to indeed never make a purchase in the category or because the consumer chooses to search other retailers for a better variant. Furthermore, in the implicit model, the value of the no-purchase option to the consumer is assumed to be independent of the other items in the assortment, whereas in an explicit model of search, a consumer's expectations and the retailer's assortment both influence the search decision. For example, the value of search to a consumer may decrease as a retailer broadens its assortment because the consumer then expects to find fewer new variants at other retailers. Indeed, if the retailer includes all possible variants in its assortment, then there is absolutely no value for a consumer to search other retailers for a better product (i.e., the consumer is already aware of every possible product by inspecting the retailer's assortment). This benefit of a broad assortment is not captured in models with implicit search.[1]

A second departure from previous work is that we do not presume a retailer knows the parameters of the actual consumer choice model, i.e., a consumer choice model in practice must be estimated from data, data that could depend on the assortment decisions of the retailer.[2] This issue is relevant in our context precisely because, as we just discussed, consumer choice is influenced by assortment decisions and search, i.e., parameters estimated with a narrow assortment may not match the parameters estimated with a broad assortment because consumer expectations between those two assortments are not the same, thereby causing differences in their search behavior. Hence, although an assortment may be optimal according to the consumer choice model estimated with data from that assortment, it may not be the optimal assortment (because the estimated choice model does not describe actual consumer choices with the true optimal assortment).[3] Furthermore, because we develop consumer choice models with search, we are able to explore the robustness of an assortment planning model that does not explicitly include search.

The remainder of this paper is organized as follows. The next section describes our three models of assortment planning. Section 3 studies the structure of the optimal assortment. Section 4 evaluates the performance of the assortment planning model that does not explicitly account for search. The final section concludes.

2. Model

A risk-neutral retailer sells a product with n possible variants to risk-neutral consumers. Let $N = \{1, 2, \ldots, n\}$ be the set of possible variants and let S be the subset of variants in the retailer's assortment. In addition to the actual product variants, we create variant 0, a faux variant, to represent the no-purchase option, i.e., a consumer who chooses variant 0 does not purchase any variant.

We use the multinomial logit (MNL) to model consumers' utilities across variants. It is an intuitive, frequently used, and successfully applied consumer choice model (see Anderson et al. 1992, Mahajan and van Ryzin 1998 for a review of the assumptions, limitations, characteristics, and properties of the MNL

[1] There is an extensive literature in economics that considers consumer search among retailers over price, e.g., Stigler (1961), Salop and Stiglitz (1977), and Stahl (1989). In our model consumers do not search for a lower price, but rather for a better product. Anderson and Renault (1999), Stahl (1982), and Wolinsky (1983, 1984) incorporate consumer search for a better product, but they assume each retailer carries only a single product. Weitzman (1979) and Morgan (1983) study optimal consumer search strategies, but do not investigate how consumer search influences retail assortment planning.

[2] Fisher and Rajaram (2000) study a model for merchandise testing to calibrate a consumer choice model (i.e., a sales forecast for each store and each product), but they do not consider which products to stock, nor do they account for consumer search.

[3] Cachon and Kok (2005), Cooper et al. (2005), and Armony and Plambeck (2005) also study the interaction between actions taken (such as which variants to include in an assortment) and the observed data used to estimate model input parameters, but in significantly different contexts.

model). Let U_i be a consumer's utility from variant $i \in [0, N]$: $U_i = (u_i - p_i) + \zeta_i$, where u_i is a constant, identical across consumers, p_i is the market price of variant i ($p_0 = 0$), and the random variable ζ_i has a zero mean Gumbel distribution. We refer to $(u_i - p_i)$ as variant i's expected net utility. We label the variants in decreasing net utility order. Let $F(x)$ be the distribution function of ζ_i,

$$F(x) = \exp[-\exp(-(x/\mu + \gamma))],$$

where γ is Euler's constant and μ is a scale parameter. Let $f(x)$ be its density function. The realizations of ζ_i are independent across consumers, so ζ_i creates consumer heterogeneity, i.e., while each consumer has the same expected net utility for each variant, their realized utilities differ.

Let $q_i(S)$ be the probability a consumer chooses variant i in the assortment S. Without loss of generality, the consumer population is normalized to one, so $q_i(S)$ is also variant i's demand. As in van Ryzin and Mahajan (1999), $q_i(S)$ does not depend on the inventory status of the variants nor do consumers substitute if their preferred variant is out of stock. However, $q_i(S)$ does depend on the retailer's assortment and consumers' expectations regarding search, as detailed in §2.1.

Define m_i as the net profit margin of variant i (price minus purchase cost). To obtain a simple structure for the optimal assortment, we assume the profit margins are identical across all variants, $m_i = m_j$, but our results are easily extended if margins are (weakly) increasing in net utility, i.e., $m_j \geq m_k$ if $u_j - p_j \geq u_k - p_k$. In the numerical study we consider the case in which margins are decreasing in net utility. Let $c(q_i(S))$ be the operational costs associated with including variant i in the assortment, i.e., the shelf space, holding, handling, and transportation costs of stocking variant i. To reflect economies of scale in operational costs, such as is common in the economic order quantity (EOQ) or newsvendor models, we assume $c(\cdot)$ is concave and increasing. (Note that because the demand rate has been normalized to one, the cost function should reflect this normalization.)

The expected profit of a variant $i \in S$ is

$$\pi_i(S) = m_i q_i(S) - c(q_i(S)).$$

The retailer's objective is to choose an assortment S to maximize expected profit

$$\max_{S \subseteq N} \pi(S) = \sum_{i \in S} \pi_i(S). \tag{1}$$

2.1. Evaluating Demand

We consider three models for consumer search. The first is the no-search model. In the no-search model, a consumer purchases a variant in the retailer's assortment if $U_{\max} = \max_{i \in S} U_i$ is greater than (or equal to) the no-purchase utility, U_0; otherwise the consumer purchases nothing. In this model, the no-purchase variant represents the fixed value of consumer search, i.e., from the retailer's perspective, it is not clear if the consumer leaves to never make a purchase or leaves to search other retailers to make some purchase. Let $q_i^m(S)$ be variant i's share of demand in this model, where the superscript "m" is used to denote the traditional MNL without explicit search. It is well known that (Anderson et al. 1992, ch. 2)

$$q_i^m(S) = \frac{v_i}{\sum_{j \in S} v_j + v_0} \quad \text{for } i = 0 \text{ and } i \in S, \tag{2}$$

where $v_j = \exp((u_j - p_j)/\mu)$ is referred to variant j's "preference."

From (2), in the no-search model, there exists a cannibalization effect: variant i's demand decreases as the assortment "expands," where expands means either that additional variants are added to the assortment or that the new assortment's total preference is larger. To be specific, we say that assortment S^+ is "deeper" than S (or expands upon S) if $\sum_{k \in S} v_k < \sum_{k \in S^+} v_k$. Clearly, S^+ is deeper than S if $S \subset S^+$, but S^+ can be a deeper assortment even with fewer variants as long as some of the variants in S^+ are not in S. (But the comparison between S and S^+ makes sense relative to variant i only if variant i is in both assortments.)

Now consider the two search models. In each one, a consumer still does not purchase if the no-purchase variant has the highest realized utility. However, the consumer may not purchase even if there is an acceptable product in the retailer's assortment (i.e., a variant with higher utility than the no-purchase utility), because the consumer may choose to continue her search (at some other retailer) for an even better variant. Whether a consumer chooses to search depends on a number of factors: the realized utility of the

best variant at the retailer, the cost of search, and the consumer's expectation on the value of search outside the retailer's assortment. It is the latter that differentiates our two search models. Roughly speaking, with the *independent assortment* search model a consumer expects that the retailer's assortment is unique, i.e., if the search option is chosen, then the consumer expects to observe different variants by searching. It follows that the consumer's expected value from search is independent of the retailer's assortment (because search yields different variants no matter which variants the retailer carries). Examples that might reasonably fit this setting include antique dealers and jewelry stores. In the *overlapping assortment* model, there is a limited number of products available in the market, as in the digital camera example. As a result, expanding the retailer's assortment reduces the value of search because search leads to fewer new variants.

Now consider the details of these models. With the independent assortment search model, a consumer expects to receive utility $U_r = u_r + \zeta_r$ if the consumer chooses to search, where u_r is a constant common to all consumers and ζ_r is a zero mean Gumbel random variable. (Note that this model is equivalent to a model that assumes search gives a consumer the possibility to purchase from another set of products in which the utilities of those variants are also the realizations of Gumbel random variables, because the Gumbel is closed under maximization.) Let b be a consumer's cost if the search option is chosen. Therefore, the consumer choice process is as follows: a consumer observes the realizations of the utility of the S variants in the retailer's assortment and the no-purchase utility; the consumer surely does not purchase a variant from the retailer if $U_{\max} < U_0$; if $U_{\max} > U_0$, then the consumer either purchases her most preferred variant from the retailer or the consumer chooses to search (thereby incurring the search cost b). From the retailer's perspective, the no-purchase variant is equivalent to the search option: in either case, the consumer does not purchase from the retailer. From the consumer's perspective, they are different: with the search option the consumer decides to forgo an acceptable variant ($U_{\max} > U_0$) for the chance of earning an even higher utility with search. ($U_{\max}$ is observed by the consumer when choosing whether to search, but the realization of U_r is not yet observed, so a realization $U_r < U_{\max}$ is possible.) The next theorem gives each variant's share of demand, $q_i^{si}(S)$, where the superscript "si" is used to denote the search model with independent assortment. All proofs can be found in the appendix.

Theorem 1. *The demand for variant i in the independent assortment search model is*

$$q_i^{si}(S) = q_i^m(S)(1 - H(\bar{U}, S)) \quad \textit{for } i \in S \tag{3}$$

and

$$q_0^{si}(S) = 1 - \sum_{i \in S} q_i^{si}(S),$$

where $\bar{U} = u_r - b$, $H(\bar{U}, S) = \exp(-\lambda(v_0 + \sum_{j \in S} v_j))$ *and* $\lambda = \exp[-(\bar{U}/\mu + \gamma)]$.

In the independent assortment search model, a consumer searches only if $U_{\max}$ (the best utility among the retailer's assortment) is less than the search threshold $\bar{U}$, which is independent of the retailer's assortment. (The value of search does not depend on the retailer's assortment because all new variants are observed if search is chosen in the independent assortment search model.) As a result, according to Theorem 1, the demand for variant i in this model is a fixed fraction, $(1 - H(\bar{U}, S))$, of the demand for variant i in the no-search model, $q_i^m(S)$. Note that fraction depends on the chosen assortment even though the search threshold $\bar{U}$ is fixed.

Unlike in the no-search model, it is not clear that the cannibalization effect always exists in the independent assortment search model: while $q_i^m(S)$ decreases as the assortment expands, the search adjustment factor, $(1 - H(\bar{U}, S))$, increases. According to Theorem 2, the former dominates the latter, i.e., the cannibalization effect does exist in the independent assortment search model.

Theorem 2. *For all S and S^+ such that $\sum_{k \in S} v_k < \sum_{k \in S^+} v_k$, $q_i^{si}(S) > q_i^{si}(S^+)$, $\forall\, i \in \{S \cap S^+\}$ in the independent assortment search model, variant i's demand is lower in a deeper assortment.*

In the overlapping assortment search model, the market contains a limited number of potential variants, e.g., digital cameras. As a result, the value of search to a consumer may very well depend

on the retailer's assortment: as the retailer deepens his assortment, a consumer may lower her expected value of search because search provides value only if new variants are discovered. To be specific, define $\bar{S}$ to be the variants outside the retailer's assortment, $\bar{S} = N - S$. The consumer choice process is now as follows: The consumer observes the no-purchase utility, U_0, and the utility for each variant in the retailer's assortment; the consumer surely does not purchase if $U_0 > U_{\max}$; if $U_0 < U_{\max}$, then the consumer either purchases the highest utility variant in S or chooses to search. If search is chosen, the consumer incurs the cost b but then has the opportunity to purchase the best variant in the entire set N, i.e., the consumer finds another retailer (or a combination of retailers) that has an assortment, which includes both S and $\bar{S}$, and the consumer's realized preferences for variants in set S remain the same (i.e., there is not a new random draw of utilities for the variants in S). The latter implies that a consumer's utility for a variant does not depend on where the consumer purchases the item. Furthermore, in this model, the retailer is effectively competing against some full assortment retailer with the same prices as our retailer. (If outside prices differed from the retailer's price, then to approximately account for this difference, adjust the search cost parameter b: lower outside prices reduce b and higher outside prices increase b.)

To the extent that the independent assortment search model represents a worst case for the search option (the consumer can end up with a less desirable variant if search is chosen), the overlapping assortment search model represents the best case for the search option (the consumer can only find a more desirable product by searching). We suspect other search models are qualitatively a mixture of these two extremes, and they are analytically more cumbersome.[4] Theorem 3 is the counterpart to Theorem 1, where the superscript "*so*" is used to denote the search model with overlapping assortment.

Theorem 3. *Variant i's demand in the overlapping assortment search model is*

$$q_i^{so}(S) = q_i^m(S)(1 - H(\bar{U}(S), S)) \quad \textit{for } i \in S,$$

where H is defined as in Theorem 1, and $\bar{U}(S)$ is the unique solution to

$$\int_{\bar{U}(S)}^{\infty} (\bar{y} - \bar{U}(S)) w(\bar{y}_S, S)\, d\bar{y} = b, \tag{4}$$

where $w(\bar{y}_S, S)$ is the density function of $\max_{i \in \bar{S}} U_i$ (the maximum utility observed in set $\bar{S} = N - S$). Furthermore, for any given S and S^+ such that $\sum_{k \in S} v_k < \sum_{k \in S^+} v_k$, $\bar{U}(S) > \bar{U}(S^+)$, the search threshold $\bar{U}(S)$ is lower in a deeper assortment.

As in the independent assortment search model, variant i's demand in the overlapping assortment search model equals a fraction of the no-search model demand (assuming the same estimated preferences for the variants). However, the key difference between the two search models is that the search threshold in the independent assortment search model is fixed, $\bar{U}$, whereas it decreases in the assortment depth in the overlapping assortment search model ($\bar{U}(S) > \bar{U}(S^+)$). (Again, depth is measured as the sum of preferences.) As a result, although the cannibalization effect exists in the no-search and in the independent assortment search models, it can be shown that the cannibalization effect does not always exist in the overlapping assortment search model: given $i \in S$ and $S \subset S^+$, it is possible that variant i's demand *increases* as the assortment expands from S to S^+, i.e., $q_i^{so}(S) < q_i^{so}(S^+)$.[5]

[4] In the construction of a consumer search model, there is clearly a choice to be made between richness and parsimony. A rich model reflects many possible nuances involved in the consumer search process. For example, there surely is heterogeneity in consumer search costs and a consumer's expectation regarding the value of search can be quite complex: it could depend on the number of previous retailers the consumer visited, whether the consumer can return to the current retailer and the consumer's belief regarding how many new variants and which variants the consumer could sample via search. We believe that incorporating all of these features into a search model would render the model analytically intractable and unimplementable (e.g., it would not be clear how all of the necessary input parameters could be empirically estimated in practice). Hence we constructed our two parsimonious models with the hope that they capture the first order effects introduced by the presence of consumer search. Furthermore, they represent two extremes regarding consumer expectations on the value of search, so we are able to use these models to better understand, from a qualitative perspective, how variation in expectations may influence our findings.

[5] For instance, in a two-variant scenario with parameters: $u_1 = 5$, $u_2 = 3$, $u_0 = 1$, and $b = 0.1$. The variant 1's demand in the assortment $\{1\}$ is 0.53, but its demand in the assortment $\{1, 2\}$ is 0.87.

In that case, not only does an expanded assortment generate incremental revenue from the added variants, the added variants actually increase the demand (and operational efficiency) of the variants already in the assortment.

3. The Optimal Assortment

The optimal assortment for a retailer must trade off the benefit of including a variant in the assortment (it generates incremental sales) with the cost of including a variant in the assortment (if the cannibalization effect is present, then it reduces demand for the other variants, thereby reducing their revenue and lowering their operational efficiency). Because of this tension, in some cases, the profit function is ill-behaved: the optimal assortment can only be found with full enumeration over the $2^N - 1$ possible assortments. In other cases, it can be shown that the optimal assortment is included in the following (small) set, which we call the popular assortment set:

$$P = \{\{\}, \{1\}, \{1, 2\}, \ldots, \{1, \ldots, n\}\},$$

where recall the variants are labeled in decreasing order of net utility ($u_i - p_i$ is decreasing in i).

This section identifies the procedure for finding the optimal assortment with each of the three models. For notational convenience, we introduce some new notation. Let $S_j = S \cup \{j\}$ and let $q_i(v_j)$ be the demand for variant i with assortment S_j. Previously we used $q_i(S_j)$ for variant i's demand, but here we wish to make explicit the relationship between variant i's demand and variant j's preference. Similarly, let $\pi_i(v_j)$ be the profit of product i when variant j is added to assortment S. Finally, let $L(v_j)$ be the change in the retailer's profit when the assortment S is expanded to include variant j:

$$L(v_j) = \sum_{i \in S} \pi_i(S) - \sum_{i \in S} \pi_i(v_j).$$

If $\pi_j(v_j) > L(v_j)$, then adding variant j to S increases the retailer's profit.

3.1. No-Search Model

van Ryzin and Mahajan (1999) consider the assortment planning problem in the no-search model with a particular cost structure derived from the newsvendor model. They find that the optimal assortment is in the popular set P by showing that for *any* assortment S (which need not include the most popular variants), the retailer's profit function is quasi-convex in the net utility of the variant added to S. Therefore, if a variant is added to the assortment, it should be the most popular variant (highest net utility). Theorem 4 generalizes their results to include any concave increasing cost function.

Theorem 4. *The function $h^m(v_j) = \pi_j^m(v_j) - L^m(v_j)$ is quasi-convex in v_j on the interval $[0, \infty)$.*

3.2. Independent Assortment Search Model

The independent assortment search model is similar to the no-search model with the exception of the adjustment factor, $1 - H(\bar{U}, S)$, which is increasing as the assortment is expanded. Nevertheless, according to Theorem 5, the retailer's profit is quasi-convex in the net utility of a variant added to any assortment S. Hence the optimal assortment is again within P.

Theorem 5. *The function $h^{si}(v_j) = \pi_j^{si}(v_j) - L^{si}(v_j)$ is quasi-convex in v_j on the interval $[0, \infty)$. In the independent assortment search model, the optimal assortment is always within the popular set P.*

3.3. Overlapping Assortment Search Model

The overlapping assortment search model is more complex than the independent assortment search model because now the consumer's search threshold, $\bar{U}(S)$, is decreasing in the depth of the assortment. As a result, a new variant does not always cannibalize demand from other variants and the retailer's profit is not necessarily quasi-convex in the net utility of an added variant. Hence the optimal assortment may not be in the set of popular assortments, P. While we did find a scenario in which the optimal assortment is outside P, that did not occur in any of the scenarios in our numerical study (with identical margins across variants), which suggests that restricting the search for the optimal assortment to P is reasonable for a wide range of parameters.[6] However, if margins are decreasing in preferences, i.e., the more popular variants have smaller margins, then the numerical study

[6] The parameters with this scenario include $m_1 = m_2 = m_3 = 2$, $u_0 = 3.4$, $u_1 = 4.1$, $u_2 = 0.7$, $u_3 = 0.6$, $b = 0.122$, and the cost function is $c(q) = 0.5q^{1/2}$. The optimal assortment is $\{1, 3\}$.

indicates it is likely that the optimal assortment is outside of P. We note that full enumeration is also necessary in the no-search model if margins are decreasing in preferences.

Two other properties of the optimal solution to the overlapping assortment search model are worth discussing. First, in that model, it may be optimal to have an item in the assortment that has negative profit so as to decrease consumer search, whereas in the other models, each variant in the optimal assortment earns a strictly positive profit (to offset its cannibalizing effect on other variants). Second, in the overlapping assortment search model, a retailer can prevent search entirely by carrying the full assortment. In fact, according to Theorem 6, there exists a sufficiently low search cost such that the full assortment is optimal for the retailer.

Theorem 6. *In the overlapping assortment search model, for any given set of preference,* $\{v_0, v_1, \ldots, v_n\}$, *if the full assortment* (*all n variants*) *yields a positive profit, then there exists a threshold consumer search cost* $\bar{b}$ *such that the full assortment is optimal for all* $b \leq \bar{b}$.

4. Implementation of the No-Search Model

Implementation of an assortment planning model begins with the estimation of the model's necessary input parameters, and then the model is evaluated to produce a recommended assortment (i.e., the assortment that is optimal according to the model). With the no-search model the input parameters include each variant's preference (v_i), margin (m_i), and the operational cost function, $c(\cdot)$. We believe the latter two are relatively straightforward to estimate, but it is not straightforward to estimate each variant's preference because that involves breaking a circularity: To estimate consumer choice, we need an assortment to observe sales data, and to choose a good assortment we need an estimate of consumer choice.[7]

We study two approaches to break this circularity. With the depth-test approach, the retailer presents the full assortment to consumers and uses the resulting sales data to estimate v_i for each variant. Those preferences are inputted into the no-search assortment planning model and the recommended assortment is implemented. Alternatively, the retailer could take an iterative approach to estimate consumer preferences: the retailer chooses some assortment, collects choice data to estimate v_is, chooses an assortment given the latest estimates, and then continually iterates through this process. We take these approaches to be stylized representations of two extreme methods for implementing an assortment planning model.[8]

Although both implementations of the no-search model (depth test or iteration) correctly estimate preference for any given assortment, they do not correctly estimate the preferences that would be observed with a different assortment because they assume preferences are independent of the assortment chosen. In other words, the no-search model correctly accounts for consumer search for a given assortment, but then fails to account for how preferences change because of search if the assortment is altered.

Even though the no-search model does not correctly model consumer preferences under all assortments (because it does not explicitly account for consumer search), it may result in optimal or near optimal assortments. To explore this issue, we study a stylized model of a retailer who follows either the depth test or the iterative implementation of the no-search assortment planning model. We assume that search is relevant to consumer choice, and in particular, we assume that either the independent search model or the overlapping search model are the correct model of consumer choice. Hence, the in our experiment, we are able to compare the assortment chosen by

[7] The retailer may not need to actually implement the assortment in all stores. The retailer could conduct merchandize testing in a few stores or the retailer could conduct focus groups. But again, the data collected from those methods are dependent on the particular assortment included in the test. This is even an issue with a more intuitive approach to assortment planning. For example, suppose the retailer asks a merchant to give an estimate of each variant's preference. In the true model, each variant's preference depends on which other variants are in the assortment, so even with this approach there is a dependence between inputs and actions even if that dependency is difficult to model (because it is not clear how to model a merchant's intuition of relative value).

[8] For example, with short life-cycle products the retailer may be able to implement some kind of depth test to measure preferences, but will probably be unable to perform many iterations to collect updated data. Hence our iterative approach represents an ideal case in which the retailer has many opportunities to collect data.

the two implementations of the no-search model with the optimal assortment (the assortment chosen by the actual consumer choice model). This comparison suggests whether it is important to explicitly account for consumer search in the assortment planning process.

With either implementation of the no-search model, let $d_i(S)$ be variant i's observed demand with assortment S, where the demands have been normalized, so that

$$\sum_{i \in S \cup 0} d_i(S) = 1. \tag{5}$$

Let $\hat{v}_i(S)$ be the estimate of variant i's preference using the sales data from assortment S and $\hat{v}(S) = \{\hat{v}_0(S), \ldots, \hat{v}_n(S)\}$. The "^" superscript is used to emphasize that $\hat{v}_i(S)$ is an estimate of variant i's preference in contrast to variant i's actual preference, v_i, which is not directly observable to the retailer. The retailer obtains $\hat{v}_i(S)$ by solving the following system of equations:

$$d_i(S) = \frac{\hat{v}_i(S)}{\sum_{j \in S \cup 0} \hat{v}_j(S)}, \quad i \in S \cup 0.$$

There does not exist a unique solution to the above equations, so without loss of generality, we set $\hat{v}_i(S) = 1$ for some $i \in S \cup 0$. We choose variant 1 to be our normalized variant: $\hat{v}_i(S) = v_1 = 1$.

Let $x(\hat{v})$ be a correspondence that returns the optimal assortment according to the no-search model given the inputted preferences, $\hat{v}$. Recall, the no-search model chooses an assortment within the popular set P, i.e., $x(\hat{v}) \in P$. With the depth-test implementation the chosen assortment is $x(\hat{v})$, where $\hat{v} = \hat{v}(N)$. With the iterative implementation we assume that the retailer begins with the full assortment, as in the depth test, and then iterates. Let x^t be the tth assortment chosen by the retailer, let $\hat{v}^t$ be the preference estimates from the tth assortment and define the depth-test assortment as $t = 1$. Thus the iterative assortment planning process begins with $x^1 = N$ and $\hat{v}^1 = \hat{v}(N)$. Subsequent iterations have $x^{t+1} = x(\hat{v}^t)$ and $\hat{v}^{t+1} = \hat{v}(x^t)$. An assortment x^* is stable if $x^* = x(\hat{v}^*)$ and $\hat{v}^* = \hat{v}(x^*)$. With a stable assortment, the assortment is optimal according to the no-search model, given the estimated preferences and the estimated preferences are observed, given the assortment. Hence, the iterative process ends when a stable assortment is chosen.

With both implementations of the no-search model, we compare the profit of the assortment chosen (x^1 with the depth test and x^* with iteration) with the true optimal assortment, x^o, given the assumed search model (independent search or overlapping search). Hence we do not consider the cost of conducting the depth test nor the profits of the assortments chosen before converging to the stable assortment. Furthermore, we assume that there is no sampling error, i.e., the sales estimates equal the expected sales estimates. Thus we present an optimistic view of the no-search model. (See Xu 2003 for a simulation study on the impact of sampling errors.)

4.1. Analytical Results

As already discussed, the no-search model may choose a nonoptimal assortment because it could yield biased estimates of the variants' preferences. Interestingly, the no-search model actually does correctly estimate the *relative* preferences of any two product variants: for all S and all $0 < i < j \leq n$, $i \in S$, $j \in S$,

$$\frac{\hat{v}_i(S)}{\hat{v}_j(S)} = \frac{v_i}{v_j}.$$

This occurs because with either search model $q_i^m(S)/q_j^m(S)$ is constant for all S that contain i and j: for all S and all $0 < i < j \leq n$, $i \in S$, $j \in S$,

$$\frac{q_i^{si}(S)}{q_j^{si}(S)} = \frac{v_i}{v_j} \quad \text{and} \quad \frac{q_i^{so}(S)}{q_j^{so}(S)} = \frac{v_i}{v_j}.$$

This is known as the Independence of Irrelevant Alternatives (IIA) property: in either search model, the relative preference between product variant i and j does not depend on which other products are in the assortment. Hence the no-search model correctly estimates $v_1, \ldots, v_n$ no matter the assortment chosen.[9]

[9] The IIA property is well known for the MNL consumer choice model and in some cases it yields results that are inconsistent with actual choices. Nested logit models are an extension of the MNL used to deal with this issue. We suspect that our results continue to hold, at least to some degree, even for more complex choice models: our results occur because the preference for one variant (the no-purchase variant) depends on which other variants are included in the assortment, so even more complex choice models are susceptible to this bias if they do not include that dependency. In fact, in our three models, the violation of the IIA property occurs precisely because of search.

But the IIA property does not hold between the no-purchase variant and any of the product variants. For example, with independent assortment search,

$$\frac{q_0^{si}(S)}{q_i^{si}(S)} = \frac{q_0^m(S) + \sum_{j\in S} q_j^m(S) H(\bar{U}, S)}{q_i^m(S)(1 - H(\bar{U}, S))},$$

which implies for $i > 0$ that $\hat{v}_0(S)/\hat{v}_i(S)$ depends on S. The analogous result also holds for the overlapping assortment model.

Because the IIA property does not hold for the no-purchase variant, the no-search model provides biased estimates of the product preferences relative to the no-purchase option. Theorem 7 describes the resulting bias.

Theorem 7. *Take any assortment S and the preferences estimated from that assortment, $\hat{v}(S) = (\hat{v}_0(S), \hat{v}_i(S)\colon i \in S, \hat{v}_j(N)\colon j \in N - S)$. With any narrower assortment S^-, i.e., $\sum_{k\in S^-} v_k < \sum_{k\in S} v_k$, the no-search model under-estimates the no-purchase demand, $d_0(S^-) \geq q_0^m(S^- \mid \hat{v}(S))$. With any deeper assortment S^+, i.e., $\sum_{k\in S} v_k < \sum_{k\in S^+} v_k$, the no-search model overestimates the no-purchase demand, $d_0(S^+) \leq q_0^m(S^+ \mid \hat{v}(S))$.*

From Theorem 7, the no-search model overestimates the retailer's demand with a narrower assortment and underestimates the retailer's demand with a deeper assortment. The retailer overestimates demand with a narrower assortment because the retailer does not account for consumer search behavior: consumer search is more likely as the assortment becomes narrower, but that behavior is not reflected in the estimates of the preferences.

By using Theorem 7, we are able to compare the stable assortment in the iterative implementation with the optimal assortment.

Theorem 8. *Let x^* be a stable assortment with the iterative implementation of the no-search assortment planning model. If search is governed by the independent assortment search model, then any stable assortment x^* has the optimal number of variants or fewer, $|x^*| \leq |x^o|$ (i.e., never more than the number of variants in the optimal assortment). If search is governed by the overlapping assortment search model and $\pi_i(x^o)$ is increasing for all $q_i \geq q_i^{so}(x^o)$, then $\sum_{j\in x^*} v_j \leq \sum_{j\in x^o} v_j$, i.e., the stable assortment is not deeper than the optimal assortment.*

Because the optimal and stable assortments with the independent assortment search model are in the popular set P, a deeper assortment (in terms of the sum of the preferences) implies an assortment with more variants. In the overlapping assortment search model, there is no guarantee that the optimal and stable assortments are in the popular set P, so Theorem 8 cannot make a statement with respect to the number of variants, only with respect to the sum of the preferences, but that is nearly as strong a statement. Furthermore, because of the lack of the cannibalization effect, the theorem's result requires that increasing demand for any variant in the optimal assortment increases its profit, which is a relatively mild condition. Finally, if the optimal and stable assortments are in the set P, then it immediately follows that the optimal assortment has at least as many variants as the stable assortment.

It is not possible to provide general results that compare the depth-test assortment and the optimal assortment, i.e., the depth-test assortment may or may not be deeper than the optimal assortment. To explain, in a stable assortment x^*, it follows that $\pi(x^*) = \pi(x^* \mid \hat{v}(x^*)) \leq \pi(x^o)$, which implies $x^* \leq x^o$. However, in the optimal depth-test assortment x^d, we have $\pi(x^d) < \pi(x^d \mid \hat{v}(N))$, according to Theorem 7 and $\pi(x^d \mid \hat{v}(N)) <> \pi(x^o)$. Similarly, it is not possible to compare the depth-test assortment with the stable assortment.

Theorem 8 applies to any stable assortment, and we find in the numerical study that multiple stable assortments are possible (but not common). It should also be noted that Theorem 8 does not establish the existence of a stable assortment, and we indeed find in the numerical study instances in which there does not exist a stable assortment.

4.2. A Numerical Study

We conducted a numerical study to further explore the performance of the two implementations of the no-search assortment planning model. In all scenarios there are eight potential variants, $n = 8$. Three functions are used to assign utilities: $u_i = 8 - \phi(1 - e^{-(i-1)})$ for $\phi \in \{1, 3, 6\}$. The no-purchase variant is assigned one of three utilities: $u_0 = u_1 - 0.25(u_1 - u_8)$, $u_0 = u_1 - 0.75(u_1 - u_8)$, and $u_0 = 0.75u_8$. The first case has a relatively attractive no-purchase utility, while

in the latter case the no-purchase utility is even lower than the least attractive product variant. Nine different cost functions are constructed: $c(q) = \nu q^{\beta}$, with $\nu \in \{0.2, 0.5, 0.8\}$ and $\beta \in \{0.2, 0.5, 0.7\}$. These functions include newsvendor-like cost functions as well as the EOQ cost function. There are nine search costs: $b \in \{0.1, 0.4, \ldots, 2.5\}$. We consider six margin patterns: $m_i \in \{3, 5, 7, 1 + 1.5(i-1), 52.5/63 + 10.5i^2/63, 11.5 - 10.5e^{i-1}\}$. The first three have identical margins across variants. In the latter three patterns, margins are linear, concave, and convex in i, respectively, but in each of them, margins are decreasing in utility and $m_1 = 1$ and $m_8 = 11.5$. The variants have identical prices, normalized to an arbitrary constant. With independent assortment search, there are three utilities for the outside alternative, $u_r \in \{u_1, u_1 - 0.5(u_1 - u_8), u_8\}$. In total, there are 13,122 scenarios with independent assortment search, 4,374 scenarios with overlapping assortment search, and 17,496 scenarios in total.

With each scenario, we evaluated the optimal assortment and the assortment chosen either with the depth-test implementation or the iterative implementation of the no-search model. For scenarios in which the optimal assortment may not be included among the set of popular assortments, P, we enumerated all assortments to determine the optimal one. With identical margins, the optimal assortment was found to always be contained within P, but with decreasing margins, there are many scenarios with optimal assortments outside of P.

We removed 277 scenarios from the data set: 78 were removed because the optimal assortment contains no variants, i.e., there does not exist a profitable assortment because operating costs are too high relative to demand; and 199 were removed because they did not contain a stable assortment nor a cycle of assortments (a series of assortments that the iterative implementation cycles through continuously). The majority of the remaining scenarios contain at least one stable assortment: 92% of the 17,219 remaining scenarios. In scenarios with an assortment cycle, we averaged the realized profit across the assortments. In scenarios with more than one cycle or more than one stable assortment, we assume the expected profit equals the maximum across the choices. Table 1 summarizes these data.

While we demonstrated that the cannibalization effect of a deeper assortment implies that every variant must make a strictly positive profit with independent assortment search, we conjectured that the optimal assortment with overlapping assortment search might include some money losing variants. In fact, we found a few scenarios (69 with identical margins and 130 with decreasing margins) in which the optimal assortment with overlapping assortment search included variants with a negative expected profit. Search costs were low in all of these scenarios, so the retailer is willing to carry some money losing variants to prevent consumer search. Nevertheless, even with overlapping assortment search, it appears in most cases it is optimal for each item to earn a positive profit. In those cases, the condition in Theorem 8 is satisfied, so there are no stable assortments that are deeper than the optimal assortment.

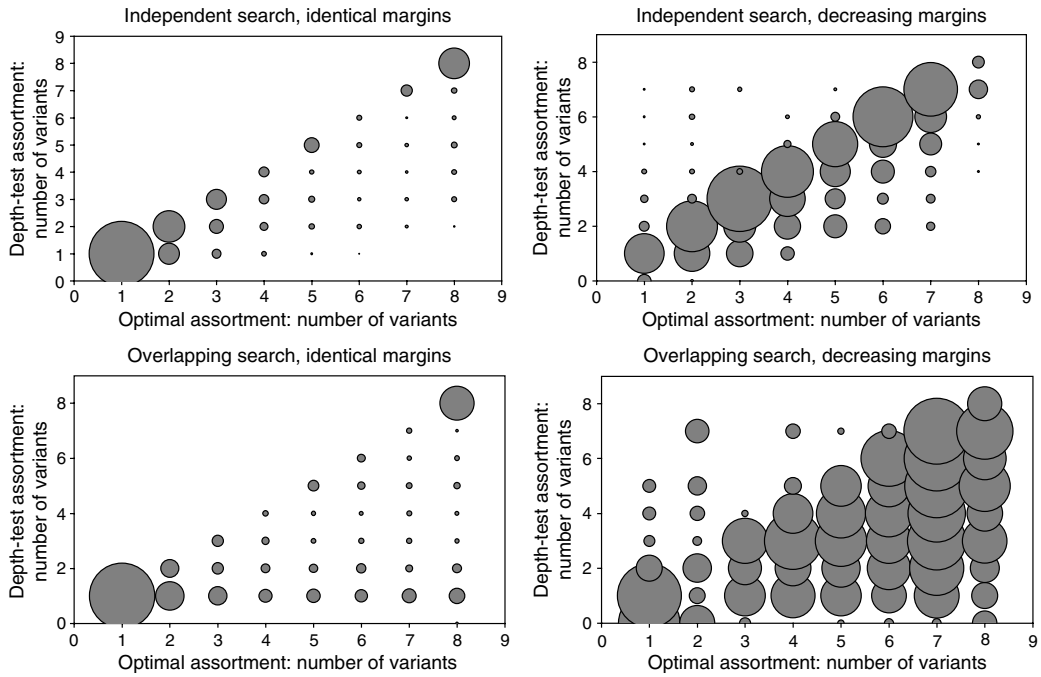
Figures 1 and 2 display the relationship between the number of variants in the optimal assortment and the number of variants in the no-search model. With independent assortment search and identical margins, we know (Theorem 8) that the stable assortments have no

Table 1 Scenarios in the Data Set and the Number of Stable Assortments

		Scenarios dropped from the data set		Scenarios kept in the data set				
				Number of stable assortments				
Search model	Margin pattern	Zero variants in optimal assortment	No stable assortment, no assortment cycle*	0	1	2	3	Total
Independent	Identical	0	7	187	6,367			6,554
	Decreasing	45	19	483	6,006	8		6,497
Overlapping	Identical	10	38	214	1,919	3	3	2,139
	Decreasing	23	135	471	1,496	62		2,029

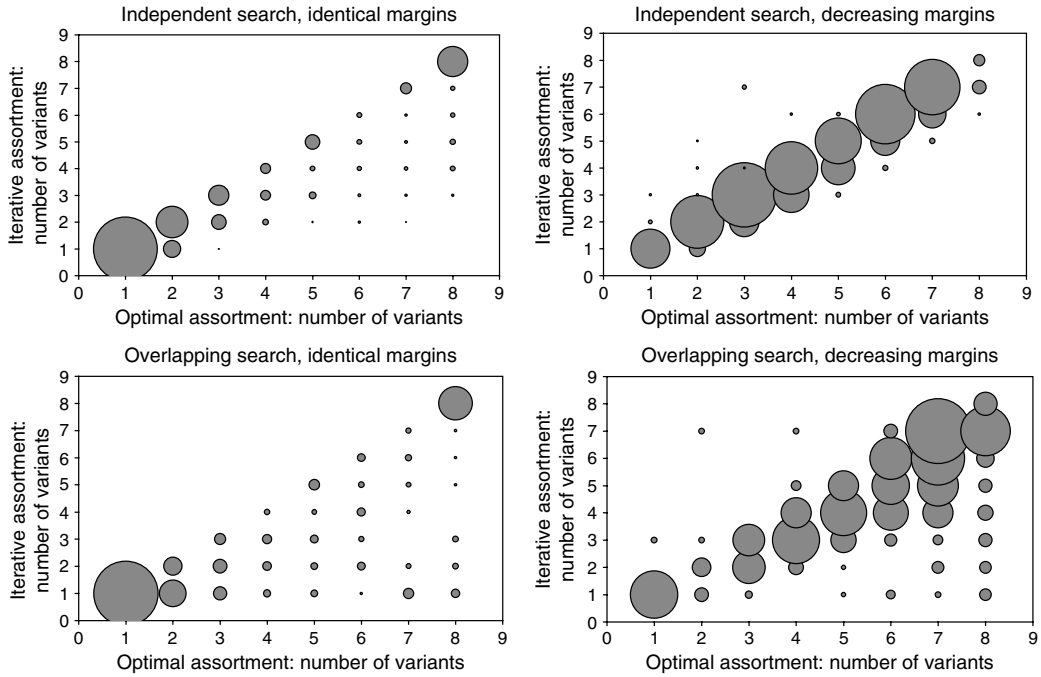
*An assortment cycle is a series of assortments that the iterative implementation cycles through repeatedly.

Figure 1 **The Number of Variants in the Optimal Assortment Relative to the Number of Variants in the Depth-Test Implementation of the No-Search Model**



Note. Bubble areas provide relative frequency.

Figure 2 **The Number of Variants in the Optimal Assortment Relative to the Number of Variants in the Highest Profit Stable Assortment from the Iterative Implementation of the No-Search Model**



Note. Bubble areas provide relative frequency.

more than the optimal number of variants, and this is reflected in the data displayed in the figure. We find the same result with overlapping search and identical margins because the condition in Theorem 8 is generally satisfied and the stable and optimal assortments are always (for these scenarios) in the popular set P. Even with decreasing margins, we generally find that with the no-search model the number of variants chosen does not exceed the number in the optimal assortment. Note that with decreasing margins even if the same number of variants is chosen as the optimal assortment, it is possible that the no-search model's profit is less than optimal because the wrong variants are chosen. We conclude from these figures that the no-search model rarely chooses more than the optimal number of variants and often chooses fewer than the optimal number of variants. (We also note that these results occur even if the optimal assortment is not part of the popular set.)

Table 2 provides some summary statistics for the performance of the no-search model under eight different combinations of implementation (depth test or iterative), margin (identical or decreasing), and search (overlapping or independent). The depth-test implementation generally performs worse than the iterative implementation. On average, the no-search model performs well with independent search and identical margins, but does not always perform well with overlapping search. In particular, the average profit loss is substantial in the identical margin scenarios when the optimal number of variants in the assortment is large (say, five or more), and in the decreasing margin scenarios when the optimal number of variants is either small (1 or 2) or large (6, 7, 8). The profit loss can be large even with few variants in the optimal assortment because the no-search model chooses the wrong variants to include in the assortment.

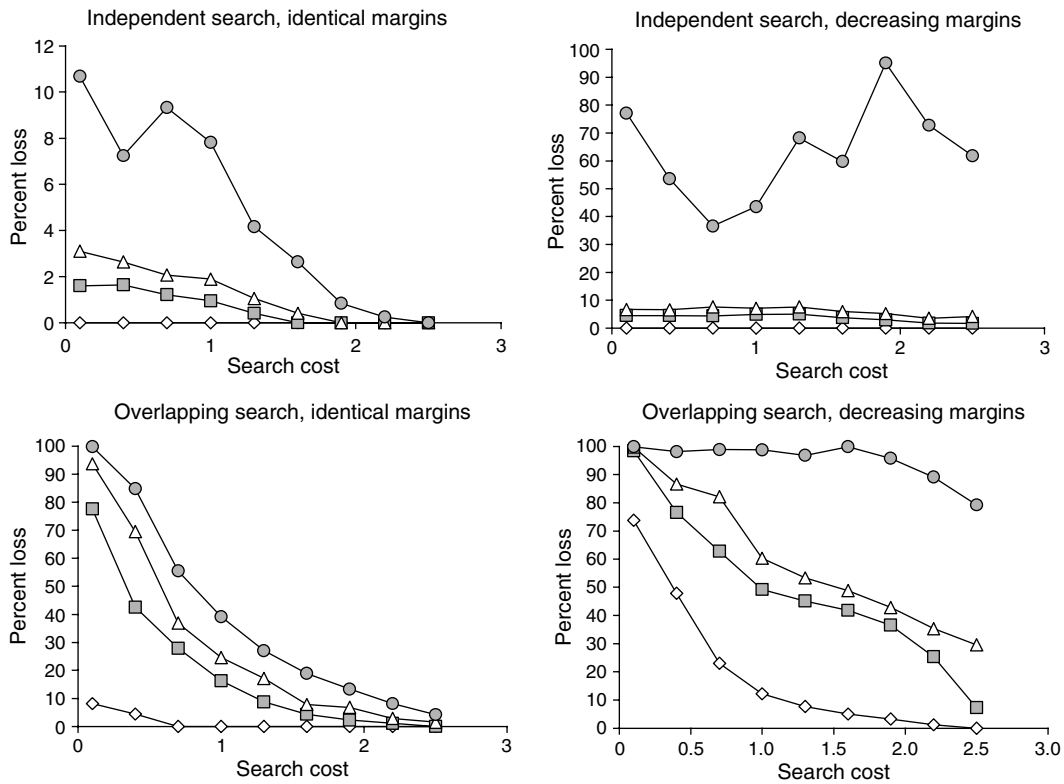
Table 2 Average Profit Loss (as a % of the Optimal Profit)

		Average % profit loss			
		Independent search		Overlapping search	
Margin pattern	Number of variants in the optimal assortment	Depth test	Iterative	Depth test	Iterative
Identical	1	0.0	0.0	0.0	0.0
	2	1.0	0.5	3.7	3.3
	3	1.6	0.4	8.5	5.7
	4	2.1	0.6	15.9	9.6
	5	1.3	0.4	20.2	12.1
	6	1.7	0.6	27.4	13.4
	7	0.9	0.3	52.7	38.1
	8	0.2	0.1	21.5	14.0
Overall		0.5	0.2	9.9	6.4
Decreasing	1	14.3	5.5	52.4	28.9
	2	6.5	1.3	70.9	36.3
	3	5.5	0.9	18.7	7.4
	4	5.7	1.0	24.7	12.1
	5	7.0	1.3	26.1	9.5
	6	5.5	0.7	28.7	13.5
	7	10.5	1.3	37.6	15.3
	8	12.3	6.8	70.8	48.9
Overall		7.3	1.4	41.0	21.3

Figure 3 displays how the profit loss with the iterative implementation depends on the search cost. (The comparable figures with the depth-test implementation are similar.) It is clear from the figure that ignoring consumer search is not a problem with independent search and identical margins. With overlapping search or decreasing margins, the profit loss from not explicitly accounting for search can be substantial. In some scenarios with low search costs, the optimal assortment includes every variant, while the no-search model's stable assortment has no variants. Even if the no-search model expands the assortment as search costs decrease, it generally does not expand the assortment enough. This is illustrated in Figure 4 for one scenario: as search costs decrease, the no-search model indeed adds variants, but it does not add enough variants, leading to a significant profit loss.

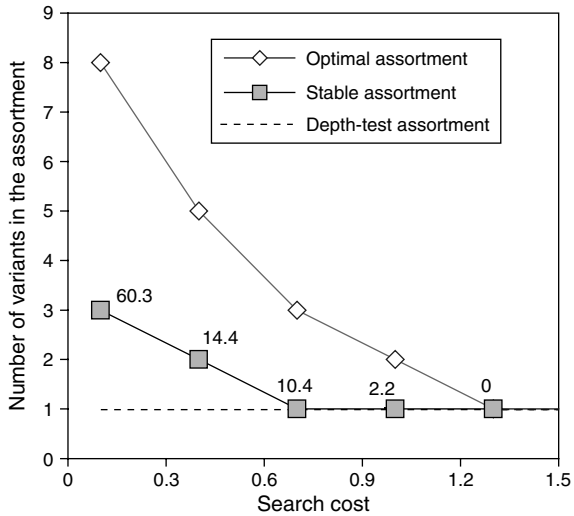
To summarize, we find that with either implementation of the no-search model, the chosen assortment tends to include too few variants (i.e., too narrow). The profit consequence of these narrow assortments is generally not substantial in the independent search model, but is often substantial with the overlapping search model. Although these results are obtained using three specific assortment planning models, we conjecture that our qualitative conclusions should apply more generally: Ignoring search in an assortment planning model is likely to lead to assortments that are too narrow because doing so removes a key benefit from expanding an assortment, the benefit of preventing consumer search.

Figure 3 **Average Profit Loss (as a Percentage of the Optimal Profit) of the Best Stable Assortment**



Note. Diamonds = 50th fractile, squares = 90th fractile, triangles = 95th fractile, and circle = maximum.

Figure 4 **Assortments Chosen for Five Scenarios that Differ Only in Search Cost**



Note. The profit loss (as a percentage of the optimal profit) are indicated in the graph for the stable assortment. Parameters: overlapping search, constant margin (5), $\phi = 3$, $u_0 = u_1 - 0.75(u_1 - u_8)$, $v = 0.5$, $\beta = 0.5$.

5. Conclusion

The focus of this research is on the question of whether it is necessary to explicitly account for consumer search in the assortment planning process. We define three assortment planning models. The no-search planning model, which is a slight generalization of the model studied by van Ryzin and Mahajan (1999), is based on the MNL model of consumer choice and does not explicitly account for consumer search. In particular, the no-search model assumes the consumers' preference for a variant is fixed, i.e., not dependent on the other variants in the assortment. The other two models expand upon the no-search model to include some form of consumer search. In the first one, which we call the independent assortment search model, there is essentially an unlimited pool of variants, which implies a low probability that the same product variant is carried in the assortment of two retailers. In the second one, called the overlapping assortment search model, there is a limited pool of variants. Hence, expanding the

retailer's assortment reduces the value of search to a consumer. In the extreme, a retailer with a full assortment (all available variants) eliminates all search.

It is important to explicitly model search in the assortment planning process if the no-search model recommends assortments that are substantially less profitable than the optimal assortment. We identify the optimal assortment by assuming the true consumer choice model is one of our search models. We find that the no-search model performs well on average with independent assortment search: in categories (such as antiques or jewelry) where the potential set of variants is so large that it is unlikely any two retailers carry the same assortment, the no-search model calibrates itself reasonably well to indirectly account for consumer search. However, the no-search model does not perform well with overlapping assortment search, which is more appropriate in environments with a limited, but potentially large, amount of variety (e.g., digital cameras). In particular, the no-search model leads the retailer to choose too few variants in the assortment. In some extreme cases, the no-search model recommends not carrying a single item in the assortment even though the full assortment is profitable. Intuitively, the no-search model performs worse as search costs decrease, i.e., as it becomes cheaper for consumers to search, it becomes more important to explicitly account for their search preferences.

We recognize that our observations are derived from three stylized models, which were chosen for analytical tractability and consistency with the existing literature. Nevertheless, we suspect our qualitative results do not depend critically on the details of our models. For example, even without MNL consumer preferences, there is likely to be a trade-off between the expanded variety of a broader assortment and the additional operational costs of more variety, and a consumer's value from searching for a better product is likely to decrease as a retailer expands its assortment. And even if a retailer chooses to implement a different analytical assortment planning process, that planning process must begin with a consumer choice model and some estimation of demand parameters. Hence, any implementation of an analytical model must be careful if its estimated inputs are somehow influenced by the recommended actions the model outputs. (In other words, we feel this issue is relevant even if a retailer were to implement an assortment planning model that differed from the ones we consider.)

Our results also contribute to the debate on whether the trend in product proliferation is excessive or not. If an analytical model is applied to a retailer's assortment and that analytical model does not explicitly account for consumer search, then the model may recommend to narrow the assortment even though the true optimal assortment is broader. In other words, the trend toward broad assortments may be because of an intuitive understanding by retail managers that broad assortments prevent consumer search. Interestingly, this bias toward narrow assortments from no-search analytical models can occur even if the firm iteratively updates its demand estimates as it changes its assortment. It can even occur if the demand estimates and the assortment are stable (i.e., the assortment is optimal given the demand estimates and the demand estimates are observed given the chosen assortment). Furthermore, the consequence for failing to account for consumer search in an assortment planning model can be substantial: the no-search model may recommend closing down a category (include no variants in the assortment) even though there exists an assortment with positive profit (a 100% profit loss).

Acknowledgments

The authors thank Erica Plambeck, Garrett van Ryzin, the associate editor, and the reviewers for their helpful comments. Financial support from the Wharton e-Business Initiative and the Fishman-Davidson Center is gratefully ackowledged.

Appendix

Proof of Theorem 1. Search is worthwhile if, and only if,

$$\int_{U_{\max}-u_r}^{\infty} (u_r + x - U_{\max}) f(x)\, dx - \int_{-\infty}^{U_{\max}-u_r} (U_{\max} - u_r - x) f(x)\, dx \geq b,$$

where the first term is the expected gain from search, the second term is the expected loss from search, and the third term is the cost of search. After rearranging terms and recognizing $E[\zeta_r] = 0$, the above simplifies to $u_r - b \geq U_{\max}$: a consumer benefits from search if $U_{\max} < \overline{U}$, where $\overline{U} = u_r - b$.

It remains to evaluate $q_i^{si}(S)$. Variant i is chosen if $U_i = U_{\max}$ (it is the best variant in the assortment), $U_i > \bar{U}$ (its utility is above the search threshold), and $U_i > U_0$ (its utility is better than no purchase). Let ϕ be the realization of ζ_i. Thus, variant i is chosen if $\phi \geq \bar{U} - (u_i - p_i)$ $(U_i > \bar{U})$ and $U_i \geq \max\{U_{\max}, U_0\}$, which has probability $\prod_{j=0,\, j\in S/i} F((u_i - p_i) + \phi - (u_j - p_j))$. Overall, the probability variant i is selected, $q_i^{si}(S)$, is

$$q_i^{si}(S) = \int_{\bar{U}-(u_i-p_i)}^{\infty} f(\phi) \prod_{\substack{j=0\\ j\in S/i}} F((u_i - p_i) + \phi - (u_j - p_j))\, d\phi.$$

Following the process of deriving choice probabilities in the traditional MNL, substitute the cumulative distribution function (CDF) of the Gumbel distribution and conduct the change of variables $\delta = \exp[-(\phi/\mu + \gamma)]$ and $v_j = \exp((u_j - p_j)/\mu)$ to obtain

$$q_i^{si}(S) = \int_0^{\alpha} \exp\left[-\delta\left(\frac{v_0 + \sum_{j\in S} v_j}{v_i}\right)\right] d\delta,$$

where $\alpha = \exp[-((\bar{U} - u_i)/\mu + \gamma)]$. The above integral simplifies to

$$\begin{aligned} q_i^{si}(S) &= \frac{v_i}{v_0 + \sum_{j\in S} v_j}\left\{1 - \exp\left[-\left(v_0 + \sum_{j\in S} v_j\right)\right.\right. \\ &\qquad\qquad \left.\left. \cdot \exp[-(\bar{U}/\mu + \gamma)]\right]\right\} \\ &= q_i^m(S)(1 - H(\bar{U}, S)). \quad \square \end{aligned}$$

Proof of Theorem 2. Define function $T(\omega) = \omega(1 - \exp(-\lambda/\omega))$, where λ is a positive constant. The first derivative of $T(\omega)$ is

$$T'(\omega) = 1 - \left(1 + \frac{\lambda}{\omega}\right)\exp\left(-\frac{\lambda}{\omega}\right).$$

$T(\omega)$ is an increasing function on the interval $[0, \infty)$ because it follows from the Taylor expansion of $\exp(\lambda/\omega)$ that $1 + \lambda/\omega < \exp(\lambda/\omega)$ for $\omega \geq 0$ and $\lambda \geq 0$. From (3), we have

$$q_i^{si}(S) = v_i T\left(\left(\sum_{k\in S} v_k + v_0\right)^{-1}\right).$$

If $\sum_{k\in S} v_k < \sum_{k\in S^+} v_k$, it implies that $\sum_{k\in S} v_k + v_0 < \sum_{k\in S^+} v_k + v_0$. Thus, by the fact that $T(\omega)$ is increasing, it follows that $q_i^{si}(S) > q_i^{si}(S^+)$, $\forall\, i \in \{S \cap S^+\}$. $\square$

Proof of Theorem 3. Let $\bar{y}_S$ be the realized maximum utility from the set $\bar{S}$, and let $w(\bar{y}_S, S)$ be its density function. (Because the realizations of ζ_i are independent for the products in $\bar{S}$, it is straightforward to evaluate $w(\bar{y}_S, S)$.) A consumer searches if

$$\int_{U_{\max}}^{\infty} (\bar{y}_S - U_{\max}) w(\bar{y}_S, S)\, d\bar{y}_S \geq b: \tag{6}$$

the consumer is assured of at least $U_{\max}$ utility, so the first term is the expected incremental gain over $U_{\max}$ from search. The left-hand side of (6) is decreasing in $U_{\max}$, so there exists a unique threshold utility, $\bar{U}(S)$, such that

$$\int_{\bar{U}(S)}^{\infty} (\bar{y}_S - \bar{U}(S)) w(\bar{y}_S, S)\, d\bar{y}_S = b.$$

A consumer searches if, and only if, $U_{\max}$ is less than the threshold $\bar{U}(S)$. Note that the same holds in the independent assortment search model except the threshold is independent of the assortment in that model. Therefore the analysis to determine $q_i^{so}(S)$ follows the approach in Theorem 1 to determine $q_i^{si}(S)$.

Let $\bar{y}_{S^+}$ be the realized maximum utility from the set $\bar{S}^+ = N - S^+$ and $g(\bar{y}_{S^+}, S^+)$ be its density function. Note that the CDF of $\max_{i\in S} U_i$ is given as the function H defined in Theorem 1. Since $\sum_{k\in S} v_k < \sum_{k\in S^+} v_k$, which implies $\sum_{k\in N-S} v_k > \sum_{k\in N-S^+} v_k$, by the definition of H, $\bar{y}_S$ is stochastically larger than $\bar{y}_S^+$. Therefore, according to the Proposition 9.1.2 in Ross (1996), for any given y, we have

$$\int_y^{\infty} (\bar{y}_S - y) w(\bar{y}_S, S)\, d\bar{y}_S > \int_y^{\infty} (\bar{y}_{S^+} - y) g(\bar{y}_{S^+}, S^+)\, d\bar{y}_{S^+}. \tag{7}$$

Since both sides of (7) are decreasing in y, therefore we have $\bar{U}(S) > \bar{U}(S^+)$, where $\bar{U}(S)$ and $\bar{U}(S^+)$ are the two values of y that make both sides of (7) equal to b, respectively. $\square$

Proof of Theorem 4. Let $V_S = v_0 + \sum_{i\in S} v_i$. From differentiation,

$$h^{m\prime}(v_j) = \frac{\left[mV_S - c'\left(\frac{v_j}{v_j+V_S}\right)V_S\right] - \left[m\sum_{i\in S} v_i - \sum_{i\in S} c'\left(\frac{v_i}{v_j+V_S}\right)v_i\right]}{(v_j + V_S)^2}. \tag{8}$$

The numerator is increasing in v_j because $v_j/(v_j + V_S)$ is increasing in v_j, $v_i/(v_j + V_S)$ is decreasing in v_j and $c(\cdot)$ is concave. It follows (because $(v_j + V_S)^2 > 0$) that there is at most one v_j such that $h^{m\prime}(v_j) = 0$. $\square$

Proof of Theorem 5. Differentiate $h^{si}(v_j)$,

$$\begin{aligned} h^{si\prime}(v_j) &= mq_j^{si\prime}(v_j) - c'(q_j^{si}(v_j))q_j^{si\prime}(v_j) \\ &\quad + \sum_{i\in S}[mq_i^{si\prime}(v_j) - c'(q_i^{si}(v_j))q_i^{si\prime}(v_j)] \\ &= J(v_j)f(v_j) + N(v_j)g(v_j), \end{aligned} \tag{9}$$

where

$$\bar{H}(v_j) = 1 - H(\bar{U}, S_j) \qquad f(v_j) = m - c'\left(\frac{\bar{H}(v_j)v_j}{v_j + V_S}\right)$$

$$V_S = v_0 + \sum_{i\in S} v_i \qquad N(v_j) = \frac{\bar{H}'(v_j)(v_j + V_S) - \bar{H}(v_j)}{(v_j + V_S)^2}$$

$$J(v_j) = \frac{\bar{H}(v_j)V_S + \bar{H}'(v_j)v_j(v_j + V_S)}{(v_j + V_S)^2}$$

$$g(v_j) = m\sum_{i\in S} v_i - \sum_{i\in S} c'\left(\frac{\bar{H}(v_j)v_i}{v_j + V_S}\right)v_i.$$

We wish to show that there exists at most one v_j such that $h^{si\prime}(v_j)=0$. It can be shown that $J(v_j)>0$, $N(v_j)<0$, $f(v_j)$ is increasing and $g(v_j)$ is decreasing. Unfortunately, neither J nor N are monotone in v_j, so we need additional results.

Because $J(v_j)>0$ and $N(v_j)<0$, if $h^{si\prime}(v_j)=0$, then $f(v_j)$ and $g(v_j)$ must have the same sign. The following statement is stronger: for all $\{v_j^1, v_j^2\}$ such that $v_j^1<v_j^2$ and $h'(v_j^1)=h'(v_j^2)=0$, it holds that $f(v_j^1)f(v_j^2)>0$. (Proof by contradiction. Suppose $f(v_j^1)<0$ and $g(v_j^1)<0$, while $f(v_j^2)>0$ and $g(v_j^2)>0$: in that case, $g(v_j^2)>g(v_j^1)$, which contradicts $g(v_j)$ is decreasing. Suppose $f(v_j^1)>0$ and $g(v_j^1)>0$, while $f(v_j^2)<0$ and $g(v_j^2)<0$: in that case, $f(v_j^2)<f(v_j^1)$, which contradicts $f(v_j)$ is increasing.) Given that the sign of $f(v_j)$ is the same for all v_j such that $h^{si\prime}(v_j)=0$, there are two cases to consider: either $f(v_j)<0$ or $f(v_j)>0$.

Case 1: $f(v_j)<0$ and $g(v_j)<0$. We have

$$\begin{aligned} h^{si\prime}(v_j) &= \frac{e^{-\lambda(v_j+V_S)}}{(v_j+V_S)^2}[D(v_j)f(v_j)-K(v_j)g(v_j)] \\ &= \frac{e^{-\lambda(v_j+V_S)}}{(v_j+V_S)^2}D(v_j)(-f(v_j))\left[\frac{g(v_j)}{f(v_j)}\frac{K(v_j)}{D(v_j)}-1\right], \end{aligned} \tag{10}$$

where

$$D(v_j)=\left(e^{\lambda(v_j+V_S)}-1+\frac{\lambda v_j}{V_S}(v_j+V_S)\right)V_S$$

$$K(v_j)=e^{\lambda(v_j+V_S)}-1-\lambda(v_j+V_S).$$

Both $D(v_j)$ and $K(v_j)$ are positive and increasing in v_j and $K(v_j)/D(v_j)$ is increasing. Furthermore, for $f(v_j)<0$ and $g(v_j)<0$, $g(v_j)/f(v_j)$ is increasing. Hence, there is at most one v_j such that $h^{si\prime}(v_j)=0$.

Case 2: $f(v_j)\geq 0$ and $g(v_j)\geq 0$. We are unable to rewrite (10) in a multiplicative form of a positive term and an increasing term in this case. However, it is sufficient to show that for all v_j such that $h^{si\prime}(v_j)=0$, then $h^{si\prime\prime}(v_j)>0$. Differentiate

$$\begin{aligned} &h^{si\prime\prime}(v_j) \\ &\quad = mq_j^{si\prime\prime}(v_j)-c''(q_j^{si}(v_j))(q_j^{si\prime}(v_j))^2-c'(q_j^{si}(v_j))q_j^{si\prime\prime}(v_j) \\ &\qquad +\sum_{i\in S}[mq_i^{si\prime\prime}(v_j)-c''(q_i^{si}(v_j))(q_i^{si\prime}(v_j))^2-c'(q_i^{si}(v_j))q_i^{si\prime\prime}(v_j)]. \end{aligned}$$

Because $c(\cdot)$ is concave,

$$\begin{aligned} h^{si\prime\prime}(v_j) &\geq mq_j^{si\prime\prime}(v_j)-c'(q_j^{si}(v_j))q_j^{si\prime\prime}(v_j) \\ &\qquad +\sum_{i\in S}[mq_i^{si\prime\prime}(v_j)-c'(q_i^{si}(v_j))q_i^{si\prime\prime}(v_j)] \\ &= \tilde{J}(v_j)f(v_j)+\widetilde{N}(v_j)g(v_j), \end{aligned}$$

where

$$\tilde{J}(v_j)=\frac{2V_S[\bar{H}'(v_j)(v_j+V_S)-\bar{H}(v_j)]+\bar{H}''(v_j)v_j(v_j+V_S)^2}{(v_j+V_S)^3},$$

$$\widetilde{N}(v_j)=\frac{[\bar{H}''(v_j)(v_j+V_S)^2-2\bar{H}'(v_j)(v_j+V_S)+2\bar{H}(v_j)]}{(v_j+V_S)^3}.$$

Thus it is sufficient to show that

$$\tilde{J}(v_j)f(v_j)+\widetilde{N}(v_j)g(v_j)>0. \tag{11}$$

From $\tilde{J}(v_j)\leq 0$ (because $\bar{H}''(v_j)\leq 0$), $f(v_j)\geq 0$, $g(v_j)\geq 0$, and

$$h^{si\prime}(v_j)=J(v_j)f(v_j)+N(v_j)g(v_j)=0,$$

(11) holds if

$$\frac{-\widetilde{N}(v_j)}{\tilde{J}(v_j)}>\frac{-N(v_j)}{J(v_j)}. \tag{12}$$

Simplifying and rearranging terms in (12) yields

$$\begin{aligned} &2-2\exp[-\lambda(v_j+V_S)]-\lambda(v_j+V_S) \\ &\qquad \cdot\exp[-\lambda(v_j+V_S)]-\lambda(v_j+V_S)<0, \end{aligned}$$

which holds for all $v_j+V_s>0$.

Applying the similar process in the proof of Theorem 1 in van Ryzin and Mahajan (1999), the quasi-convexity of $h^{si}(v_j)$ implies that the optimal assortment is always within the set P. □

Proof of Theorem 6. We need to show that every assortment $S\neq\{1,2,\ldots,n\}$ leads to less profit than the full assortment, which generates a positive profit. The expected profit of assortment S is

$$\begin{aligned} \pi(S) &= \sum_{i\in S}\big[m_iq_i^{so}(S)-c(q_i^{so}(S))\big] \\ &= \sum_{i\in S}\big[m_iq_i^m(S)(1-H(\bar{U}(S),S)) \\ &\qquad -c(q_i^m(S)(1-H(\bar{U}(S),S)))\big]. \end{aligned} \tag{13}$$

From the search rule specified in Theorem 3, the search threshold $\bar{U}(S)$ is monotonically decreasing in the search cost b. Furthermore, $\lim_{b\to\infty}\bar{U}(S)=-\infty$ and $\lim_{b\to 0}\bar{U}(S)=\infty$. Therefore, from (13), there exists a sufficiently low b such that $H(\bar{U}(S),S)$ is close enough to 1 to ensure that $\pi(S)$ is less than the full assortment profit for any S. □

Proof of Theorem 7. We show the result for the independent assortment search model. To establish $d_0(S^-)\geq q_0^m(S^-\mid\hat{v}(S))$, it is sufficient to show

$$\hat{v}_0(S^-)/\hat{v}_i(S^-)\geq\hat{v}_0(S)/\hat{v}_i(S).$$

Recall that

$$d_i(S)=\hat{v}_i(S)\Big/\sum_{j\in S}\hat{v}_j(S),$$

so

$$\frac{\hat{v}_0(S)}{\hat{v}_i(S)}=\frac{d_0(S)}{d_i(S)}=\frac{q_0^m(S)+\sum_{j\in S}q_j^m(S)H(\bar{U},S)}{q_i^m(S)(1-H(\bar{U},S))}$$

and we need to establish

$$\frac{q_0^m(S^-)+\sum_{j\in S^-}q_j^m(S^-)H(\bar{U},S^-)}{q_i^m(S^-)(1-H(\bar{U},S^-))}\geq\frac{q_0^m(S)+\sum_{j\in S}q_j^m(S)H(\bar{U},S)}{q_i^m(S)(1-H(\bar{U},S))},$$

which can be, using (2) and the expressions for $H()$, written as

$$\frac{v_0 + (\sum_{j\in S^-} v_j)\exp(-\lambda \sum_{j=0,\, j\in S^-} v_j)}{1-\exp(-\lambda \sum_{j=0,\, j\in S^-} v_j)} \geq \frac{v_0 + (\sum_{j\in S} v_j)\exp(-\lambda \sum_{j\in S} v_j)}{1-\exp(-\lambda \sum_{j=0,\, j\in S} v_j)}. \tag{14}$$

Define

$$Z(\delta) = \frac{v_0 + \delta\exp(-\lambda(v_0+\delta))}{1-\exp(-\lambda(v_0+\delta))},$$

so we only need $Z'(\delta) \leq 0$, where

$$\begin{aligned} Z'(\delta) = {} & \big([1-\lambda(v_0+\delta) - \exp(-\lambda(v_0+\delta)) \\ & - (1-\lambda)(v_0+\delta\exp(-\lambda(v_0+\delta)))]\exp(-\lambda\delta)\big) \\ & \cdot \big([1-\exp(-\lambda(v_0+\delta))]^2\big)^{-1}. \end{aligned}$$

By Taylor expansion of $\exp(-\lambda(v_0+\delta))$, we have $1-\lambda(v_0+\delta)-\exp(-\lambda(v_0+\delta)) \leq 0$ for any $\lambda(v_0+\delta) \geq 0$, so we indeed have $Z'(\delta) \leq 0$. The proof for the overlapping assortment search model is similar, except now λ is a function of $\bar{U}(S)$. But from Theorem 3, $\bar{U}(S^-) > \bar{U}(S)$, so $\lambda(\bar{U}(S^-)) < \lambda(\bar{U}(S))$. We see that $Z(\delta, \lambda)$ is decreasing in λ, so the condition (14) is easier to satisfy in the overlapping assortment model. A similar process demonstrates $d_0(S^+) \leq q_0^m(S^+ \mid \hat{v}(S))$. □

Proof of Theorem 8. First, consider the independent assortment search model. Suppose we have $|x^*| > |x^o|$, which because $x^* \in P$ and $x^o \in P$, implies $\sum_{j\in x^o} v_j < \sum_{j\in x^*} v_j$. Given $\hat{v}(x^*)$ from Theorem 7, the retailer overestimates the expected profit of the assortment x^o, i.e., $\pi(x^o \mid \hat{v}(x^*)) > \pi(x^o)$: because $\pi_i(S)$ is convex in $q_i^{si}(S)$, and the cannibalization effect implies $\pi_i(x^o) > 0$ for all $i \in x^o$, it follows that for each variant in x^o, its profit increases if its preference is increased. Note that $\pi(x^*) = \pi(x^* \mid \hat{v}(x^*))$. However, combining these results yields

$$\begin{aligned} \pi(x^*) &= \pi(x^* \mid \hat{v}(x^*)) \\ &= \max_{x\in P}\{\pi(x \mid \hat{v}(x^*))\} \geq \pi(x^o \mid \hat{v}(x^*)) > \pi(x^o), \end{aligned}$$

which contradicts that x^o is the optimal assortment (i.e., we know that $\pi(x^o) > \pi(x^*)$). Hence it must be that $|x^*| \leq |x^o|$. The proof for the overlapping assortment search model is similar. Now the cannibalization effect is not guaranteed to be present, so there is an additional condition that $\pi_i(x^o)$ is increasing for all $q_i \geq q_i^{so}(x^o)$, which implies the estimates $\hat{v}(x^*)$ overestimates the expected profit of the assortment x^o, i.e., $\pi(x^o \mid \hat{v}(x^*)) > \pi(x^o)$. □

References

Anderson, S., R. Renault. 1999. Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model. *RAND J. Econom.* **30**(4) 719–735.

Anderson, S., A. de Palma, J. F. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, MA.

Armony, M., E. Plambeck. 2005. The impact of duplicate orders on demand estimation and capacity investment. *Management Sci.* Forthcoming.

Aydin, G., W. H. Hausman. 2002. Supply chain coordination and assortment planning. Working paper, Stanford University, Stanford, CA.

Bayus, B., W. P. J. Putsis. 1999. Product proliferation: An empirical analysis of product line determinants and market outcomes. *Marketing Sci.* **18**(2) 137–153.

Cachon, G., G. Kok. 2005. Implementation of the newsvendor model with clearance pricing: How to (and how not to) estimate a salvage value. Working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Chong, J. K., T. H. Ho, C. S. Tang. 2001. A modeling framework for category assortment planning. *Manufacturing Service Oper. Management* **3**(3) 191–210.

Cooper, W., T. Homem-de-Mello, A. Kleywegt. 2005. Models of the spiral-down effect in revenue management. *Oper. Res.* Forthcoming.

Fisher, M., K. Rajaram. 2000. Accurate retail testing of fashion merchandise: Methodology and application. *Marketing Sci.* **19**(3) 266–278.

Fisher, M., R. Raman, A. McClelland. 2000. Rocket science retailing is almost here—Are you ready? *Harvard Bus. Rev.* **78**(4) 115–124.

Kok, G., M. Fisher. 2004. Demand estimation and assortment optimization under substitution: Methodology and application. Working paper, Fuqua School of Business, Duke University, Durham, NC.

Mahajan, S., G. van Ryzin. 1998. Retail inventories and consumer choice. S. Tayur, R. Ganeshan, M. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA, 491–551.

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

Morgan, P. B. 1983. Search and optimal sample sizes. *Rev. Econom. Stud.* **50** 659–675.

Pashigian, B. P. 1988. Demand uncertainty and sales: A study of fashion and markdown pricing. *Amer. Econom. Rev.* **78**(5) 936–953.

Quelch, J., D. Kenny. 1994. Extend profits, not product lines. *Harvard Bus. Rev.* **72**(5) 153–160.

Ross, S. 1996. *Stochastic Processes*. John Wiley and Sons, New York.

Salop, S., J. Stiglitz. 1977. Bargains and ripoffs: A model of monopolistically competitive price dispersion. *Rev. Econom. Stud.* **44** 493–510.

Smith, S. 2002. Choosing optimal retail assortments using conjoint preference data. *INFORMS Conf.*, San Jose, CA.

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

Stahl, D. O. 1989. Oligopolistic pricing with sequential consumer search. *Amer. Econom. Rev.* **79**(4) 700–712.

Stahl, K. 1982. Differentiated products, consumer search, and locational oligopoly. *J. Indust. Econom.* **31** 97–114.

Stigler, G. J. 1961. The economics of information. *J. Political Econom.* **69**(3) 213–225.

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

Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica* **47**(3) 641–654.

Wolinsky, A. 1983. Retail trade concentration due to consumers' imperfect information. *Bell J. Econom.* **14**(1) 275–282.

Wolinsky, A. 1984. Product differentiation with imperfect information. *Rev. Econom. Stud.* **51** 53–61.

Xu, Y. 2003. Essays in process improvement and assortment planning. Doctoral dissertation, The Wharton School, University of Pennsylvania, Philadelphia, PA.