Approximation Algorithms for Product Framing and Pricing

Guillermo Gallego, Anran Li, Van-Anh Truong, Xinshang Wang
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, USA,
gmg2@columbia.edu,al2942@columbia.edu, vatruong@ieor.columbia.edu,xw2230@columbia.edu

We propose one of the first models of "product framing" and pricing. Product framing refers to the way consumer choice is influenced by how the products are framed, or displayed. We present a model where a set of products are displayed, or framed, into a set of virtual web pages. We assume that consumers consider only products in the top pages, with different consumers willing to see different numbers of pages. Consumers select a product, if any, from these pages following a general choice model. We show that the product framing problem is NP-hard. We derive algorithms with guaranteed performance relative to an optimal algorithm under reasonable assumptions. Our algorithms are fast and easy to implement.

We also present structural results and design algorithms for pricing under framing effects for the multinomial logit model. We show that for profit maximization problems, at optimality, products are displayed in descending order of their value gap and in ascending order of their markups.

1. Introduction

In this paper, we propose one of the first models of *product framing* and pricing. Framing refers to the way in which the choice among available alternatives is influenced by how the alternatives are framed, or displayed (Tversky and Kahneman 1986). For example, empirical works by Agarwal, Hosanagar and Smith (2009) and Ghose and Yang (2009) in online advertising show that ads that are placed higher on a webpage attract more clicks from consumers. Johnson, Moe, Fader, Bellman and Lohse (2004) examine the average number of websites, sorted by product categories, that are actively visited by households each month. They observe that in a typical search session, consumers search from fewer than two stores. Their data show that 70% of CD shoppers, 70% of book shoppers, and 42% of travel shoppers are loyal to just one site. Brynjolfsson, Dick, and Smith (2010) find, in the context of a website that catalogs price and product information from multiple retailers, that only 9% of users select offers that are listed beyond the first page. In related search contexts, Baye, Gatti, Kattuman and Morgan (2009) have found that a consumer's likelihood of visiting a firm and purchasing from it is strongly related to the order in which the firm is listed on a webpage by a search engine. They find that a firm receives about 17% fewer clicks for every competitor listed above it on the screen, all other things being equal.

This well-documented framing effect is a natural outcome of the cognitive burden of processing larger and larger assortments. During online shopping, it is cognitively harder for a typical consumer to visit sellers who are listed at the bottom of a web page, before or in addition to visiting those who are listed at the top (Animesh, Viswanathan and Agarwal 2011). In the context of online retailing, it has been observed that consumers' attention to a display decreases exponentially with the display's distance to the top (Feng, Bhargava, and Pennock 2007). Thus, positioning a brand or product at a top position on a listing can improve both consumer attention to the brand, and consequently, consumer selection of the brand (Chandon, Hutchinson, Bradlow, and Young 2009).

1.1. Model Overview

Despite substantial evidence suggesting the impact of framing on consumers' choice outcome, there are very few models that have attempted to capture these effects. In this paper, we introduce one of the first models for product framing and the first one for pricing that accounts explicitly for these effects.

We base our model on the notion of *consideration set*. A consideration set is the set of products over which a consumer will make utility comparisons before arriving at the final purchase decision. Consideration sets have gained considerable acceptance since their introduction in the seminal work of Howard and Sheth (1969). A widely used approach to modeling choice in psychology and marketing is to assume that a consumer will first form a consideration set. Then she will choose from among the alternatives in the set. Consideration sets explain, behaviorally, consumers' limited ability to process or acquire information (Manrai and Andrews 1998). Methodologically, it has been shown that ignoring consideration sets may lead to biased parameter estimates (Chiang, Chib and Narasimhan 1999), whereas including consideration sets improves the predictability of choice models (Hauser and Gaskin 1984, Silk and Urban 1978). As an example, Hauser (1978) finds that a disproportionate 78% of the explainable uncertainty in consumer choice can be accounted for by consideration sets, whereas the Multinomial Logit Model (MNL) can only capture the remaining 22%.

We model the effect of framing on the formation of consideration sets as follows. Products are organized into virtual pages. Each page can hold a finite number, say p, of products. A consumer will examine only the first X pages, where X is a random variable that may be *personalized* to the consumer's profile. The consumer forms a consideration set consisting of only products in the examined pages. From this consideration set, the consumer makes a choice according to a *general choice model*. Thus, products that are placed in earlier pages are more likely to be considered, and therefore purchased, than those that are placed in later pages.

Given the behavior described above, we study two problems that are faced by an e-retailer who is managing n different products in a particular product category. The retailer's *product framing problem* is how to determine an assortment and a distribution of the products in the assortment into the different pages in order to maximize the expected revenue. The retailer's *joint pricing and framing problem* is to determine both the framing and pricing of the products in order to maximize the expected revenue.

We further model *location preference* effects that come into play after the consideration sets have been formed. Location preference works as follows. First, given that a certain page enters into a consumer's consideration set, products that are displayed higher on the page are more likely to be chosen than those displayed lower on the same page, all other factors being equal. Second, given that two different pages enter into a consumer's consideration set, products that are listed in the earlier page are more likely to be chosen than products that are listed in the later page, all other factors being equal.

1.2. Results and Implications for Retailers

Our contributions in this paper are the followings:

- We propose one of the first models of framing effects. Our model allows for a general choice model and a general framing structure.
- We prove that the product framing problem is NP-hard, even when there are only 2 pages and the choice model is the MNL model.
- We propose fast, easy-to-implement algorithms with worst-case performance guarantees. The ease and simplicity of the algorithms mean that they can be personalized on-line for each arriving consumer.
- We show that an optimal joint pricing and framing solution for the MNL model has the following properties: (i) Every page is fully filled with products until all the products are displayed. (ii) All products on the same page have the same page-level markup (price minus cost), which increases monotonically with the page index. (iii) Products with higher value gaps are displayed earlier. The value gap of a product is the net utility when the product is sold at the unit wholesale cost. This implies that more attractive products (i.e., those having higher value gaps and lower markups) are displayed in earlier positions. This last finding is contrary to the findings of Arbatskaya (2007), who argues that consumers with lower search costs will search longer and obtain better deals.

Our model also applies to brick-and-mortar retailers with a suitable interpretation of what consumers are willing to look at. Some consumers would only look at the most prominent displays (see Chandon, Hutchinson, Bradlow and Young 2009, Corstjens and Corstjens 2012), while others may look at some aisles or the entire the store. In this context, our location-preference helps to

model the values of having products at eye level and the value of end-of-aisle locations. Our pricing results suggest that the most prominent displays should have the highest utility products at the lowest markups.

2. Literature Review

2.1. Relation to Assortment Optimization

Our paper falls within the literature on assortment optimization. Assortment optimization began with a stylized model introduced by van Ryzin and Mahajan (1999). They show that under the MNL model, an optimal assortment consists of a certain number of highest-utility products when the products are equally profitable. When the products' prices are given exogenously and the choice model is the MNL model, Talluri and van Ryzin (2004) prove that an optimal assortment includes a certain number of products with the highest revenues.

The assortment optimization problem is easy to solve for the MNL model over a given consideration set. Davis, Gallego and Topaloglu (2013) show that this problem can be formulated as a linear program with totally unimodular constraints. Davis, Gallego and Topaloglu (2014) also propose that under the nested logit (NL) model, the assortment optimization problem can be solved by a linear program when the nest-dissimilarity parameters of the choice model are less than one, and each consumer always makes a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard.

The assortment optimization problem is NP-hard for general choice models. Indeed, Bront, Mendes-Diaz and Vulcano (2009) show that under the mixed multinomial logit (MMNL) choice model, the assortment optimization problem with a fixed number of mixtures is NP-hard. Desir and Goyal (2013) show that this problem is even NP-hard to approximate within a factor of $O(n^{1-\epsilon})$, for any fixed $\epsilon > 0$. They give approximation schemes that tradeoff running time with solution quality, but the running time for their approach grows exponentially with the number of mixtures.

Few papers have studied assortment optimization with location effects. Davis, Gallego and Topaloglu (2013) model location effects by introducing location-dependent item weights to the MNL model, but they do not model consideration sets. The resulting assortment optimization problem reduces to a linear program with totally unimodular constraints.

Assortment optimization under consideration-set-based choice models have been studied by a number of authors. One stream works with endogenous consideration sets that arise as a result of search. Cachon (2005) shows that ignoring consumer search will lead to less assortment variety, since in equilibrium, the seller needs a larger assortment to attract more consumers. Sahin and Wang (2015) also study the assortment optimization problem with search costs. They assume consumers are homogeneous, and their search sequences are predetermined by all the products'

expected utilities that are known to the consumers before they search. Feldman and Topaloglu (2015) study a model in which consumers choose products according to the MNL model, but consumers of different types have different consideration sets, and the sets are fixed and nested. They devise a fully polynomial-time approximation scheme for this problem.

To our knowledge, there are only two papers that model a framing-dependent formation of consideration sets. Davis, Topaloglu and Williamson (2015) study a problem in which a firm must sequentially add products to its assortment over time, thereby monotonically increasing consumers' consideration sets. They provide an algorithm with constant relative performance. The decision space for this problem is more constrained than ours and the application context is very specific. Aouad and Segev (2016) consider a variant of our model, where the number of products that can be displayed on each page is one, the choice model is the MNL model, and all products must be displayed even if doing so is suboptimal.

2.2. Relation to Assortment Pricing

Our work is also related to assortment pricing. Hanson and Martin (1996) are among the first to notice that the expected revenue function fails to be concave in pricing problems, even under the MNL model. Song and Xue (2007) show that the expected revenue is concave with respect to the market shares. Under the MNL model with uniform price-sensitivity parameter, the markup, defined as price minus cost, has been shown to be constant across all products at optimality (Anderson, de Palma and Thisse 1992, Hopp and Xu 2005, and Gallego and Stefanescu 2011). By assuming that the price sensitivities of the products are constant within each nest and the nest dissimilarity parameters are restricted to the unit interval, Li and Huh (2011) extend the concavity result to the NL model. Gallego and Wang (2014) consider the general NL model with product-differentiated price-sensitivity parameters and arbitrary nest coefficients. They find that the adjusted nest-level markup is also constant across all the nests.

We extend the assortment pricing literature by modeling framing effects. Under the MNL profit-maximization model, we find that the constant markup property holds at the page level. We also show that optimal markups are higher for products having lower value gaps, which is contrary to the findings of Arbatskaya (2007).

3. Product Framing Problem

Throughout this paper, we let $[k]$ denote the set $\{1, 2, \ldots, k\}$ for any positive integer k.

Consider n products. Product i has unit profit or revenue r_i, $i \in [n]$. This interpretation allow us to handle both profit optimization and revenue maximization problems in a single framework, perhaps after the transformation $r_i \leftarrow r_i - c_i$ where c_i is the unit cost of product i. Products are

organized into m virtual pages. Each page can hold up to p products. Potentially all of the products may be offered, but offering all of the products is not a hard requirement. The consideration set of a consumer is the set of all products in pages $1, 2, \ldots, X$, where $X \in [m]$ is the random number of pages that the consumer is willing to view. From this consideration set, the consumer purchases at most one product according to a general choice model. Let $\lambda(x) = \mathbf{P}[X = x]$, and $\Lambda(x) = \mathbf{P}[X \geq x]$ for all $x \in [m]$.

We *do not know* the number of pages X that a consumer is willing to view when she arrives into the system. We assume, however, that the distribution of X is known and is independent of the framing of the products. Knowledge of X can be acquired from observing click data and by computing the frequency of consumers who examine $x \in [m]$ pages. By the law of large numbers, these frequencies converge to the probability distribution of X. Although we will refer to a single random variable X, it is easy to see that X can be personalized to heterogeneous consumer types based on available information about the distribution of pages they are willing to see. Information that may change the distribution of X includes, but is not limited to, prior purchases, zip code, age, and gender.

We first assume that consumers choose according to a general choice model that is independent of X. Later in Section 6, we will remove this assumption and show that under mild conditions we can still design framing algorithms with provable performance guarantees.

The *product framing problem* is to distribute the n products among the m pages to maximize the expected revenue that can be obtained from an arriving consumer. Our model does not impose any hard constraints on whether a particular page should be fully filled, i.e., should contain exactly p products. However, all the solutions generated by our algorithms have a desirable property that, for any page $x \in [m]$ that is not empty, all the previous pages $1, 2, \ldots, x-1$ are full. In other words, our algorithms do not leave holes in the displayed pages.

The product framing problem can be formulated in terms of decision variables $f \in \{0,1\}^{n \times m}$, where f_{ix} indicates whether product $i \in [n]$ is displayed on page $x \in [m]$. Let $P(i, S)$ denote the purchase probability of product i when the consideration set is $S \subseteq [n]$, with $P(i, S) = 0$ if $i \notin S$. The formulation in terms of the variables f_{ix} is given by

$$\begin{aligned} V^{\mathsf{OPT}} = \max_{f \in \{0,1\}^{n \times m}} & \sum_{x \in [m]} \lambda(x) \sum_{i \in [n]} r_i P(i, \{k \in [n] : \sum_{l=1}^{x} f_{kl} = 1\}) \\ \text{s.t.} \ & \sum_{x \in [m]} f_{ix} \leq 1, \ \forall i \in [n] \\ & \sum_{i \in [n]} f_{ix} \leq p, \ \forall x \in [m] \\ & f_{ix} \in \{0,1\}, \ \forall i \in [n], \ x \in [m]. \end{aligned} \tag{1}$$

Here V^{OPT} is the expected revenue of an optimal framing algorithm OPT. For other sub-optimal algorithms Π, we denote the expected revenue of Π as V^{Π}.

3.1. Hardness of the Framing Problem

We show that problem (1) is NP-hard even in the special case that $m = 2$ and the choice model is the MNL model. We do this by reducing the well-known 2-PARTITION problem to a special case of our model. The 2-PARTITION problem is defined as follows

Definition 1 (2-PARTITION). Given a set of n non-negative numbers $w_1, w_2, \ldots, w_n$, determine whether there is a set $S \subseteq [n]$ such that $\sum_{i \in S} w_i = \sum_{i \notin S} w_i$.

Our reduction works as follows. Starting with any instance of 2-PARTITION, we design an instance of problem (1). We show that the solution to the continuous relaxation of this problem takes a certain value if and only if there is a solution to the 2-PARTITION problem.

Theorem 1. *Problem* (1) *is NP-hard even when there are two pages and all consumers follow the same MNL model.*

3.2. Upper Bound on the Optimal Revenue

Since the problem of finding an optimal framing solution is NP-hard, it is interesting to characterize upper bounds on V^{OPT} that are easy to compute. We now propose such an upper bound, which we will use to establish approximation ratios of our framing algorithms.

Consider the following assortment optimization problem, which constrains the number of products in the assortment to be at most c.

$$\begin{aligned} G(c) \ &= \max_{S \subseteq [n]} \sum_{i \in S} r_i P(i, S) \\ \text{s.t. } &|S| \leq c. \end{aligned} \tag{2}$$

Define $U(x) \equiv G(x \cdot p)$, for all $x \in [m]$, as the optimal expected revenue from consumers who see $x \in [m]$ pages. If we had the luxury of knowing the number of pages X upon the arrival of a consumer, we would offer her an optimal assortment that yields expected revenue $U(X)$, and earn expected total revenue

$$\mathbf{E}[U(X)] = \sum_{x \in [m]} \lambda(x) U(x). \tag{3}$$

The following result states that this $\mathbf{E}[U(X)]$ is an upper bound on the optimal expected revenue V^{OPT}.

Theorem 2. $\mathbf{E}[U(X)] \geq V^{OPT}$.

4. Approximation Algorithms for Product Framing

In this section, we propose framing algorithms having guaranteed constant performance ratios relative to OPT. The performance guarantee we prove in this section is tight relative to the upper bound (3), in the sense that our proofs exhibit ways to construct instances in which the bounds are achieved by our algorithms.

4.1. Assumptions for Analysis of Algorithms

For the purpose of proving theoretical performance guarantees, we make three innocuous assumptions:

Assumption A1 *$P(i,S) \geq P(i,T)$ for all $i \in S$ and $S \subseteq T \subseteq [n]$.*

Assumption A2 *In polynomial time, we can obtain a solution with expected revenue $\bar{G}(c)$ to problem (2) such that $\bar{G}(c) \geq (1-\epsilon)G(c)$ for some constant $\epsilon > 0$.*

Assumption A3 *X has new better than used in expectation (NBUE) distribution.*

We stress that the framing problem is still NP-hard under these three assumptions, because the reduction in the proof of Theorem 1 incorporates all these assumptions.

Assumption A1 is very general as it holds for all random utility models. Davis, Topaloglu and Williamson (2015) have shown that Assumption A1 leads to the following results, which we will use in the analysis.

Lemma 1. *(Davis et al. 2015) For any set $S \subseteq [n]$ with $|S| \geq 2$, there exists $i \in S$ such that*

$$\frac{\sum_{k \in S, k \neq i} r_k P(k, S \setminus \{i\})}{|S|-1} \geq \frac{\sum_{k \in S} r_k P(k,S)}{|S|}.$$

Lemma 2. *(Davis et al. 2015) $U(x)/x$ is decreasing in x, for $x \in [m]$.*

Lemma 1 means that it is always possible to raise the per-product revenue by removing one of the products from an assortment. This is intuitive as we only need to remove the product $k \in S$ having the smallest contribution $r_k P(k,S)$. Lemma 2 is an analogous version of Lemma 1 at page level, i.e., the optimal per-page revenue decreases as consumers consider more pages.

Assumption A2 states that we can solve the capacity-constrained problem (2) within a constant approximation ratio. For the MNL and NL models, the capacity-constrained problem can be exactly solved in polynomial time ($\epsilon = 0$) (Gallego and Topaloglu 2014). For the Mixed Multinomial Logit model with a constant number of mixtures, the capacity-constrained problem can be solved within any constant error $\epsilon > 0$ in polynomial time (Mittal and Schulz 2013, Desir and Goyal 2013). For ease of exposition, in the rest of the paper we assume that $\epsilon = 0$, but our algorithms and bounds can be easily extended to the case of $\epsilon > 0$ by replacing an optimal solution to (2) with an approximate one, and scaling the corresponding bound by $(1-\epsilon)$.

Assumption A3 means that the additional expected number of pages that a consumer will see is no more than the expected number of pages that she would like to see before the search. If we let $q(x) = \mathbf{E}[X - x + 1 | X \geq x]$ for all $x \in [m]$, then Assumption A3 is equivalent to $q(x) \leq q(1)$ for all $x \in [m]$.

4.2. NEST Algorithm

We first propose a family of algorithms, called $\mathsf{NEST}(y)$, which start by truncating the number of pages to an arbitrary integer y, and selecting an optimal assortment, say $S(y)$, for consumers who view the first y pages. Then they select an assortment for the first $y-1$ pages, say $S(y-1)$, by looking only at the products in $S(y)$. This procedure continues until the content of all the y pages have been determined.

For all $x \in [m]$, let $S(x) \subseteq [n]$ denote the set of products to be displayed in the first x pages. Let $R(S)$ denote the expected revenue when S is the consideration set of a consumer.

$\mathsf{NEST}(y)$ Algorithm (for $y \in [m]$):

1. Solve the assortment optimization problem (2) with cardinality bound $c = y \cdot p$. Set $S(y)$ to be the assortment solution.

2. For $x = y-1$ down to 1, choose $S(x) \subseteq S(x+1)$ and $|S(x)| = \min(|S(x+1)|, x \cdot p)$ such that

$$\frac{R(S(x))}{|S(x)|} \geq \frac{R(S(x+1))}{|S(x+1)|}. \tag{4}$$

Lemma 1 ensures that we can always find such a set $S(x)$. Note that if $|S(x+1)| > x \cdot p$, then $|S(x)| = x \cdot p$, meaning that the first x pages are fully filled.

3. Leave pages $y+1, y+2, \ldots, m$ blank. Hence, $S(x) = S(y)$ for all $x > y$.

Now for any choice of y, there is a corresponding expected total revenue $V^{\mathsf{NEST}(y)}$. We let NEST be the algorithm that chooses the y that brings the highest expected revenue. That is,

$$V^{\mathsf{NEST}} = \max_{y \in [m]} V^{\mathsf{NEST}(y)}.$$

4.3. Constant Approximation Ratio

We now show that NEST achieves at least $6/\pi^2 \approx 0.608$ times the optimal expected revenue. First, we derive lower bounds on $V^{\mathsf{NEST}(y)}$.

Proposition 1. *$V^{NEST(y)} \geq \frac{U(y)}{y}\mathbf{E}[\min(X,y)]$, for all $y \in [m]$.*

Consequently, as NEST chooses y to maximize the expected total revenue, we must have

$$V^{\mathsf{NEST}} = \max_{y \in [m]} V^{\mathsf{NEST}(y)} \geq \max_{y \in [m]} \frac{U(y)}{y}\mathbf{E}[\min(X,y)].$$

We prove the performance guarantee of NEST by comparing this lower bound on V^{NEST} against the upper bound (3) on V^{OPT}. The idea of the proof is to examine the worst-case structure of $\max_{y \in [m]} \frac{U(y)}{y}\mathbf{E}[\min(X,y)]$ over all functions $U(\cdot)$ satisfying Lemma 2, and over all distributions of X satisfying Assumption A3. For convenience we will scale $U(\cdot)$ without loss of generality so $\mathbf{E}[U(X)] = 1$. This leads to the following min max problem:

$$
\begin{aligned}
\gamma \quad &= \quad \min_{U,\Lambda} \max_{x\in[m]} \frac{U(x)}{x}\mathbf{E}[\min(X,x)] \\
\text{s.t.} \quad & 1=\Lambda(1)\geq\Lambda(2)\geq\cdots\geq\Lambda(m)\geq 0, \\
& \Lambda(x+1)\times\sum_{y=1}^{m}\Lambda(y)\geq\sum_{y=x+1}^{m}\Lambda(y), \quad \forall x=1,2,\ldots,m-1, \\
& U(x)\leq U(x+1), \quad \forall x=1,2,\ldots,m-1, \\
& \frac{U(x)}{x}\geq\frac{U(x+1)}{x+1}, \quad \forall x=1,2,\ldots,m-1, \\
& \mathbf{E}[U(x)]=1, \\
& U(x)\geq 0, \quad \forall x\in[m].
\end{aligned} \tag{5}
$$

The first constraint ensures that Λ corresponds to a valid tail distribution. The second constraint ensures that X has the NBUE property. The third and fourth constraints ensure that $U(\cdot)$ is increasing and satisfies Lemma 2. The fifth constraint normalizes $\mathbf{E}[U(X)]$ to 1, and the last ensures that $U(\cdot)$ is non-negative.

To solve this bound-revealing problem (5), we will first characterize the functions $U(\cdot)$ and $\Lambda(\cdot)$ in the worst case. In the process of establishing the bounds, we will not use special notations, say $\gamma^*, U^*(\cdot)$ or $\Lambda^*(\cdot)$, to denote an optimal solution to problem (5). This comes at a small cost of ambiguity, but makes the exposition much cleaner.

The following proposition gives the worst-case structure for $U(\cdot)$. It states that, in the worst case, $\mathbf{E}[\min(X,x)]U(x)/x$ is constant for all $x\in[m]$.

Proposition 2. $\gamma=\frac{U(x)}{x}\mathbf{E}[\min(X,x)]$ *for all* $x\in[m]$.

For $U(x)$ of the form given by Proposition 2, we next show how γ depends on the distribution of X, by using the fact that $\mathbf{E}[U(X)]=1$.

Proposition 3. *Let Y be a random variable that is independently and identically distributed as X. Then Problem* (5) *can be equivalently re-formulated as*

$$
\begin{aligned}
\frac{1}{\gamma} &= \max_{\Lambda}\mathbf{E}\left[\frac{X}{\mathbf{E}[\min(X,Y)|X]}\right] \\
s.t. \quad & 1=\Lambda(1)\geq\Lambda(2)\geq\cdots\geq\Lambda(m)\geq 0, \\
& \Lambda(x+1)\times\sum_{y=1}^{m}\Lambda(y)\geq\sum_{y=x+1}^{m}\Lambda(y), \quad \forall x=1,2,\ldots,m-1.
\end{aligned} \tag{6}
$$

We prove the performance guarantee of NEST by showing that $\mathbf{E}[\frac{X}{\mathbf{E}[\min(X,Y)|X]}]\leq\pi^2/6$ among all non-negative, NBUE distributions of X.

Theorem 3. $V^{NEST}\geq\frac{6}{\pi^2}V^{OPT}$.

In our analysis, we used an exponential distribution to achieve the bound for Problem (5). The reader may wonder whether there is a discrete distribution over non-negative integers such that the bound for Problem (5) is also tight. The following proposition asserts that this is indeed the case.

PROPOSITION 4. *The performance ratio $6/\pi^2$ with respect to the upper bound (3) is attained when m tends to infinity, X follows a geometric distribution with mean $1/(1-\alpha)$, and α tends to 1 from below.*

4.4. Practical Improvement to the Algorithm

We can refine $\mathsf{NEST}(y)$ in various ways to improve its empirical performance. First, we can greedily add products to pages beyond y. As long as the expected revenue of assortment $S(x)$ is at least $U(y)$ for all $x = y+1, y+2, \ldots, m$, the lower bound given by Proposition 1 will still hold, and thus the approximation ratio $6/\pi^2$ will still be valid. Second, in Step 2 of $\mathsf{NEST}(y)$, instead of looking for a set $S(x)$ that satisfies equation (4), we can find an optimal $S(x)$ from $S(x+1)$ by solving

$$\max_{S(x)\subseteq S(x+1)} \sum_{i\in S(x)} r_i P(i, S(x))$$
$$\text{s.t. } |S(x)| = \min(|S(x+1)|, x\cdot p).$$

This method is favorable when the choice model allows for efficient assortment optimization algorithms.

We use $\mathsf{NEST}{+}(y)$ to denote the variant of $\mathsf{NEST}(y)$ resulting from the above modifications. We let $\mathsf{NEST}{+}$ be the algorithm that chooses y to maximize the expected revenue of $\mathsf{NEST}{+}(y)$. We will test both NEST and $\mathsf{NEST}{+}$ in our numerical experiments in Section 8.

COROLLARY 1. *If the capacitated assortment problem* (2) *has a unique optimal solution for each c, and all the optimal solutions for different c's are nested, i.e., $T(1) \subseteq T(2) \subseteq \cdots \subseteq T(n)$, where $T(c)$ is the unique optimal assortment for $G(c)$, then* $\mathsf{NEST}{+}$ *returns an optimal solution to the product framing problem.*

5. Product Framing with Location Preferences

In online retailing, consumers may be more likely to choose products that are displayed at the top among search results, since consumers tend to associate high valuation with products that are displayed earlier (Chandon, Hutchinson, Bradlow, and Young 2009). In this section, we augment our model to capture the phenomenon that a consumer is more likely to buy a product displayed earlier, even if the consumer has determined her consideration set. We call this the *location preference* effect.

We model location preference by using the MNL choice model and introducing location-dependent preference weights for all products. We use ν_{ixq} to denote the preference weight of product $i \in [n]$ when this product is displayed at location $q \in [p]$ on page $x \in [m]$.

To capture the product framing decisions, we use $f = \{f_{ixq} : i \in [n], x \in [m], q \in [p]\} \in \{0,1\}^{n \times m \times p}$ to denote the framing solution, where $f_{ixq} = 1$ if we offer product i at location q of page x; otherwise $f_{ixq} = 0$. Under framing solution f, the expected revenue from consumers who view x pages is

$$\frac{\sum_{i\in[n]} \sum_{l=1}^{x} \sum_{q\in[p]} f_{ilq} \nu_{ilq} r_i}{1 + \sum_{i\in[n]} \sum_{l=1}^{x} \sum_{q\in[p]} f_{ilq} \nu_{ilq}}.$$

Here, we have assumed without loss of generality that the preference weight for the no-purchase option is 1.

Under the MNL choice model, the upper-bound problem (2) is formulated as

$$\begin{aligned} U(x) = \max_{f} \quad & \frac{\sum_{i\in[n]} \sum_{l\in[x]} \sum_{q\in[p]} f_{ilq} \nu_{ilq} r_i}{1 + \sum_{i\in[n]} \sum_{l\in[x]} \sum_{q\in[p]} f_{ilq} \nu_{ilq}} \\ \text{s.t.} \quad & \sum_{i\in[n]} f_{ilq} \le 1 \quad \forall l \in [x], q \in [p] \\ & \sum_{l\in[x]} \sum_{q\in[p]} f_{ilq} \le 1 \quad \forall i \in [n] \\ & f_{ilq} \in \{0,1\}, \quad \forall i \in [n], l \in [x], q \in [p] \end{aligned} \tag{7}$$

where the first set of constraints ensure that each product is offered in at most one location and the second set of constraints ensure that each location is used by at most one product. The constraint matrix is that of an assignment problem, which is totally unimodular (see Corollary 2.9 in Chapter III.1 of Nemhauser and Wolsey 1988). With the linear fractional objective function, we know the problem is easily solvable (see Davis, Gallego and Topaloglu 2013).

We prove the following generalization of Lemma 2:

Lemma 3. *Suppose $\nu_{ixq} \le \nu_{ix'q'}$ for all $i \in [n]$, $x, x' \in [m]$ and $q, q' \in [p]$ such that $x \cdot p + q > x' \cdot p + q'$, i.e., the preference weight will decrease if the product is displayed in a later position. Then $\frac{U(x)}{x} \le \frac{U(x')}{x'}$ for any $x, x' \in [m]$ and $x > x'$.*

With Lemma 3 and the NBUE property (Assumption A3), it immediately follows that the $\frac{6}{\pi^2}$ performance bound still holds.

Corollary 2. *Suppose $\nu_{ixq} \le \nu_{ix'q'}$ for all $i \in [n]$, $x, x' \in [m]$ and $q, q' \in [p]$ such that $x \cdot p + q > x' \cdot p + q'$. Then all bounds proved in Sections 4 continue to hold for the model with location preference.*

6. Product Framing with Type-Dependent Consumer Choice Models

In this section, we relax the requirement that the consumer choice model must be the same for all the values of X. This relaxation allows us to use the number of pages viewed to differentiate keen consumers, who like to explore more pages and have a stronger desire to buy, or picky consumers, who like to look around but are less likely to make purchases.

To this end, we say that a consumer is of *type* x if she views x pages. We allow the choice model to be type dependent. Accordingly, for a consumer of type x, we let $P_x(i,S)$ be the purchase probability of product i in assortment S. We will show that under mild assumptions, we can still design an algorithm to guarantee an expected revenue of 1/3 of the upper bound.

Define

$$R_x(S) = \sum_{i \in S} r_i P_x(i,S) \tag{8}$$

to be the expected revenue from presenting assortment S to a consumer of type x, for all $x \in [m]$.

We will still use $U(x)$ to denote the optimal revenue from the capacitated assortment problem for consumers of type x, namely,

$$\begin{aligned} U(x) = \max_{S \subseteq [n]} \; & R_x(S) \\ \text{s.t.} & |S| \le x \cdot p. \end{aligned} \tag{9}$$

It is easy to verify that Theorem 2 is still valid, i.e., $\mathbf{E}[U(X)]$ is still an upper bound on the optimal expected revenue, as we only need to replace $P(\cdot,\cdot)$ with $P_x(\cdot,\cdot)$ in the proof of Theorem 2.

We make the following assumptions:

Assumption B1 *$R_x(S) \le R_y(S)$ for all $x \le y$ and $S \subseteq [n]$ with $|S| \le x \cdot p$.*

Assumption B2 *$U(x)/x$ is decreasing in x, for $x \in [m]$.*

Assumption B3 *Same as Assumption A2.*

Assumption B4 *X has increasing failure rate (IFR).*

Assumption B1 was implicitly true in previous sections as $R_x(S) = R_y(S)$ when all consumers follow the same choice model. The motivation behind Assumption B1 is as follows. When there are keen consumers, who like to shop around and have a stronger desire to buy, we would expect $P_x(i,S) \le P_y(i,S)$ for $x < y$, which leads to Assumption B1. When there is a mix of keen and picky consumers, $P_x(i,S) \le P_y(i,S)$ may not hold, but Assumption B1 could still serve as a reasonable approximation.

Assumption B2 is weaker than Assumption A1 in the previous section because, according to Lemma 2, the former is a result of the latter.

Assumption B4 is sufficient for Assumption A3 (Shaked and Shanthikuma 2007). It is equivalent to $h(x) \equiv \frac{\lambda(x)}{\Lambda(x)}$ increasing in $x \in [m]$. Assumption B4 means that the probability that a consumer will view the next page is decreasing in the number of pages she has viewed.

To gauge the appropriateness of our algorithms in settings where our assumptions might not hold, we perform computational experiments in these settings in Section 8.1.2. The experiments indicate that our algorithms significantly outperform greedy heuristics even in settings where the choice model changes drastically with x and the monotonicity of $R_x(S)$ in x is violated.

6.1. Truncation (TRUNC) Algorithms

Under Assumptions B1 to B4, we propose a family of algorithms $\mathsf{TRUNC}(y)$, which simply optimize for the set of products to be included in the first y pages, for $y \in [m]$. In other words, $\mathsf{TRUNC}(y)$ truncates the number of pages to exactly y, and does not try to optimize for the placement of these products within the y pages. The idea is to cater only to consumers who will view at least y pages.

$\mathsf{TRUNC}(y)$ Algorithm (for $y \in [m]$):

1. Same as Step 1 of $\mathsf{NEST}(y)$ (i.e., setting $S(y)$ to be the assortment solution to problem (9)).
2. Use any heuristic to fill in the first y pages such that the set of all products in the first y pages is $S(y)$.
3. Leave pages $y+1, y+2, \ldots, m$ blank.

Therefore, the only difference between $\mathsf{TRUNC}(y)$ and $\mathsf{NEST}(y)$ is that given a choice of y, $\mathsf{TRUNC}(y)$ does not optimize the display in the first y pages.

Let TRUNC be the algorithm that chooses the y maximizing $V^{\mathsf{TRUNC}(y)}$. We next show that TRUNC achieves at least 1/3 of the optimal expected revenue.

The main idea of the proof is as follows. According to Assumption B1, we have $U(x) \geq U(y)$ for all $x \geq y$. Therefore, $U(y)\Lambda(y)$ is a lower bound on $V^{\mathsf{TRUNC}(y)}$, because every consumer who views $x \geq y$ pages yields expected revenue $U(x) \geq U(y)$. Consequently, $\max_{y \in [m]} U(y)\Lambda(y)$ is a lower bound on V^{TRUNC}. We will minimize the quantity $\max_{y \in [m]} U(y)\Lambda(y)$ over all IFR distributions of X and all increasing functions $U(\cdot)$ satisfying Assumption B2. We will scale $U(\cdot)$ without loss of generality so the upper bound $\mathbf{E}[U(X)]$ is normalized to 1. We will then show that the smallest value of $\max_{y \in [m]} U(y)\Lambda(y)$ is at least 1/3. Doing this is equivalent to proving the same lower bound on the

following optimization problem:

$$\begin{aligned}
\gamma \quad &= \quad \min_{U,\Lambda} \ \max_{x\in[m]} U(x)\Lambda(x) \\
\text{s.t.} \quad & 1 = \Lambda(1) \geq \Lambda(2) \geq \cdots \geq \Lambda(m) \geq 0, \\
& \Lambda(x+1)\Lambda(x-1) \leq \Lambda(x)^2, \quad \forall x = 2,3,\ldots,m-1, \\
& U(x) \leq U(x+1), \quad \forall x = 1,2,\ldots,m-1, \\
& \frac{U(x)}{x} \geq \frac{U(x+1)}{x+1}, \quad \forall x = 1,2,\ldots,m-1, \\
& \mathbf{E}[U(x)] = 1, \\
& U(x) \geq 0, \quad \forall x \in [m].
\end{aligned} \tag{10}$$

All the constraints follow the same logic as those in Problem (5). Note that the third constraint that $U(\cdot)$ be increasing is a necessary condition of Assumption B1.

The following theorem states our main result for TRUNC. Please refer to Appendix A.4 for the detailed analysis.

THEOREM 4. *Under Assumptions B1, B2, B3, and B4,* $V^{TRUNC} \geq 1/3V^{OPT}$.

7. Joint Pricing and Framing Problem

In practice, retailers care not only about how to select and display the products, but also how to price them to maximize the expected revenue. In this section, we consider the problem faced by a retailer who is jointly framing and pricing all products. The framing policy still determines a consumer's consideration set, i.e., the set of products in the first several pages the consumer views. But now the retailer also adopts a *pricing policy* that sets the price r_i for each product $i \in [n]$, which influences consumers' utility and thus influences the choice probability of the product.

We assume that the relationship between the pricing policy and consumers' choice probabilities is given by the MNL choice model. Specifically, we assume that the choice probability of a consumer with consideration set S is

$$P(i,S) = \begin{cases} \frac{e^{u_i}}{1+\sum_{k\in S} e^{u_k}}, & i \in S \\ 0, & \text{otherwise,} \end{cases}$$

where u_i is consumers' mean utility for product i. We assume that the mean utility of a product scales linearly with its price: $u_i = a_i - \beta r_i$, for all $i \in [n]$. Here, a_i is the price-independent quality of product i, $\beta > 0$ is the price sensitivity parameter, and r_i is the *pricing decision*. We use product 0 to refer to the outside alternative, and assume that $u_0 = 0$. This formulation is commonly used in the economics, marketing and psychology literature (Berry, Levinsohn and Pakes 1995; Fader and Hardie 1996; Shugan 1980).

Next, in Section 7.1, we establish structural results of an optimal solution to the joint pricing and framing problem. Based on these structural results, we propose an approximation algorithm for this joint optimization problem in Section 7.2.

7.1. Structural Results

We first consider the case where the framing policy is fixed and the only task is to find the optimal prices to maximize revenues. More precisely, suppose we are given a fixed framing policy that displays products $S(x) \subseteq [n]$ in the first x pages, for all $x \in [m]$. The pricing problem is

$$\max_{r=(r_1,r_2,\ldots,r_n)} \mathbf{E}[R(r|S(X))],$$

where

$$R(r|S) \equiv \sum_{i \in S} r_i P(i,S)$$

is the expected revenue from a consumer with consideration set S, when $(r_1,\ldots,r_n)$ is the pricing policy. We prove the following structural result.

THEOREM 5. *Assume that the framing policy is given. To maximize the expected revenue, all products in the same page should have the same price. That is, there are parameters $\theta_1,\ldots,\theta_m$ such that $r_i = \theta_x$ for all products i displayed on page x. Moreover, we must have $\theta_1 \le \theta_2 \le \cdots \le \theta_m$.*

This result extends the classical MNL pricing structure. It is known that the optimal pricing strategy under the MNL revenue-optimization problem is to set the same price for all products (Anderson, de Palma and Thisse 1992, Hopp and Xu 2005 and Gallego and Stefanescu 2011). Theorem 5 states that, in our framing problem, this constant-price structure is preserved for products in any fixed page.

The monotonicity of page-dependent prices θ_x can be explained as follows. In the classic MNL pricing problem, if a consumer is willing to view more products, hence expanding her consideration set, then an optimal pricing policy would charge her a higher price for every product. Our model inherits this property in the sense that, if a consumer expands her consideration set by viewing more pages, she will see new products having higher and higher prices.

This pricing structure is in sharp contrast to that in an *oligopoly market* (Arbatskaya 2007). Arbatskaya (2007) studies a model in which multiple retailers compete to sell the same product. They find that retailers in later positions of the list have to price lower in order to attract more demand. Our model, however, considers the *monopoly market*, in which there is no need for the single retailer to compete on prices.

As a result of Theorem 5, we can reduce the number of decision variables from n prices, one for each product, to m prices $(\theta_1,\theta_2,\ldots,\theta_m)$, one for each page. Let $P(T,S) = \sum_{i \in T} P(i,S)$ denote the probability that a consumer with consideration set S buys any product in T. The pricing problem in terms of $(\theta_1,\theta_2,\ldots,\theta_m)$ is

$$\max_{\theta=(\theta_1,\theta_2,\ldots,\theta_m)} \mathbf{E}[R(\theta|S(X))],$$

where

$$R(\theta|S(x)) \equiv \sum_{l=1}^{x} \theta_l P\big(S(l) \setminus S(l-1), S(x)\big)$$

is the expected revenue from a consumer who considers $S(x)$ (we let $S(0)=\emptyset$).

The first order condition yields the system of equations

$$\frac{\partial \mathbf{E}[R(\theta|S(X))]}{\partial \theta_x} = \sum_{l=x}^{m} \lambda(l) P\big(S(x) \setminus S(x-1), S(l)\big)\big(1-\beta\theta_x + \beta R(\theta|S(l))\big) = 0 \quad \forall x \in [m],$$

which are equivalent to

$$\theta_x = \frac{1}{\beta} + \frac{\sum_{l=x}^{m} \lambda(l) P\big(S(x) \setminus S(x-1), S(l)\big) R(\theta|S(l))}{\sum_{l=x}^{m} \lambda(l) P\big(S(x) \setminus S(x-1), S(l)\big)}, \quad \forall x \in [m].$$

In other words, the price for products listed on page x is a constant $1/\beta$ plus a weighted average of the expected revenue from consumers who view page x.

In general, the expected revenue $\mathbf{E}[R(\theta|S(X))]$ may not be jointly concave in the θ vector, as illustrated in the following example.

Example 1. There are two products and two pages. Each page can hold at most one product. Suppose that $\mathbf{P}[X=1]=56\%$ and $\mathbf{P}[X=2]=44\%$. Product 1 has quality $a_1=4$ and product 2 has $a_2=2$. The price sensitivity is $\beta=1$. Figure 1 shows the expected revenue $\mathbf{E}[R(\theta|S(X))]$ as a function of the two prices. We see that this function is not jointly concave.

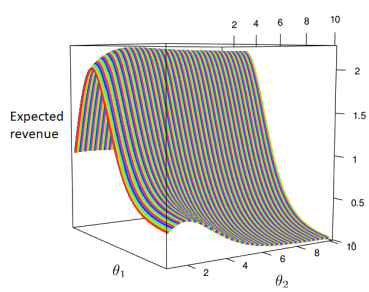

Figure 1 Pricing example.

We now look into the joint pricing and framing problem. By the presumed utility structure, we see that products are differentiated only by their quality parameters $a_i,\ i \in [n]$.

Theorem 6. *At optimality, each page is fully filled with products until all products are displayed. The products are displayed in descending order of quality a_i.*

In essence, Theorems 5 and 6 tell us that higher-quality products should be displayed first and given lower prices, with the lowest price enjoyed by products in the first page that is seen by everybody. One of the insights given by the two theorems is that lower-quality products should be charged higher prices so as to steer consumers to higher-quality products.

7.1.1. Profit-Maximization Problem. The structural results we have proved so far can be adapted to the profit-maximization problem, in which the expected *profit* from a consumer with consideration set S is $\sum_{i\in S}(r_i - c_i)P(i,S)$, where c_i is the cost of product i.

We can solve this profit-maximization problem by change of variables: $a_i - \beta c_i = a_i'$, $r_i - c_i = r_i'$. The resulting formulation in terms of a_i' and r_i' is that of the revenue-maximization problem (Gallego, Li and Beltran 2016). This naturally leads to the following result.

Corollary 3. *For the profit-maximization problem, at optimality, products are displayed in decreasing order of value gap, defined as $a_i - \beta c_i$ for all $i \in [n]$. Also, all products are priced with page-dependent markups. That is, $r_i - c_i = \theta_x$ for all products i displayed on page x. Moreover, the page-dependent markup θ_x increases with respect to the page index x.*

7.1.2. Location Preference Effect. Another relevant question is how to incorporate the location preference effect (see Section 5) into the joint pricing and framing problem. We can prove the following structural result based on Theorems 5 and 6.

Corollary 4. *Suppose that each product $i \in [n]$ has mean utility $a_i + \omega_{xq} - \beta r_i$ when it is placed at the q-th position on page x, where w_{xq} is the add-on utility for position (x,q), for all $x \in [m]$ and $q \in [p]$. Also suppose that $\omega_{xq} \le \omega_{x'q'}$ for all $i \in [n]$, $x, x' \in [m]$ and $q, q' \in [p]$ such that $x \cdot p + q > x' \cdot p + q'$. At optimality, each page is fully filled with products until all products are displayed. The products are displayed in descending order of their quality a_i. Moreover, all products on the same page have the same price (for the revenue maximization problem) or the same markup (for the profit optimization problem). The page-dependent price or markup increases with respect to the page index.*

7.2. Approximate Pricing Algorithm

In this section, we propose an efficient pricing heuristic, which we denote as NEST-P, that borrows the idea from NEST, and prove that the heuristic also guarantees a performance ratio $6/\pi^2$. The algorithm's idea is to pick a number y of pages, find a price that maximizes the revenue from consumers who view y pages, and set all products with this price.

NEST-P(y) Algorithm (for $y \in [m]$):

1. Display all products in decreasing order of their quality a_i, $i \in [n]$. Let $S(x)$ denote the set of products displayed in the first x pages, for all $x \in [m]$.

2. For all products in $S(y)$, solve an MNL pricing problem to maximize the expected revenue from consumers who view y pages.

3. For $x = y+1, \ldots, m$, solve an MNL pricing problem to maximize the expected revenue from consumers who view x pages, when the prices of products in $S(x-1)$ are fixed.

The first step of NEST-P(y) is essentially motivated by the structural result of Theorem 6. The second and third steps are designed so that the solution of NEST-P(y) inherits all the structures that we have proved for NEST(y). From Anderson, de Palma and Thisse (1992), we know that after applying NEST-P(y), all products in the first y pages have the same price, and products in the same page $x = y+1, \ldots, m$ have the same price.

Let NEST-P be the algorithm that picks the y maximizing the expected total revenue of NEST-P(y). We can easily derive the $6/\pi^2$ performance ratio of NEST-P by verifying that the same lower bound in Proposition 1 continues to hold.

THEOREM 7. *For the joint pricing and framing problem, the expected revenue of NEST-P is at least* $\frac{6}{\pi^2}$ *times the optimal expected revenue.*

8. Computational Experiments

In this section, we numerically test the performance of our framing and pricing algorithms, and compare them to other simple heuristics.

In all of our test problems, we assume that consumer choice is governed by the Multinomial Logit Model. If the set of products $S(x)$ is displayed in the first x pages, then conditional on a consumer with $X = x$, she will buy product $i \in S(x)$ with probability

$$\frac{v_i}{v_0 + \sum_{i' \in S(x)} v_{i'}}.$$

By convention, we set $v_0 = 1$ and $v_i = e^{a_i - \beta r_i}$, where a_i is the product quality.

8.1. Results of Framing Algorithms

We randomly generate $n = 300$ products in each test case. Every r_i is randomly drawn from a log-normal distribution such that $\log \frac{r_i}{50}$ has mean 0 and variance 1. We set $\beta = 1.02$ and $a_i = r_i + \epsilon_i$, where $\epsilon_i \in [-0.3, 0.3]$ is a noise added to the quality of product i.

We test three different distributions of X: geometric, uniform and Poisson. The distribution is truncated to $m = 20$ pages, so that $\lambda(20) = \Lambda(20)$. We further differentiate the test cases by different values of p and $\mathbf{E}[X]$. We report the result of each test case based on 1000 simulation replicates.

We test the following algorithms:

- NEST. Recall that NEST(y) leaves pages $x > y$ empty.
- NEST+ (Section 4.4), which is designed to improve the empirical performance of NEST.
- A sorting heuristic SORT_1, which first solves for S an assortment optimization problem with capacity $m \cdot p$, and then displays all products in S in increasing order of price.
- A sorting heuristic SORT_2, which is similar to SORT_1 but displays all products in S in decreasing order of mean utility.
- A sorting heuristic SORT_3, which is similar to SORT_1 but displays all products $i \in S$ in decreasing order of $r_i v_i$.
- A bottom-up greedy heuristic BU, which starts with the first page ($x = 1$) and sequentially fills in products that would maximize the expected revenue for consumers who view x pages, for $x = 2, \ldots, m$, such that this assortment includes all products in the first $x - 1$ pages.
- A top-down greedy heuristic TD, which starts by fitting all products into m pages. Then in the k-th step, $k = 1, \ldots, m-1$, the heuristic finds an assortment that maximizes revenue for consumers who view $m - k$ pages, from products in the first $m - k + 1$ pages.

Refer to Tables 1 to 3 for results of framing algorithms. We defer results of additional test cases, in which X has different distributions, to Tables 11 to 16 in Appendix B. In general, SORT_3 outperforms SORT_1 and SORT_2, and unsurprisingly NEST+ outperforms NEST. NEST+ dominates all other heuristics SORT_1, SORT_2, SORT_3, TD and BU in most test cases. The average optimality gap of NEST+ is just 1.42% in the worst case, compared to 12.41% for SORT_1, 12.25% for SORT_2, 10.85% for SORT_3, 10.81% for TD, and 3.51% for BU.

According to the numerical results, when both the number of products per page p and the expected number of pages seen $\mathbf{E}[X]$ become large, algorithms such as NEST, NEST+, TD and BU tend to achieve negligible optimality gaps. This can be explained as follows. As p increases, $U(y)$ would approach $U(m)$ for any $y \in [m]$, because consumers would be less likely to choose the no-purchase option when there are more products in the consideration set. On the other hand, $\frac{\mathbf{E}[\min(X,y)]}{y}$ tends to 1 as $\mathbf{E}[X]$ increases. Altogether, we can expect that the lower bound $U(y)\mathbf{E}[\min(X,y)]/y$ (see Proposition 1) tends to $U(m)$, an upper bound on the expected revenue from any consumer, as both p and $\mathbf{E}[X]$ increase. Consequently, as long as an algorithm tries to optimize the assortment in the first y pages for some properly chosen y, the optimality gap of the algorithm would be small in settings with large p and $\mathbf{E}[X]$.

8.1.1. Running Time Comparison. In Table 4, we report the running time of different framing algorithms (implemented using C# on a 4Ghz CPU), when the test scenario is the same as that of Table 2. The most time-consuming part of our algorithms is solving the cardinality constrained assortment optimization problem (2). Table 5 summarizes the number of times that each of the algorithms has to solve this assortment problem.

Table 1 Performance of framing algorithms when $\mathbf{E}[X] = 2$ and X follows a geometric distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	2.44%	0.78%	12.41%	11.25%	10.85%	10.76%	0.79%	6.6	20.0	20.0	20.0
3	2.54%	1.24%	11.99%	6.19%	5.35%	6.51%	0.92%	16.8	60.0	60.0	60.0
9	1.40%	0.77%	5.49%	3.03%	2.09%	1.36%	0.57%	65.4	72.3	71.8	74.8
15	0.43%	0.33%	3.55%	2.34%	1.68%	0.38%	0.26%	71.0	71.7	71.5	72.9

Table 2 Performance of framing algorithms when $\mathbf{E}[X] = 4$ and X follows a geometric distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	3.12%	1.14%	7.65%	6.57%	6.15%	6.10%	2.02%	12.6	20.0	20.0	20.0
3	2.73%	1.42%	7.79%	4.26%	3.38%	3.37%	2.09%	41.5	60.0	59.9	60.0
9	0.66%	0.54%	3.34%	2.01%	1.56%	0.63%	0.87%	71.4	71.4	71.4	74.5
15	0.21%	0.16%	1.95%	1.38%	1.06%	0.18%	0.25%	71.5	71.7	71.8	73.0

Table 3 Performance of framing algorithms when $\mathbf{E}[X] = 8$ and X follows a geometric distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	2.73%	1.12%	4.28%	3.42%	3.20%	3.11%	3.35%	18.8	20.0	20.0	20.0
3	1.73%	1.08%	4.71%	2.68%	2.09%	1.67%	2.92%	59.6	60.0	59.9	60.0
9	0.30%	0.28%	1.82%	1.20%	0.94%	0.30%	0.82%	71.7	72.1	71.7	74.7
15	0.10%	0.08%	1.04%	0.72%	0.58%	0.08%	0.23%	71.6	71.6	71.1	72.8

Table 4 Average running time (milliseconds) of framing algorithms when $\mathbf{E}[X] = 4$ and X follows a geometric distribution.

p	TRUNC	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU
1	7.1	7.3	16.6	< 1	< 1	< 1	< 1	9.1
3	7.0	7.7	19.5	< 1	< 1	< 1	1.3	8.2
9	6.9	8.4	18.7	< 1	< 1	< 1	< 1	2.9
15	7.3	8.4	16.3	< 1	< 1	< 1	< 1	1.6

Under the MNL choice model, we can solve (2) using a bisection method that iteratively finds the optimal expected revenue. In each iteration, the bisection method first makes a guess R on the optimal expected revenue (i.e., a guess on the optimal objective value of (2)), and then in $O(n)$ time the method looks for a subset of products i having the largest values of $(r_i - R)v_i$. These products help decide whether R is larger or smaller than the optimal expected revenue (Rusmevichientong, Shen and Shmoys 2010). The total time complexity of this bisection method is $O(n \log \rho)$, where $\log \rho$ is the number of bisection iterations and is determined by the lengths of the initial and final

search intervals. For instance, ρ can be

$$\frac{\max_{i\in[n]} r_i}{\max_{i\in[n]} \frac{r_i v_i}{v_0+v_i}},$$

where $\max_{i\in[n]} r_i$ and $\max_{i\in[n]} \frac{r_i v_i}{v_0+v_i}$ are upper and lower bounds, respectively, on the optimal objective value of (2). (We obtain the $O(n \log \rho)$ time complexity by using in each iteration a standard $O(n)$ algorithm for finding a subset of largest numbers; in practice, one may simply use a sorting method in each iteration to find a subset of largest numbers, which results in $O(n \log n \log \rho)$ total time complexity for the assortment optimization problem.)

Table 5 Number of times the cardinality-constrained assortment optimization problem needs to be solved.

TRUNC(y)	TRUNC	NEST(y)	NEST	NEST+(y)	NEST+	SORT$_1$, SORT$_2$, SORT$_3$	TD	BU
1	m	1	m	$O(m)$	$O(m^2)$	1	m	m

The time complexity of NEST is m times that of NEST(y), because in our implementation, NEST exhaustively computes NEST(y) for all $y \in [m]$, and then picks the best y. In practice, one could alternatively let NEST pick $y = \arg\max_{x\in[m]} \frac{U(x)}{x} \mathbf{E}[\min(X, x)]$ (i.e., pick the y maximizing the lower bound; see Proposition 1). In this way, NEST would keep the $6/\pi^2$ approximation ratio, and have the same time complexity as NEST(y). The same argument applies to TRUNC and NEST+.

In general, the running time of NEST is better than that of NEST+ because it is faster to solve (4) than the assortment optimization problem. Nevertheless, the time complexities of both NEST and NEST+ scale linearly with n, because the time complexity of solving the assortment optimization problem (2) scales linearly with n. Figure 2 illustrates how the average running time depends on n.

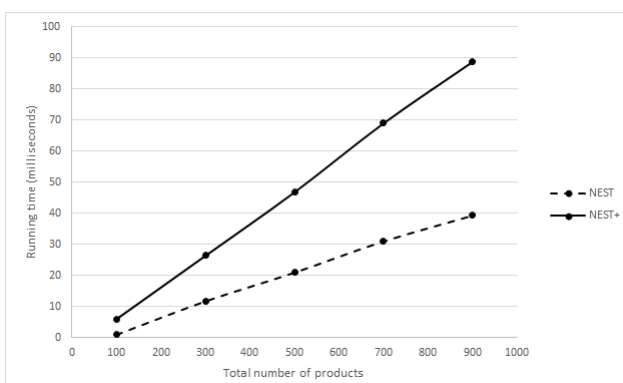

Figure 2 The average running time of NEST and NEST+, when n takes values in $\{100, 300, 500, 700, 900\}$. Other parameters are the same as those of Table 2.

8.1.2. Picky Consumers. We investigate in this section whether our algorithms continue to make good decisions in settings in which $v_0(x)$ increases linearly or exponentially in x. Specifically, in the linear case, we set $v_0(x) = 1 + (x-1)\omega$, for page index $x \in [m]$. In the exponential case, we set $v_0(x) = e^{(x-1)\omega}$, for page index $x \in [m]$. We explore the range of values $[0.2, 2]$ for ω in the linear case and $[0.1, 1]$ in the exponential case.

We can adapt NEST+ to this setting by choosing $v_0 = v_0(y)$ in computing NEST+(y). In this way, NEST+ inherits the structure of TRUNC. Since NEST+ is optimized for empirical performance, we could expect that NEST+ outperforms TRUNC in this setting.

Tables 6 and 7 display the results in the linear and exponential case, respectively. The results show that our leading algorithm NEST+ continues to have good performance in all cases. These experiments indicate that our model serves as a good approximation of the more complex setting where the choice model may change drastically with the consumer type. Thus, our model is a good starting point for an investigation of framing decisions involving heterogeneous consumers.

Table 6 Performance of framing algorithms when choice probabilities depend on X. $\mathbf{E}[X] = 4$; p = 3; X follows a Poisson distribution; $v_0(i) = 1 + (i-1)\omega$, where i is the page index.

	Avg gap							Max gap			
ω	TRUNC	NEST+	SORT$_1$	SORT$_2$	SORT$_3$	TD	BU	TRUNC	NEST+	TD	BU
0.2	1.01%	0.08%	7.34%	3.56%	2.44%	0.09%	0.44%	2.35%	0.54%	0.98%	2.76%
0.4	0.42%	0.02%	7.53%	3.71%	2.56%	0.02%	0.11%	1.20%	0.35%	0.21%	2.03%
0.6	0.18%	0.01%	7.33%	3.68%	2.59%	0.01%	0.03%	0.72%	0.08%	0.13%	0.48%
0.8	0.09%	0.00%	7.26%	3.56%	2.51%	0.00%	0.01%	0.39%	0.06%	0.07%	0.39%
1	0.05%	0.00%	6.82%	3.48%	2.46%	0.00%	0.00%	0.34%	0.05%	0.05%	0.19%
1.2	0.03%	0.00%	6.48%	3.40%	2.21%	0.00%	0.00%	0.22%	0.03%	0.04%	0.13%
1.4	0.02%	0.00%	6.19%	3.25%	2.07%	0.00%	0.00%	0.11%	0.03%	0.02%	0.05%
1.6	0.01%	0.00%	6.09%	3.18%	1.96%	0.00%	0.00%	0.08%	0.02%	0.02%	0.03%
1.8	0.01%	0.00%	5.92%	3.07%	1.80%	0.00%	0.00%	0.06%	0.02%	0.01%	0.09%
2	0.01%	0.00%	5.88%	3.03%	1.59%	0.00%	0.00%	0.05%	0.02%	0.01%	0.04%

8.2. Results of Joint Pricing and Framing Algorithms

In this section, we report test results of joint pricing and framing algorithms. We still choose $n = 300$ and $m = 20$. For product quality, we draw each a_i from $[50, 70]$ uniformly at random, so that no single product would have a much larger quality value compared to the rest. We test three different algorithms: NEST-P, NEST-P(m) and NEST-P(1). Notice that NEST-P(m) and NEST-P(1) are essentially top-down and bottom-up greedy heuristics, respectively.

Tables 8 to 10 summarize the test results of these algorithms. Please refer to Appendix B for more numerical results, in which X is assumed to follow different distributions. Among all test

Table 7 Performance of framing algorithms when choice probabilities depend on X. $\mathbf{E}[X]=4$; p = 3; X follows a Poisson distribution; $v_0(i)=e^{(i-1)\omega}$, where i is the page index.

	Avg gap							Max gap			
ω	TRUNC	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	TRUNC	NEST+	TD	BU
0.1	0.35%	0.13%	7.16%	3.50%	2.51%	0.14%	0.67%	0.87%	0.95%	1.15%	3.61%
0.2	0.24%	0.02%	7.57%	3.89%	2.79%	0.37%	0.13%	0.51%	0.18%	1.90%	1.49%
0.3	0.77%	0.03%	7.84%	4.07%	3.05%	1.09%	0.03%	1.42%	0.17%	4.67%	0.44%
0.4	1.45%	0.07%	7.17%	3.89%	2.86%	0.94%	0.02%	2.72%	0.34%	5.28%	0.22%
0.5	2.03%	0.13%	6.37%	3.61%	2.31%	0.68%	0.04%	3.34%	0.58%	2.84%	0.19%
0.6	2.30%	0.14%	5.98%	3.35%	1.84%	0.45%	0.06%	3.93%	0.73%	2.74%	0.39%
0.7	2.40%	0.14%	6.25%	3.79%	1.73%	0.33%	0.06%	5.18%	1.06%	2.66%	0.45%
0.8	2.37%	0.14%	7.11%	4.79%	2.05%	0.23%	0.06%	4.61%	0.96%	1.61%	0.59%
0.9	2.12%	0.12%	8.77%	6.70%	2.39%	0.20%	0.06%	4.56%	1.11%	1.46%	0.52%
1	1.95%	0.12%	10.78%	8.72%	2.38%	0.14%	0.05%	4.25%	1.00%	1.15%	0.44%

cases, the largest optimality gap of NEST-P is only 0.40%, compared to 9.02% for NEST-P(m) and 0.87% for NEST-P(1).

In all test scenarios, the average running time of NEST-P is less than two milliseconds, which is much smaller compared to NEST and NEST+. This is mainly because the MNL pricing problem can be reduced to a simple single-variable optimization problem. In particular, each time the common price of an assortment is updated by a search procedure, it takes only $O(1)$ time to re-compute the expected revenue of the assortment (Anderson, de Palma and Thisse 1992, Hopp and Xu 2005, Gallego and Stefanescu 2011).

Table 8 Performance of joint pricing and framing algorithms when $\mathbf{E}[X]=2$ and X follows a geometric distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.15%	5.94%	0.15%	0.17%	9.02%	0.17%
3	0.11%	2.18%	0.13%	0.15%	3.84%	0.16%
9	0.07%	0.34%	0.09%	0.10%	0.81%	0.11%
15	0.03%	0.10%	0.04%	0.05%	0.27%	0.08%

Table 9 Performance of joint pricing and framing algorithms when $\mathbf{E}[X]=4$ and X follows a geometric distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.24%	3.55%	0.36%	0.29%	5.16%	0.41%
3	0.17%	1.25%	0.27%	0.23%	2.36%	0.35%
9	0.06%	0.17%	0.13%	0.12%	0.45%	0.23%
15	0.03%	0.05%	0.06%	0.06%	0.13%	0.12%

Table 10 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 8$ and X follows a geometric distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.33%	2.01%	0.60%	0.40%	2.95%	0.70%
3	0.20%	0.67%	0.42%	0.28%	1.23%	0.55%
9	0.06%	0.09%	0.17%	0.13%	0.23%	0.27%
15	0.02%	0.03%	0.08%	0.05%	0.08%	0.15%

9. Conclusion and Future Work

In this work, we propose one of the first models of "framing effects" for pricing and assortment optimization. We introduce a model in which a set of products must be organized sequentially into a set of virtual pages and priced appropriately. Each consumers will only consider a random number of pages, and will select an product, if any, from these pages following a general choice model. We show that this product framing problem is NP-hard. We derive algorithms with guaranteed relative performance. Our algorithms are fast and easy to implement. We also show new structural results for pricing under framing effects. Directions for future research include to endogenize the number of pages consumers are willing to see. In the context of dynamic search, it would be convenient to allow for correlations between X and the choice model.

References

Agarwal, Ashish, Kartik Hosanagar, Michael D. Smith. 2011. Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of Marketing Research* **48**(6) 1057–1073. doi:10.1509/jmr.08.0468. URL `http://dx.doi.org/10.1509/jmr.08.0468`.

Amos Tversky, Daniel Kahneman. 1986. Rational choice and the framing of decisions. *The Journal of Business* **59**(4) S251–S278. URL `http://www.jstor.org/stable/2352759`.

Anderson, Simon P, Andre De Palma, Jacques François Thisse. 1992. *Discrete choice theory of product differentiation*. MIT press.

Animesh, Animesh, Siva Viswanathan, Ritu Agarwal. 2011. Competing "creatively" in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance. *Information Systems Research* **22**(1) 153–169. doi:10.1287/isre.1090.0254. URL `http://pubsonline.informs.org/doi/abs/10.1287/isre.1090.0254`.

Aouad, Ali, Danny Segev. 2016. Display optimization for vertically differentiated locations under multinomial logit choice preferences. Tech. rep., Working Paper.

Arbatskaya, Maria. 2007. Ordered search. *The RAND Journal of Economics* **38**(1) 119–126. URL `http://www.jstor.org/stable/25046295`.

Baye, Michael R., J. Rupert J. Gatti, Paul Kattuman, John Morgan. 2009. Clicks, discontinuities, and firm demand online. *Journal of Economics & Management Strategy* **18**(4) 935–975. doi:10.1111/j.1530-9134.2009.00234.x. URL `http://dx.doi.org/10.1111/j.1530-9134.2009.00234.x`.

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile Prices in Market Equilibrium. *Econometrica* **63**(4) 841–90. URL `https://ideas.repec.org/a/ecm/emetrp/v63y1995i4p841-90.html`.

Bront, J.J.M., I. Mendez-Diaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57**(3) 769–784.

Brynjolfsson, Erik, Astrid A. Dick, Michael D. Smith. 2010. A nearly perfect market? : differentiation vs. price in consumer choice.

Cachon, Gérard P., Christian Terwiesch, Yi Xu. 2005. Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* **7**(4) 330–346. doi:10.1287/msom.1050.0088. URL `http://dx.doi.org/10.1287/msom.1050.0088`.

Chandon, Pierre, J. Wesley Hutchinson, Eric T. Bradlow, Scott H. Young. 2009a. Does in-store marketing work? effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing* **73**(6) 1–17. doi:10.1509/jmkg.73.6.1. URL `http://dx.doi.org/10.1509/jmkg.73.6.1`.

Chandon, Pierre, J. Wesley Hutchinson, Eric T. Bradlow, Scott H. Young. 2009b. Does in-store marketing work? effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing* **73**(6) 1–17. doi:10.1509/jmkg.73.6.1. URL `http://dx.doi.org/10.1509/jmkg.73.6.1`.

Chiang, Jeongwen, Siddhartha Chib, Chakravarthi Narasimhan. 1998. Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* **89**(1–2) 223 – 248. doi:http://dx.doi.org/10.1016/S0304-4076(98)00062-1. URL `http://www.sciencedirect.com/science/article/pii/S0304407698000621`.

Corstjens, J., M. Corstjens. 2012. *The battlefield for Mindspace and Shelfspace*. John Wiley and Sons, Inc., 69–82. doi:10.1002/9781119207832.ch5. URL `http://dx.doi.org/10.1002/9781119207832.ch5`.

Davis, James, Guillermo Gallego, Huseyin Topaloglu. 2013. Assortment planning under the multinomial logit model with totally unimodular constraint structures. Tech. rep., Working paper.

Davis, James M., Guillermo Gallego, Huseyin Topaloglu. 2014. Assortment optimization under variants of the nested logit model. *Operations Research* **62**(2) 250–273. doi:10.1287/opre.2014.1256. URL `http://dx.doi.org/10.1287/opre.2014.1256`.

Davis, James M., Huseyin Topaloglu, David P. Williamson. 2015. Assortment optimization over time. *Operations Research Letters* **43**(6) 608 – 611. doi:http://dx.doi.org/10.1016/j.orl.2015.08.007. URL `http://www.sciencedirect.com/science/article/pii/S0167637715001108`.

Desir, Antoine, Vineet Goyal. 2013. Near-optimal algorithms for capacity constrained assortment optimization. Tech. rep., Working paper.

Feldman, J., H. Topaloglu. 2015. Assortment optimization under the multinomial logit model with nested consideration sets. Tech. rep., Working Paper.

Feng, Juan, Hemant K. Bhargava, David M. Pennock. 2007. Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms. *INFORMS Journal on Computing* **19**(1) 137–148. doi:10.1287/ijoc.1050.0135. URL `http://dx.doi.org/10.1287/ijoc.1050.0135`.

Gallego, Guillermo, Anran Li, Jose Beltran. 2016. Product line design and pricing under logit model. Tech. rep., working paper.

Gallego, Guillermo, Catalina Stefanescu. 2009. Upgrades, upsells and pricing in revenue management. Tech. rep., Working Paper.

Gallego, Guillermo, Huseyin Topaloglu. 2014. Constrained assortment optimization for the nested logit model. *Management Science* **60**(10) 2583–2601. doi:10.1287/mnsc.2014.1931. URL `http://dx.doi.org/10.1287/mnsc.2014.1931`.

Gallego, Guillermo, Ruxian Wang. 2014. Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* **62**(2) 450–461. doi:10.1287/opre.2013.1249. URL `http://dx.doi.org/10.1287/opre.2013.1249`.

Ghose, Anindya, Sha Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* **55**(10) 1605–1622. doi:10.1287/mnsc.1090.1054. URL `http://dx.doi.org/10.1287/mnsc.1090.1054`.

Hauser, John R. 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach. *Operations Research* **26**(3) 406–421. URL `http://www.jstor.org/stable/169752`.

Hauser, John R., Steven P. Gaskin. 1984. Application of the defender consumer model. *Marketing Science* **3**(4) 327–351. doi:10.1287/mksc.3.4.327. URL `http://dx.doi.org/10.1287/mksc.3.4.327`.

Hopp, Wallace J., Xiaowei Xu. 2005. Product line selection and pricing with modularity in design. *Manufacturing & Service Operations Management* **7**(3) 172–187. doi:10.1287/msom.1050.0077. URL `http://dx.doi.org/10.1287/msom.1050.0077`.

Howard, J.A., J.N. Sheth. 1969. *The theory of buyer behavior*. Wiley marketing series, Wiley. URL `https://books.google.com.hk/books?id=HOmRAAAAIAAJ`.

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, Gerald L. Lohse. 2004. On the depth and dynamics of online search behavior. *Manage. Sci.* **50**(3) 299–308. doi:10.1287/mnsc.1040.0194. URL `http://dx.doi.org/10.1287/mnsc.1040.0194`.

Li, Hongmin, Woonghee Tim Huh. 2011. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management* **13**(4)

549–563. doi:10.1287/msom.1110.0344. URL http://pubsonline.informs.org/doi/abs/10.1287/msom.1110.0344.

Manrai, Ajay K, Rick L Andrews. 1998. Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions. *European Journal of Operational Research* **111**(2) 193 – 215. doi:http://dx.doi.org/10.1016/S0377-2217(98)00145-3. URL http://www.sciencedirect.com/science/article/pii/S0377221798001453.

Mittal, Shashi, Andreas S. Schulz. 2013. A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. *Operations Research* **61**(2) 386–397. doi:10.1287/opre.1120.1093. URL http://dx.doi.org/10.1287/opre.1120.1093.

Nemhauser, George L., Laurence A. Wolsey. 1988. *Integer and Combinatorial Optimization*. Wiley-Interscience, New York, NY, USA.

Peter S. Fader, Bruce G. S. Hardie. 1996. Modeling consumer choice among skus. *Journal of Marketing Research* **33**(4) 442–452. URL http://www.jstor.org/stable/3152215.

Sahin, Ozge, Ruxian Wang. 2014. The impact of consumer search cost on assortment planning and pricing. Tech. rep., Working Paper.

Shaked, Moshe, J. George Shanthikumar. 2007. *Stochastic Orders*. Springer New York. doi:10.1007/978-0-387-34675-5.

Shugan, Steven M. 1980. The cost of thinking. *Journal of Consumer Research* **7**(2) 99–111. URL http://www.jstor.org/stable/2489077.

Silk, Alvin J, Glen L Urban. 1978. Pre-test-market evaluation of new packaged goods: A model and measurement methodology. *Journal of marketing Research* 171–191.

Song, Jing-sheng, Zhengliang Xue. 2009. Demand management and inventory control for substitutable products. Tech. rep., Working Paper.

Talluri, K., G. Van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* **50**(1) 15–33.

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

Ward Hanson, Kipp Martin. 1996. Optimizing multinomial logit profit functions. *Management Science* **42**(7) 992–1003. URL http://www.jstor.org/stable/2634363.

Appendix A: Collection of Proofs

A.1. Proofs in Section 3

In order to prove Theorem 1, we define the following function $f(\cdot,\cdot)$ for a fixed integer $d \geq 1$:

$$f(l,\theta) \equiv 0.5\frac{59(1-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}+0.5\frac{59(2-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}$$

for all $l = 0,1,\ldots,2d$ and $\theta \in \mathbb{R}$. The idea of the proof is to design a special case of the framing problem in which the expected revenue can be expressed using the $f(\cdot,\cdot)$ function. We first establish some technical properties of the $f(\cdot,\cdot)$ function in Lemmas 4, 5 and 6.

Lemma 4. *$f(l,0)+\frac{1}{200d^2} < f(d,0)$ for all $l \neq d$, $l \in \{0,1,...,2d\}$.*

Proof. Suppose l is relaxed to a non-negative continuous variable. Then the only solution for $\frac{\partial f(l,0)}{\partial l} = 0$ is

$$l = \frac{-21d-25000d^3+25d^2\left(-427+4\sqrt{-441-7350d+360000d^2}\right)}{7(3+25d)(-1+200d)} \geq 0.$$

It is easy to check that this is a local maximizer for $f(l,0)$. Therefore, $f(l,0)$ is quasi-concave in l for $l \geq 0$.

Given that l can only take non-negative integral values, we can deduce that, as long as $d \geq 1$,

$$f(d,0)-f(d-1,0) = \frac{-49+9600d+5000d^2}{2(1-200d+600d^2)(1-200d+800d^2)} > \frac{1}{200d^2},$$

$$f(d,0)-f(d+1,0) = \frac{-49+9600d+75000d^2}{2(-1+200d+600d^2)(-1+200d+800d^2)} > \frac{1}{200d^2}.$$

Therefore, since $f(l,0)$ has at most one local maximizer for $l \geq 0$, $l = d$ must be the unique maximizer for $f(l,0)$ when l is a non-negative integer. Consequently, the above two conditions ensure that $f(d,0)-f(l,0) > \frac{1}{200d^2}$ for all non-negative integers l such that $l \neq d$.

□

Lemma 5. *$f(d,\theta) < f(d,0)$ for all $\theta \neq 0$, $\theta \in [-1,1]$.*

Proof. Plugging $l = d$ into the definition of $f(l,\theta)$, we obtain

$$f(d,\theta) = \frac{8}{\theta+4}-\frac{4.5}{\theta+3}+34$$

$$\Longrightarrow \frac{\partial f(d,\theta)}{\partial\theta} = \frac{4.5}{(\theta+3)^2}-\frac{8}{(\theta+4)^2}.$$

Given $\theta \in [-1,1]$, the unique solution of $\partial f(d,\theta)/\partial\theta = 0$ is $\theta = 0$. It is easy to check that $\theta = 0$ is a maximizer for $f(d,\theta)$. This prove that $f(d,\theta) < f(d,0)$ for all $\theta \neq 0$ and $\theta \in [-1,1]$.

□

Lemma 6. *For all $l \neq d$, $l \in \{0, 1, ..., 2d\}$, and θ such that*

$$|\theta| \leq \frac{1}{2f(d,0)+34} \cdot \frac{1}{200d^2},$$

we have $f(l,\theta) < f(d,0)$.

Proof. It is easy to verify that, as long as $d \geq 1$,

$$\frac{1}{2f(d,0)+34} \cdot \frac{1}{200d^2} \leq 0.5.$$

Given $|\theta| \leq \frac{1}{2f(d,0)+34} \cdot \frac{1}{200d^2} \leq 0.5$ and $l \in \{0, 1, \ldots, 2d\} \setminus d$, we can deduce that

$$\begin{aligned}
f(l,\theta) =& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\left(\frac{l}{d}+\theta\right)}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 0.5\cdot 34|\theta| + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 0.5\cdot 34|\theta| \\
& \text{(because } 2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1 \text{ and } 3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1) \\
=& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta} + 34|\theta| \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}}\left(1+\frac{2|\theta|}{2-\frac{l-d}{200d^2}+\frac{l}{d}}\right) + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}\left(1+\frac{2|\theta|}{3-\frac{l-d}{200d^2}+\frac{l}{d}}\right) + 34|\theta| \\
& \text{(because } \theta \in [-0.5, 0.5]) \\
\leq& 0.5\frac{59(1-\frac{l-d}{200d^2})+34\frac{l}{d}}{2-\frac{l-d}{200d^2}+\frac{l}{d}}(1+2|\theta|) + 0.5\frac{59(2-\frac{l-d}{200d^2})+34\frac{l}{d}}{3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta}(1+2|\theta|) + 34|\theta| \\
& \text{(because } 2-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1 \text{ and } 3-\frac{l-d}{200d^2}+\frac{l}{d}+\theta \geq 1) \\
=& f(l,0) + (2f(l,0)+34)|\theta| \\
\leq& f(l,0) + (2f(d,0)+34)|\theta| \quad \text{(by Lemma 4)} \\
\leq& f(l,0) + (2f(d,0)+34)\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2} \\
=& f(l,0) + \frac{1}{200d^2} \\
<& f(d,0) \quad \text{(by Lemma 4).}
\end{aligned}$$

This completes the proof of the lemma.

□

Proof of Theorem 1.

Proof. Fix any instance of a 2-PARTITION problem with d numbers $w_1, w_2, ..., w_d$. We reduce this instance to a special case of our model with $m = 2$, $p = 200d^2$ and $n = 400d^2 + 2d$. The attractiveness of the 'no-purchase' option is 1. The revenues and attractiveness of the n products are as follows:

- Each of the first d products corresponds to a number in the two-partition problem. For $i = 1, 2, ..., d$, we set $r_i = 34$ and $\nu_i = M + \epsilon w_i$, where ϵ is some small value that we will define shortly, and M is determined by ϵ via

$$2d \cdot M + \epsilon \sum_{i=1}^{d} w_i = 2.$$

- For $i = d+1, d+2, ..., 2d$, we set $r_i = 34$ and $\nu_i = M$.
- For $i = 2d+1, 2d+2, ..., 3d$, we set $r_i = 59$ and $\nu_i = 2\delta$ where

$$\delta \equiv \frac{1}{p} = \frac{1}{200d^2}.$$

- For $i = 3d+1, 3d+2, ..., n$, we set $r_i = 59$ and $\nu_i = \delta$.

Given the special case constructed above, we argue that for an optimal framing solution, it is critical to decide which of the first $2d$ products with revenue 34 should be offered on the first page. We prove that the total expected revenue is a quasi-concave function of the total attractiveness of the first $2d$ products offered on the first page. In particular, if the 2-PARTITION problem has a solution, then we are able to recover that solution from the maximizer of the quasi-concave function (i.e., the optimal solution of our model). Therefore, we can solve the 2-PARTITION problem by optimizing the expected total revenue of our model.

We first observe the following structural properties of the special case of our model:

1. Since $r_i \leq 59$ for all $i \in [n]$, the optimal expected revenue must be strictly less than 59 (due to the no-purchase option). Thus, it is never optimal to leave any space in the two pages unfilled, as there are plenty of products with revenue 59.

2. It is easy to check that it always improves revenue to greedily replace a product with revenue 59 and attractiveness δ by, if any, a *spare* product with revenue 59 and attractiveness 2δ. Furthermore, whenever there is a product with revenue 59 and attractiveness δ on the first page and a product with revenue 59 and attractiveness 2δ on the second page, it is better to greedily swap the two products. Thus, products with revenue 59 and attractiveness 2δ should all be put on the first page.

3. Starting with an optimal solution, if we remove all the (at most $2d$) products with revenue 34 from the first two pages, we end up with at least $2p - 2d$ products with revenue 59 remaining in the two pages (among which only d products have attractiveness 2δ). Since $2p - 2d \approx 2p$, it is easy to check that the resulting expected revenue for consumers who view two pages is at least

$$\frac{59 \cdot d \cdot 2\delta + 59 \cdot (2p - 2d) \cdot \delta}{1 + d \cdot 2\delta + (2p - 2d) \cdot \delta} \approx 39.33.$$

Thus, when we put those products with revenue 34 back into the solution, the expected revenue for consumers who view two pages must be strictly greater than 34. This implies that in the optimal solution, no product with revenue 34 should be put on the second page, because 34 is lower than the expected revenue of the assortment consisting of products on the two pages.

In summary, an optimal framing solution must (i) put some of the products with revenue 34 on the first page, (ii) put all products with revenue 59 and attractiveness 2δ on the first page, and (iii) fill in all other spots using products with revenue 59 and attractiveness δ.

In the rest of this proof, we will focus on framing solutions that satisfy the above conditions. Then each framing solution can be characterized by a set $S \subseteq \{1, 2, ..., 2d\}$ of products with revenue 34 that are put in the first page.

We set $\lambda(1) = \lambda(2) = 0.5$. The total expected revenue under decision S is

$$\begin{aligned} R(S) =& \lambda(1)\frac{34\sum_{i\in S}\nu_i + 59\cdot 2\delta\cdot d + 59\cdot\delta\cdot(p-|S|-d)}{1+\sum_{i\in S}\nu_i + 2\delta\cdot d + \delta\cdot(p-|S|-d)} + \lambda(2)\frac{34\sum_{i\in S}\nu_i + 59\cdot 2\delta\cdot d + 59\cdot\delta\cdot(2p-|S|-d)}{1+\sum_{i\in S}\nu_i + 2\delta\cdot d + \delta\cdot(2p-|S|-d)} \\ =& 0.5\frac{59(1-(|S|-d)\delta) + 34\sum_{i\in S}\nu_i}{2-(|S|-d)\delta + \sum_{i\in S}\nu_i} + 0.5\frac{59(2-(|S|-d)\delta) + 34\sum_{i\in S}\nu_i}{3-(|S|-d)\delta + \sum_{i\in S}\nu_i}. \end{aligned} \tag{11}$$

By definition of $f(\cdot,\cdot)$, we have

$$\begin{aligned} R(S) &= f(|S|, \sum_{i\in S}\nu_i - |S|/d) \\ &= f(|S|, |S|M - |S|/d + \sum_{i\in S\cap\{1,2,\ldots,d\}} \epsilon w_i) \\ &= f(|S|, |S|\frac{2-\epsilon\sum_{i=1}^d w_i}{2d} - |S|/d + \epsilon\sum_{i\in S\cap\{1,2,\ldots,d\}} w_i) \\ &= f(|S|, \epsilon\left[\frac{-|S|\sum_{i=1}^d w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]). \end{aligned}$$

We set

$$\epsilon = \frac{1}{2\sum_{i=1}^d w_i}\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}.$$

Then,

$$\begin{aligned} &\left|\epsilon\left[\frac{-|S|\sum_{i=1}^d w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]\right| \\ =& \left|\frac{1}{2\sum_{i=1}^d w_i}\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}\left[\frac{-|S|\sum_{i=1}^d w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right]\right| \\ \le& \frac{1}{2\sum_{i=1}^d w_i}\cdot\frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}\left[\sum_{i=1}^d w_i + \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i\right] \\ \le& \frac{1}{2f(d,0)+34}\cdot\frac{1}{200d^2}. \end{aligned}$$

By Lemmas 4, 5 and 6, we have

$$R(S) = f(|S|, \sum_{i\in S}\nu_i - |S|/d)\begin{cases} = f(d,0), & \text{if } |S| = d, \sum_{i\in S}\nu_i = 1 \\ < f(d,0), & \text{otherwise.} \end{cases} \tag{12}$$

This implies that $R(S) = f(d,0)$ if and only if $|S| = d, \sum_{i\in S} \nu_i = 1$. Furthermore, if $R(S) = f(d,0)$, then S is an optimal solution to our model (not vice versa).

Let S^* be an optimal solution to our model. Finally, we prove the theorem by showing that the following two conditions are equivalent:

- $R(S^*) = f(d,0)$, i.e., the optimal expected revenue of our model is $f(d,0)$.
- The 2-PARTITION problem has a solution.

This equivalence helps reduce the framing problem to the 2-PARTITION problem as follows. If we can solve the framing problem, then we can find the optimal expected revenue. If the optimal expected revenue is $f(d,0)$, then we can conclude that there is a solution to the 2-PARTITION problem. If the optimal value is not $f(d,0)$, then we can conclude that there is no solution to the 2-PARTITION problem.

To prove the equivalence, first suppose that the 2-PARTITION problem has a solution $T \subseteq \{1,2,...,d\}$ such that

$$\sum_{i\in T} w_i = \frac{1}{2}\sum_{i=1}^{d} w_i.$$

We construct a solution $S \subset \{1,2,...,2d\}$ to our model as

$$S = T \cup \{d+1, d+2, ..., 2d - |T|\}.$$

We can check that $|S| = d$ and

$$\sum_{i\in S} \nu_i = dM + \sum_{i\in T} \epsilon w_i = dM + \frac{1}{2}\epsilon \sum_{i=1}^{d} w_i = 1.$$

Therefore, according to (12), S is an optimal solution to our model that gives expected revenue $R(S) = f(d,0)$.

On the other hand, suppose S^* is an optimal solution to our model and $R(S^*) = f(d,0)$. According to (12), we must have $|S^*| = d$ and $\sum_{i\in S^*} \nu_i = 1$, which gives

$$\begin{aligned}
&\sum_{i\in S} \nu_i = 1 \\
\Longrightarrow\ & |S|M + \sum_{i\in S\cap\{1,2,\ldots,d\}} \epsilon w_i = 1 \\
\Longrightarrow\ & d\frac{2 - \epsilon\sum_{i=1}^{d} w_i}{2d} + \sum_{i\in S\cap\{1,2,\ldots,d\}} \epsilon w_i = 1 \\
\Longrightarrow\ & \sum_{i\in S\cap\{1,2,\ldots,d\}} w_i = \frac{1}{2}\sum_{i=1}^{d} w_i.
\end{aligned}$$

This proves that the 2-PARTITION problem has a solution.

□

Proof of Theorem 2.

Proof. Suppose f^* is an optimal solution to (1). We must have, for any $x \in \mathcal{M}$,

$$U(x) \geq \sum_{i \in \mathcal{N}} r_i P(i, \{k \in \mathcal{N} : \sum_{l=1}^{x} f^*_{kl} = 1\})$$

$$\Longrightarrow \sum_{x \in \mathcal{M}} \lambda(x) U(x) \geq V^{\mathsf{OPT}}.$$

□

Proof of Corollary 1.

Proof. If we apply $\mathsf{NEST}+(m)$, then $S(m)$ is the optimal solution to Problem (2) with cardinality constraint $m \cdot p$. Since the optimal assortments are nested, and $S(x)$ is optimally selected from $S(x+1)$ for all $x = 1, 2, \ldots, m-1$, we must have $R(S(x)) = U(x)$ for all $x \in [m]$. Therefore, the expected revenue achieved by $\mathsf{NEST}+(m)$ is equal to the upper bound $\mathbf{E}[U(X)]$. Consequently, $\mathsf{NEST}+(m)$ (and thus $\mathsf{NEST}+$) is optimal.

□

A.2. Proofs in Section 4

Proof of Proposition 1.

Proof. Fix any $y \in [m]$. Under algorithm $\mathsf{NEST}(y)$, let $\tilde{R}(x) = R(S(x))$ be the expected revenue from a consumer who views x pages, for all $x \in [m]$. According to (4), if $|S(x)| = x \cdot p$, then

$$\frac{\tilde{R}(x)}{x \cdot p} = \frac{\tilde{R}(x)}{|S(x)|} \geq \frac{\tilde{R}(x+1)}{|S(x+1)|} \geq \frac{\tilde{R}(x+1)}{(x+1)p}$$

$$\Longrightarrow \frac{\tilde{R}(x)}{x} \geq \frac{\tilde{R}(x+1)}{x+1}.$$

If $|S(x)| = |S(x+1)|$, then we must have $S(x) = S(x+1)$, and thus

$$\tilde{R}(x) = \tilde{R}(x+1) \Longrightarrow \frac{\tilde{R}(x)}{x} \geq \frac{\tilde{R}(x+1)}{x+1}.$$

Consequently, for all $x = 1, 2, \ldots, y-1$,

$$\frac{\tilde{R}(x)}{x} \geq \frac{\tilde{R}(y)}{y} = \frac{U(y)}{y}, \tag{13}$$

where the equality is because $S(y)$ is an optimal solution to problem (2), and thus $\tilde{R}(y)$ is the corresponding optimal expected revenue.

Since all consumers who view $x \geq y$ pages see the same assortment $S(y)$ with expected revenue $U(y)$, we can obtain

$$\begin{aligned} V^{\mathsf{NEST}(y)} &= \sum_{x\in[m]} \tilde{R}(x)\lambda(x) \\ &= \sum_{x=1}^{y-1} \tilde{R}(x)\lambda(x) + U(y)\Lambda(y) \\ &\geq \frac{U(y)}{y} \sum_{x=1}^{y-1} x\lambda(x) + U(y)\Lambda(y) \\ &= \frac{U(y)}{y} \left[\sum_{x=1}^{y-1} x\lambda(x) + y\Lambda(y) \right] \\ &= \frac{U(y)}{y} \mathbf{E}[\min(X,y)], \end{aligned} \tag{14}$$

where the inequality is by (13). □

Proof of Proposition 2.

Proof. For ease of notation, define $g(x) = \frac{U(x)}{x}\mathbf{E}[\min(X,x)]$. We want to prove that $g(x)$ is constant for all $x \in [m]$. Suppose for a contradiction that there is a smallest x, $x > 1$, such that $g(x-1) \neq g(x)$.

- First, consider the case when $g(x-1) < g(x)$. We can revise the function $U(\cdot)$ as follows. For all $y \in \{1, 2, \ldots, x-1\}$, we scale up $U(y)$ such that $U(y) \leftarrow (1+\epsilon)U(y)$ for some $\epsilon > 0$. For all $y \in \{x, x+1, \ldots, m\}$, we scale down $U(y)$ such that $U(y) \leftarrow (1-\epsilon')U(y)$ for some $\epsilon' > 0$. We properly choose ϵ and ϵ' to maintain $g(x-1) \leq g(x)$ and $\mathbf{E}[U(X)] = 1$. This revision keeps the solution $U(\cdot)$ feasible because the only constraint that might be violated is $U(x-1) \leq U(x)$. However, $g(x-1) < g(x)$ implies $U(x-1) < U(x)$, as $\frac{\mathbf{E}[\min(X,x-1)]}{x-1} \geq \frac{\mathbf{E}[\min(X,x)]}{x}$. Therefore, the constraint $U(x-1) \leq U(x)$ will not be violated as long as ϵ and ϵ' are small enough.

As a result, we can strictly decrease $\max_{y\in[m]} g(y)$, because we can strictly decrease $g(y)$ at points $y = x, x+1, \ldots, m$, and $g(1) = g(2) = \cdots = g(x-1) \leq g(x)$. This contradicts the condition that the original $U(\cdot)$ is an optimal solution to (5), as we can strictly decrease the value of $\gamma = \max_{y\in[m]} g(y)$.

- Next, consider the case when $g(x-1) > g(x)$. We can revise the function $U(\cdot)$ as follows. Let $\mathcal{Y} = \{x, x+1, ..., z\}$ denote the set of consecutive indices such that (recall that $U(\cdot)$ is non-decreasing due to the constraints of (5))

  — $U(y) = U(x)$ for all $y \in \mathcal{Y}$;

  — $z = m$ or $U(z) < U(z+1)$.

For indices $y \in \mathcal{Y}$, we scale up $U(y)$ such that $U(y) \leftarrow (1+\epsilon)U(y)$ for some $\epsilon > 0$. For indices $y \in [m] \setminus \mathcal{Y}$, we scale down $U(y)$ such that $U(y) \leftarrow (1-\epsilon')U(y)$ for some $\epsilon' > 0$. We properly choose the values of ϵ and ϵ' such that $\mathbf{E}[U(X)] = 1$ and $g(x-1) \geq g(x)$ are maintained.

The only two constraints that could potentially be violated by this revision are $\frac{U(x-1)}{x-1} \geq \frac{U(x)}{x}$ and $U(z) \leq U(z+1)$ (if $z < m$). However, $g(x-1) > g(x)$ implies $\frac{U(x-1)}{x-1} > \frac{U(x)}{x}$, which means the constraint $\frac{U(x-1)}{x-1} \geq \frac{U(x)}{x}$ will not be violated as long as ϵ and ϵ' are chosen small enough. Moreover, if $z < m$, we have $U(z) < U(z+1)$ before the revision by definition of $\mathcal{Y}$. So again, as long as ϵ and ϵ' are chosen small enough, the constraint $U(z) \leq U(z+1)$ will not be violated.

Since this revision strictly decreases $g(y)$ for all $y \in [m] \setminus \mathcal{Y}$, and $g(y) \leq g(x-1)$ for all $y \in \mathcal{Y}$, the revision strictly decreases $\max_{y\in[m]} g(y)$. Thus the optimality condition meets contradiction.

In sum, we have proved that the function $g(x) = \frac{U(x)}{x}\mathbf{E}[\min(X,x)]$ must be non-increasing and non-decreasing.

□

Proof of Proposition 3.

Proof. By Proposition 2, $U(x) = \gamma \frac{x}{\mathbf{E}[\min(Y,x)]}$, where Y has the same distribution as X. Then

$$1 = \mathbf{E}[U(X)] = \gamma \sum_{x\in[m]} \frac{x}{\mathbf{E}[\min(Y,x)]}\lambda(x) = \gamma\ \mathbf{E}\left[\frac{X}{\mathbf{E}[\min(X,Y)|X]}\right].$$

The result follows after dividing both sides by γ. □

Before proving Theorem 3, we first establish an upper bound on $\mathbf{E}[\frac{X}{\mathbf{E}[\min(X,Y)|X]}]$ in Lemma 7 and Corollaries 5 and 6.

Lemma 7. *If Z is an exponential random variable with mean $\mathbf{E}[Z] = \mathbf{E}[X]$, then $\mathbf{E}[g(X)] \leq \mathbf{E}[g(Z)]$ for any increasing convex function $g(\cdot)$.*

Proof. Let $\mu = \mathbf{E}[Z] = \mathbf{E}[X]$. Let $\bar{F}_X$ denote the complementary cumulative distribution function (CCDF) of X, and $\bar{F}_Z$ denote the CCDF of Z. Proving this lemma is equivalent to proving $\int_a^\infty \bar{F}_X(v)dv \leq \int_a^\infty \bar{F}_Z(v)dv = \mu e^{-a/\mu}$ for any $a \geq 0$ (see Shaked and Shanthikumar 2007). Since X has the NBUE property, we have $\mathbf{E}[X-t|X>t] \equiv \frac{\int_t^\infty \bar{F}_X(v)dv}{\bar{F}_X(t)} \leq \mu$ for all $t \geq 0$, which gives

$$\frac{\bar{F}_X(t)}{\int_t^\infty \bar{F}_X(v)dv} \geq \frac{1}{\mu}.$$

Notice that the left hand side can be written as $-d\ln \int_t^\infty \bar{F}_X(v)dv$. We integrate both sides over $t \in [0,a]$ to obtain

$$\ln \int_0^\infty \bar{F}_X(v)dv - \ln \int_a^\infty \bar{F}_X(v)dv \geq \frac{a}{\mu}.$$

Since $\ln \int_0^\infty \bar{F}_X(v)dv = \ln \mathbf{E}[X] = \ln \mu$, the above reduces to the desired expression:

$$\int_a^\infty \bar{F}_X(v)dv \leq \mu e^{-a/\mu} = \int_a^\infty \bar{F}_Z(v)dv.$$

□

Corollary 5. *If Z is an exponential random variable with mean $\mathbf{E}[Z] = \mathbf{E}[X]$, then $\mathbf{E}[\min(X,x)] \geq \mathbf{E}[\min(X,x)]$ for all $x \geq 0$.*

Proof. Applying Lemma 7 to X and Z, we obtain $\mathbf{E}[(X-x)^+] \leq \mathbf{E}[(Z-x)^+]$. Thus, $\mathbf{E}[\min(X,x)] = \mathbf{E}[X] - \mathbf{E}[(X-x)^+] \geq \mathbf{E}[Z] - \mathbf{E}[(Z-x)^+] = \mathbf{E}[\min(Z,x)]$.

□

Corollary 6. *If Y is independently and identically distributed as X, and W and Z are independent exponential random variables with mean $\mathbf{E}[W] = \mathbf{E}[Z] = \mathbf{E}[X]$, then*

$$\mathbf{E}\left[X/\mathbf{E}[\min(X,Y)|X]\right] \leq \mathbf{E}\left[W/\mathbf{E}[\min(Z,W)|W]\right].$$

Proof. Define $h(x) = x/\mathbf{E}[\min(Z,x)]$.

By Corollary 5, we have for all $x \in [m]$,

$$\frac{x}{\mathbf{E}[\min(Y,x)]} \leq \frac{x}{\mathbf{E}[\min(Z,x)]} = h(x)$$

$$\Longrightarrow \mathbf{E}\left[\frac{X}{\mathbf{E}[\min(Y,X)|X]}\right] \leq \mathbf{E}[h(X)].$$

Since $h(x)$ is increasing convex, it follows from Lemma 7 that

$$\mathbf{E}\left[\frac{X}{\mathbf{E}[\min(Y,X)|X]}\right] \leq \mathbf{E}[h(X)] \leq \mathbf{E}[h(W)] = \mathbf{E}\left[\frac{W}{\mathbf{E}[\min(Z,W)|W]}\right].$$

□

Proof of Theorem 3.

Proof. By Proposition 3 and Corollary 6,

$$\frac{1}{\gamma} = \max_{\Lambda} \mathbf{E}\left[\frac{X}{\mathbf{E}[\min(X,Y)|X]}\right] \leq \mathbf{E}\left[\frac{W}{\mathbf{E}[\min(Z,W)|W]}\right],$$

where W and Z are independent exponential random variables with the same mean, say μ.

Since Z is exponential with mean μ, we have $\mathbf{E}[\min(W,Z)|W] = \mu(1-\exp(-W/\mu))$. Substituting this expression into the denominator, we obtain

$$\mathbf{E}\left[\frac{W}{\mu(1-\exp(-W/\mu))}\right] = \int_0^\infty \frac{w/\mu \exp(-w/\mu)}{1-\exp(-w/\mu)} dw/\mu = \int_0^\infty \frac{ue^{-u}}{1-e^{-u}} du = \sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6},$$

where the first equality follows from the distribution of W, the second from the transformation $u = w/\mu$, the third equality is a well know result from calculus, and the last equality is an important problem in number theory, posed by Mengoli in 1644. This problem remained open for 90 years until Euler solved it in 1734 at the age of 28.

Therefore,

$$\frac{1}{\gamma} \leq \mathbf{E}\left[\frac{W}{\mathbf{E}[\min(Z,W)|W]}\right] = \frac{\pi^2}{6}$$

$$\Longrightarrow \gamma \geq \frac{6}{\pi^2}.$$

Since γ lower-bounds V^{NEST} when the upper bound $\mathbf{E}[U(X)]$ on V^{OPT} is normalized to 1, we must have

$$\frac{V^{\mathsf{NEST}}}{V^{\mathsf{OPT}}} \geq \frac{V^{\mathsf{NEST}}}{\mathbf{E}[U(X)]} \geq \gamma = \frac{6}{\pi^2}.$$

□

Proof of Proposition 4.

Proof. Let X, Y be geometrically distributed with mean $\frac{1}{1-\alpha}$. That is, $\mathbf{P}[X=x]=\mathbf{P}[Y=x]=\alpha^{x-1}(1-\alpha)$. Then we can write

$$\mathbf{E}\left[\frac{X}{\mathbf{E}[\min(X,Y)|X]}\right] = \sum_{y=1}^{\infty} \frac{\alpha^{y-1}(1-\alpha)^2 y}{1-\alpha^y}.$$

For $\alpha < 1$, we have

$$\begin{aligned}
\sum_{y=1}^{\infty} \frac{\alpha^{y-1}(1-\alpha)^2 y}{1-\alpha^y} &= (1-\alpha)^2 \sum_{y=1}^{\infty} \frac{y}{\alpha} \frac{\alpha^y}{1-\alpha^y} \\
&= (1-\alpha)^2 \sum_{y=1}^{\infty} \frac{y}{\alpha} \sum_{n=1}^{\infty} \alpha^{yn} \\
&= \frac{(1-\alpha)^2}{\alpha} \sum_{n=1}^{\infty} \alpha^n \sum_{y=1}^{\infty} y(\alpha^n)^{y-1} \\
&= \frac{(1-\alpha)^2}{\alpha} \sum_{n=1}^{\infty} \alpha^n \frac{d}{d\alpha^n} \frac{1}{1-\alpha^n} \\
&= \frac{(1-\alpha)^2}{\alpha} \sum_{n=1}^{\infty} \alpha^n \frac{1}{(1-\alpha^n)^2} \\
&= \sum_{n=1}^{\infty} \alpha^{n-1} \frac{(1-\alpha)^2}{(1-\alpha^n)^2}.
\end{aligned}$$

The above series is an increasing function of α. It is maximized both locally and globally at $\alpha = 1$. To find the limit at $\alpha = 1$, we use two applications of L'Hospital's rule to obtain $\lim_{\alpha\to 1} \sum_{n=1}^{\infty} \alpha^{n-1} \frac{(1-\alpha)^2}{(1-\alpha^n)^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. □

A.3. Proofs in Section 5

Proof of Lemma 3.

Proof. It is easy to see that the value of $U(x)$ satisfies the following equation:

$$U(x) = \sum_{i\in[n]} \sum_{l\in[x]} \sum_{q\in[p]} (r_i - U(x)) \nu_{ilq} f^*_{ilq},$$

where f^* is a framing solution that corresponds to an optimal assortment for Problem (7). Any suboptimal assortment yields the left hand side larger than the right hand side.

For any $x' < x$, from all the products offered in f^*, we can pick the $x' \times p$ products i having the largest values of $(r_i - U(x))\nu_{ilq}$. Let us denote this set of products by S. Then it must be true that

$$\sum_{i\in S}\sum_{l\in[x]}\sum_{q\in[p]}(r_i - U(x))\nu_{ilq}f^*_{ilq} \geq x'\frac{U(x)}{x}.$$

Let f' be the framing solution that keeps the positions of those products $i \in S$ that are already in the first x' pages, and moves all other products in S to the first x' pages. The products in S that are now displayed earlier have higher location preference weights. So we must have

$$\sum_{i\in S}\sum_{l\in[x']}\sum_{q\in[p]}(r_i - U(x))\nu_{ilq}f'_{ilq} \geq \sum_{i\in S}\sum_{l\in[x]}\sum_{q\in[p]}(r_i - U(x))\nu_{ilq}f^*_{ilq}.$$

Also, since $U(x')$ is the optimal expected revenue of an assortment displayed in the first x' pages, we must have

$$U(x') \geq \sum_{i\in S}\sum_{l\in[x']}\sum_{q\in[p]}(r_i - U(x'))\nu_{ilq}f'_{ilq}.$$

Finally, we have $U(x') \leq U(x)$ because $U(x')$ corresponds to a more constrained assortment optimization problem.

Altogether, we can obtain

$$\begin{aligned}
U(x') &\geq \sum_{i\in S}\sum_{l\in[x']}\sum_{q\in[p]}(r_i - U(x'))\nu_{ilq}f'_{ilq} \\
&\geq \sum_{i\in S}\sum_{l\in[x']}\sum_{q\in[p]}(r_i - U(x))\nu_{ilq}f'_{ilq} \\
&\geq \sum_{i\in S}\sum_{l\in[x]}\sum_{q\in[p]}(r_i - U(x))\nu_{ilq}f^*_{ilq} \\
&\geq x'\frac{U(x)}{x},
\end{aligned}$$

which completes the proof.

□

A.4. Proofs in Section 6

In this section, we prove Theorem 4 by characterizing the structure of an optimal solution to (10). Let (U, Λ) denote an optimal solution to (10). Without loss of generality, suppose that $\Lambda(x) > 0$ for all $x \in [m]$ (if this does not hold, we can reduce the value of m so that $\Lambda(m) > 0$).

Define $g(x) \equiv x\Lambda(x)$, $h(x) \equiv \lambda(x)/\Lambda(x)$ for all $x \in [m]$ (thus $h(x)$ is the failure rate). Also define

$$y \equiv \max \arg\max_{x\in[m]} U(x)\Lambda(x).$$

We first establish some elementary results concerning the function $g(x)$, which turns out to play an important role in the analysis.

LEMMA 8. *If X has an IFR distribution, then $g(x)$ is weakly unimodal. Furthermore,*

$$\max \arg\max_{x\in[m]} g(x) = \min\{x \in [m] : h(x) > 1/(x+1)\}.$$

Proof. By Assumption B4 and definition of $h(\cdot)$, we have $h(1) \leq \cdots \leq h(m) = 1$. Since $1/(x+1)$ is decreasing, there must exist some $z \in [m]$ such that

$$h(x) \leq 1/(x+1),\ \forall x = 1, \ldots, z-1,$$

$$h(x) > 1/(x+1),\ \forall x = z, \ldots, m.$$

A little algebra shows that $g(x) \leq g(x+1)$ if and only if $h(x) \leq 1/(x+1)$, for all $x = 1, 2, \ldots, m-1$. Therefore, we must have

$$g(x) \leq g(x+1),\ \forall x = 1, \ldots, z-1,$$

$$g(x) > g(x+1),\ \forall x = z, \ldots, m-1,$$

which proves that $g(\cdot)$ is weakly unimodal. It immediately follows that $z = \min\{x \in [m] : h(x) > 1/(x+1)\}$ is the largest maximizer of $g(\cdot)$.

□

Next, we show the structure of an optimal solution to Problem (10) in Propositions 5, 6 and 7.

PROPOSITION 5. $\frac{U(x)}{x} = \frac{U(y)}{y}$ *for all* $x = y, \ldots, m$.

Proof. We first prove the first statement. Recall that Assumption B2 states $\frac{U(x-1)}{x-1} \geq \frac{U(x)}{x}$ for all $x = y+1, \ldots, m$. If there is any $z \in \{y+1, \ldots, m\}$ such that $\frac{U(z-1)}{z-1} > \frac{U(z)}{z}$ (i.e., inequality is strict), we can revise $U(\cdot)$ as follows:

- $U(x) \leftarrow (1+\epsilon_1)U(x)$ for all $x = z, \ldots, m$,
- $U(x) \leftarrow (1-\epsilon_2)U(x)$ for all $x = 1, 2, \ldots, z-1$,

where $\epsilon_1 > 0$ and $\epsilon_2 > 0$ are small values such that all constraints of (10) are still satisfied.

Since y was the original largest maximizer of the objective $U(x)\Lambda(x)$ of (10), and we have reduced $U(x)$ for all $x \in [y] \subseteq [z-1]$, the revised $U(\cdot)$ function gives a smaller objective value of (10). This contradicts the fact that $U(\cdot)$ is an optimal solution. Therefore, we must have $\frac{U(x-1)}{x-1} = \frac{U(x)}{x}$ for all $x = y+1, \ldots, m$.

□

LEMMA 9. $y = \max \arg\max_{x\in[m]} g(x)$.

Proof. Since y is the largest maximizer of $U(x)\Lambda(x)$, we have $U(x)\Lambda(x) < U(y)\Lambda(y)$ for all $x > y$, and $U(x)\Lambda(x) \leq U(y)\Lambda(y)$ for all $x < y$.

For all $x < y$, we combine $U(x)\Lambda(x) \leq U(y)\Lambda(y)$ and Assumption B2 to obtain $x\Lambda(x) \leq y\Lambda(y)$. For all $x > y$, we combine $U(x)\Lambda(x) < U(y)\Lambda(y)$ and Proposition 5 to obtain $x\Lambda(x) < y\Lambda(y)$. Therefore, y is the largest maximizer of $g(x)$ over $x \in [m]$.

□

PROPOSITION 6. *$U(x)\Lambda(x) = U(y)\Lambda(y)$ for all $x = 1, \ldots, y$.*

Proof. Since y is the largest maximizer of $U(x)\Lambda(x)$, we have $U(x)\Lambda(x) \leq U(y)\Lambda(y)$ for all $x = 1, \ldots, y$.

Suppose for a contradiction that the proposition does not hold. Then let z be the smallest $x \in \{1, \ldots, y\}$ such that $U(x)\Lambda(x) < U(y)\Lambda(y)$. Since $U(z)\Lambda(z) < U(y)\Lambda(y)$ and $\Lambda(z) \geq \Lambda(y) \geq 0$, we must have $U(z) < U(y)$. Consider the following two cases depending on the value of z:

- If $z = 1$, then we could scale up $U(1)$ and scale down $U(x)$ for all $x = 2, \ldots, m$, while keeping $U(x)$ increasing in x and $\mathbf{E}[U(X)] = 1$. This reduces $\max_{x \in [m]} U(x)\Lambda(x)$, which contradicts the optimality of Problem (10).
- Suppose $z > 1$. We now argue that $U(z)/z = U(z-1)/(z-1)$. Otherwise, we could scale up $U(z)$ and scale down $U(x)$ for all $x \neq z$, without violating any constraints of Problem (10), and again get a contradiction. Since z is the smallest x such that $U(x)\Lambda(x) < U(y)\Lambda(y)$, we must have $U(z-1)\Lambda(z-1) > U(z)\Lambda(z)$. Altogether, we obtain

$$U(z-1)\Lambda(z-1) > U(z)\Lambda(z) = zU(z-1)/(z-1) \cdot \Lambda(z)$$

$$\Longrightarrow g(z-1) > g(z).$$

Because $z \leq y$ and $y = \max \arg\max_{x \in [m]} g(x)$ (by Lemma 9), $g(z-1) > g(z)$ contradicts the unimodality of $g(\cdot)$ given by Lemma 8.

Altogether, we have proved by contradiction that $U(z)\Lambda(z) = U(y)\Lambda(y)$.

□

PROPOSITION 7. *$h(x) = h(y)$ for all $x = y, \ldots, m-1$.*

Proof. This proposition implies that $\Lambda(x+1)\Lambda(x-1) = \Lambda(x)^2$ for all $x = y, \ldots, m-1$. If not, there is a z such that $\Lambda(z+1)\Lambda(z-1) < \Lambda(z)^2$, and we can increase $\Lambda(z+1), \ldots, \Lambda(m)$ by a small amount while maintaining the IFR property. This adjustment does not affect $\max_{x \in [m]} U(x)\Lambda(x)$ because $y \leq z$ is its largest maximizer. However, this adjustment has the effect of increasing $\mathbf{E}[U(X)]$. To maintain $\mathbf{E}[U(X)] = 1$ we would need to scale down the entire $U(\cdot)$ and in the process reduce $\max_{x \in [m]} U(x)\Lambda(x)$, again contradicting the optimality of Problem (10).

□

Proof of Theorem 4.

Proof. Let $\theta = h(y)$ and recall from Proposition 7 that $h(x) = \theta$ for all $x = y, \ldots, m-1$. This implies that $X-(y-1)|X \geq y$ takes value in $\{1, 2, ..., m-y+1\}$ and follows a truncated geometric distribution with success rate θ. Therefore, we have $\mathbf{E}[X|X \geq y] = y-1+\mathbf{E}[X-(y-1)|X \geq y] \leq y-1+1/\theta$.

Consequently,

$$
\begin{aligned}
1 = \mathbf{E}[U(X)] =& \sum_{x=1}^{y-1} U(x)\lambda(x) + \sum_{x=y}^{m} U(x)\lambda(x) \\
=& \sum_{x=1}^{y-1} \frac{U(y)\Lambda(y)}{\Lambda(x)}\lambda(x) + \sum_{x=y}^{m} U(x)\lambda(x) \quad \text{(by Proposition 6)} \\
=& \sum_{x=1}^{y-1} \frac{U(y)\Lambda(y)}{\Lambda(x)}\lambda(x) + \sum_{x=y}^{m} \frac{xU(y)}{y}\lambda(x) \quad \text{(by Proposition 5)} \\
=& U(y)\Lambda(y)\left[\sum_{x=1}^{y-1} h(x) + \frac{1}{y}\sum_{x=y}^{m} x\frac{\lambda(x)}{\Lambda(y)}\right] \\
\leq& U(y)\Lambda(y)\left[\sum_{x=1}^{y-1} h(y-1) + \frac{1}{y}\sum_{x=y}^{m} x\frac{\lambda(x)}{\Lambda(y)}\right] \quad \text{(by Assumption B4)} \\
=& U(y)\Lambda(y)\left[(y-1)h(y-1) + \frac{1}{y}\mathbf{E}[X|X > y-1]\right] \\
\leq& U(y)\Lambda(y)\left[(y-1)h(y-1) + \frac{1}{y}(y-1+1/\theta)\right].
\end{aligned}
$$

From Lemma 8 and Lemma 9, we know that $y = \min\{x \in [m] : h(x) > 1/(x+1)\}$. It follows that $\theta = h(y) > 1/(y+1)$ and $h(y-1) \leq 1/y$. Substituting back into the above expression, we obtain

$$
\begin{aligned}
1 = \mathbf{E}[U(X)] \leq& U(y)\Lambda(y)\left[(y-1)h(y-1) + \frac{1}{y}(y-1+1/\theta)\right] \\
\leq& U(y)\Lambda(y)\left[(y-1)/y + \frac{1}{y}(y-1+y+1)\right] \\
\leq& U(y)\Lambda(y)\cdot 3.
\end{aligned}
$$

$$
\Longrightarrow U(y)\Lambda(y) \geq 1/3.
$$

This proves the theorem because $U(y)\Lambda(y)$ is the optimal objective value of Problem (10).

□

A.5. Proofs in Section 7

Proof of Theorem 5.

Proof. Given the framing decision $S(x),\ \forall x \in [m]$, we maximize solely over prices. That is, we want to compute

$$\max_{r=(r_1,\ldots,r_n)} \sum_{x=1}^{m} \lambda(x) R(r|S(x)) = \sum_{x=1}^{m} \lambda(x) \sum_{i \in S(x)} r_i P\big(i, S(x)\big),$$

where $R(r|S(x))$ is the expected revenue from consumers with consideration set $S(x)$, and $P\big(i, S(x)\big) = \frac{\exp(a_i - \beta r_i)}{1+\sum_{k \in S(x)} \exp(a_k - \beta r_k)}$ is the probability of choosing product i if $i \in S(x)$. Taking the partial derivative of $P\big(i, S(x)\big)$ with respect to r_i and r_k, respectively, we have

$$\frac{\partial P\big(i, S(x)\big)}{\partial r_i} = \beta P\big(i, S(x)\big)\Big(P\big(i, S(x)\big) - 1\Big),$$

$$\frac{\partial P\big(i, S(x)\big)}{\partial r_k} = \beta P\big(i, S(x)\big) P\big(k, S(x)\big).$$

Taking the first order derivative of the expected revenue $R(r|S(x))$ with respect to r_i, we obtain

$$\begin{aligned}
\frac{\partial R(r|S(x))}{\partial r_i} &= P\big(i, S(x)\big) + r_i \frac{\partial P\big(i, S(x)\big)}{\partial r_i} + \sum_{k \neq i} r_k \frac{\partial P\big(k, S(x)\big)}{\partial r_i} \\
&= \beta P\big(i, S(x)\big)\{\frac{1}{\beta} + \sum_{k \in S(x)} r_k P\big(k, S(x)\big) - r_i\} \\
&= \beta P\big(i, S(x)\big)\{\frac{1}{\beta} + R(r|S(x)) - r_i\}.
\end{aligned}$$

Let $x(i)$ denote the page index where product i is displayed (let $x(i) = m+1$ if product i is not displayed in any of the m pages). Taking partial derivative of the total expected revenue with respect to r_i, we obtain

$$\frac{\partial \mathbf{E}[R(r|S(X))]}{\partial r_i} = \beta \sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big)\{\frac{1}{\beta} + R(r|S(l)) - r_i\}. \tag{15}$$

Setting the partial derivative to zero, we obtain

$$\sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big)\{\frac{1}{\beta} + R(r|S(l))\} = \sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big) r_i. \tag{16}$$

Notice that equation (16) is satisfied either when $P\big(i, S(l)\big) = 0,\ \forall l \in [m]$, (i.e., when $r_i = +\infty$, meaning product i is priced out of the market); or when

$$r_i = \frac{\sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big)\{\frac{1}{\beta} + R(r|S(l))\}}{\sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big)}. \tag{17}$$

Products with infinite prices can be equivalently viewed as they are not displayed at all. Thus, without loss of generality, we can assume that all the displayed products (i.e., all products in $S(m)$)

have finite prices. On the other hand, it is easy to check that in equation (17), the right hand side is the same for all i and k such that $x(i) = x(k)$. Thus, for every finitely priced product i, there must be a page-level invariant price $\theta_{x(i)}$ such that $r_i = \theta_{x(i)}$ at optimality.

For the monotonicity of the page-level prices, notice that equation (17) tells us that the price is a weighted average of $\frac{1}{\beta} + R(r|S(l))$, where $R(r|S(l))$, the expected revenue from consumers with consideration set $S(l)$, must be non-decreasing in l. To see this, suppose for a contradiction that $R(r|S(l-1)) > R(r|S(l))$ for some l. This could only be because we set prices too low for products in page l. Then we would just raise the prices of products in page l, and therefore increase the expected revenue from consumers with consideration set $S(l)$. Moreover, the expected revenue from consumers who view $l+1, \ldots, m$ pages can also be increased due to the unimodality of $R(r|S(l))$ with respect to each of the price values (Gallego, Li and Beltran 2016). Consequently, at optimality, $R(r|S(l))$ must be non-decreasing in l.

□

Proof of Theorem 6.

Proof. Suppose that the framing solution $S(x)$, for all $x \in [m]$, is fixed. Given the quality vector $a = (a_1, ..., a_n)$, we define $r(a)$ and $R(a)$ as an optimal pricing vector and its corresponding total expected revenue, respectively:

$$R(a) \equiv \sum_{x=1}^{m} \lambda(x) R(r(a)|S(x)) = \max_{r=(r_1,\ldots,r_n)} \sum_{x=1}^{m} \lambda(x) R(r|S(x)).$$

Now, we characterize how $R(a)$ changes with respect to the quality values a_i. Recall that we use $x(i)$ to denote the page where product i is displayed. According to the envelope theorem,

$$\begin{aligned} \frac{\partial R(a)}{\partial a_i} &= \sum_{l=x(i)}^{m} \lambda(l) \frac{\partial R(r(a)|S(l))}{\partial a_i} \\ &= \sum_{l=x(i)}^{m} \lambda(l) \{ r_i(a) - R\big(r(a)|S(l)\big) \} P\big(i, S(l)\big). \end{aligned} \tag{18}$$

Since $r(a)$ must satisfy the first-order condition given by (17), i.e., $\sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big) r_i(a) = \sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big) \{ \frac{1}{\beta} + R\big(r(a)|S(l)\big) \}$, we have

$$\begin{aligned} \frac{\partial R(a)}{\partial a_i} &= \sum_{l=x(i)}^{m} \lambda(l) \{ r_i(a) - R\big(r(a)|S(l)\big) \} P\big(i, S(l)\big) \\ &= \frac{1}{\beta} \sum_{l=x(i)}^{m} \lambda(l) P\big(i, S(l)\big) \geq 0. \end{aligned} \tag{19}$$

Thus, the total expected revenue increases in the quality of any displayed product. Since removing a product is equivalent to reducing its quality to negative infinity, it is never beneficial to remove

any product from the m pages. Furthermore, if any product i that is not displayed has a higher quality value than a displayed product k, we can replace product k with product i to increase the total expected revenue. Altogether, an optimal policy will either display all the products, or display the $p \cdot m$ products having the highest quality values.

To prove that products should be displayed in decreasing order of quality, suppose for a contradiction that there is a product k displayed on page x and a product i displayed on page $x+1$ such that $a_k < a_i$.

Consider two cases:

1. $a_k - \beta r_k(a) < a_i - \beta r_i(a)$. By Theorem 5, we have $r_k(a) \leq r_i(a)$, since product k is displayed prior to product i. We modify the solution by switching the positions of these two products without changing their prices. This only affects the expected revenue from consumers who view x pages. For these consumers, product k is replace with a product having higher expected utility and a higher price, so the expected revenue from them must increase. Therefore, the modification increases the total expected revenue, leading to a contradiction.

2. $a_k - \beta r_k(a) \geq a_i - \beta r_i(a)$. We must have $P(k, S(l)) \geq P(i, S(l))$ for all $l = x+1, \ldots, m$. Thus,

$$\frac{\partial R(a)}{\partial a_k} = \frac{1}{\beta} \sum_{l=x}^{m} \lambda(l) P\big(k, S(l)\big) \geq \frac{1}{\beta} \sum_{l=x+1}^{m} \lambda(l) P\big(i, S(l)\big) = \frac{\partial R(a)}{\partial a_i}.$$

This says that we can improve the total expected revenue $R(a)$ by simultaneously increasing a_k and decreasing a_i. After such modification, $a_k - \beta r_k(a) \geq a_i - \beta r_i(a)$ still holds and thus the above inequality still holds. Consequently, we can keep increasing a_k and decreasing a_i until the quality values of the two products are switched. This only increases the total expected revenue, again leading to a contradiction.

Altogether, we have proved that the quality values of products in any page x should be at least those of products in page $x+1$.

□

Proof of Corollary 4.

Proof. Given that the framing policy is fixed, the common price property can be proved using the same argument as in Theorem 5.

As for the ordering of product quality, we can slightly modify the proof of Theorem 6 as follows. Suppose for a contradiction there is a product i displayed at position (x, q) and a product i' displayed at position (x', q') such that $a_i < a_{i'}$ and $x \cdot p + q < x' \cdot p + q'$. Consider two cases:

1. $a_i + \omega_{xq} - \beta r_i < a_{i'} + \omega_{x'q'} - \beta r_{i'}$. Since position (x, q) is earlier than (x', q'), we have $w_{xq} \geq w_{x'q'}$, so $a_i + \omega_{xq} - \beta r_i < a_{i'} + \omega_{x'q'} - \beta r_{i'} \leq a_{i'} + \omega_{xq} - \beta r_{i'}$. If we switch the positions of the two products without changing their prices, consumers who view $x, \ldots, x'-1$ pages will see a product having

higher expected utility (increased from $a_i + \omega_{xq} - \beta r_i$ to $a_{i'} + \omega_{xq} - \beta r_{i'}$) and a higher price. Thus, the expected revenue from those consumers will increase, while the expected revenue from all other consumers will not change. Therefore, the total expected revenue will increase after the switching.

2. $a_i + \omega_{xq} - \beta r_i \geq a_{i'} + \omega_{x'q'} - \beta r_{i'}$. Conceptually, we can apply the same argument as in the proof of Theorem 6 by viewing $a_i + w_{xq}$ and $a_{i'} + w_{x'q'}$ as the new quality of products i and i', respectively. We simultaneously increase a_i and reduce $a_{i'}$ until the quality values are switched. Such modification only improves the total expected revenue.

Altogether, we have shown by contradiction that an optimal policy would always display all products in decreasing order of quality, even with the location preference effect.

□

Proof of Theorem 7.

Proof. From Theorem 6, we know that an optimal joint framing and pricing solution always displays all products in decreasing order of quality, until all pages are full or all products are displayed. Fix the framing policy, and let $S(x)$ denote the set of all products in the first $x \in [m]$ pages. Let $r^* = (r_1^*, \ldots, r_n^*)$ denote the optimal pricing vector. We want to show that

$$V^{\mathsf{NEST\text{-}P}} \geq \frac{6}{\pi^2} \sum_{x \in [m]} \lambda(x) R(r^* | S(x)).$$

We can naturally re-define

$$U(x) = \max_{r=(r_1,\ldots,r_n)} R(r|S(x)) \tag{20}$$

as an upper bound on $R(r^*|S(x))$.

It then suffices to show $V^{\mathsf{NEST\text{-}P}} \geq \frac{6}{\pi^2} \sum_{x \in [m]} \lambda(x) U(x)$. To this end, it remains to verify that (i) the lower bound in Proposition 1 holds for $\mathsf{NEST\text{-}P}(y)$, and (ii) all the constraints in (5) regarding $U(x)$ are still satisfied. Let $r^{\mathsf{NEST\text{-}P}(y)}$ denote the pricing vector set by $\mathsf{NEST\text{-}P}(y)$.

• We first argue that the lower bound in Proposition 1 still holds. For consumers who view y pages, we have precisely $R(r^{\mathsf{NEST\text{-}P}(y)}|S(y)) = U(y)$ by definition of $\mathsf{NEST\text{-}P}(y)$. For consumers who view $x < y$ pages, since all products in the first y pages have the same price and products on earlier pages have higher quality, we must have

$$R(r^{\mathsf{NEST\text{-}P}(y)}|S(x)) \geq \frac{x}{y} R(r^{\mathsf{NEST\text{-}P}(y)}|S(y)) = \frac{x}{y} U(y). \tag{21}$$

From Gallego, Li and Beltran (2016), we know that the price of products set by $\mathsf{NEST\text{-}P}(y)$ in pages $x = y+1, \ldots, m$ increases in x, as the sum of quality of all products considered by consumers increases in x. Therefore, $R(r^{\mathsf{NEST\text{-}P}(y)}|S(x))$ increases in x. Consequently, $R(r^{\mathsf{NEST\text{-}P}(y)}|S(x)) \geq R(r^{\mathsf{NEST\text{-}P}(y)}|S(y)) = U(y)$ for all $x \geq y$.

Altogether,

$$\begin{aligned} V^{\mathsf{NEST\text{-}P}(y)} &= \sum_{x\in[m]} \lambda(x) R(r^{\mathsf{NEST\text{-}P}(y)} | S(x)) \\ &\geq \sum_{x=1}^{y} \lambda(x) \frac{x}{y} U(y) + \sum_{x=y+1}^{m} U(y) \\ &= \frac{U(y)}{y} \mathbf{E}[\min(X, y)]. \end{aligned}$$

• Next we verify that all the constraints in (5) regarding $U(\cdot)$ are still satisfied. Clearly $U(x)$ is increasing in x. Also, after solving the upper bound problem (20), the expected revenue of the assortment consisting of products in the first $x-1$ pages must be at least $\frac{x-1}{x}U(x)$, because all the products have the same price and products in the first $x-1$ pages have higher quality values. It follows that $U(x-1) \geq \frac{x-1}{x}U(x)$. Thus $U(x)/x$ is decreasing in x.

Altogether, Problem (5) still serves as a bound-revealing problem if we normalize $\mathbf{E}[U(X)] = 1$. So the approximation ratio of NEST-P is at least $6/\pi^2$.

□

Appendix B: Additional Tables

Table 11 Performance of framing algorithms when $\mathbf{E}[X] = 2$ and X follows a uniform distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	0.56%	0.43%	12.28%	10.76%	10.89%	10.81%	0.71%	3.0	20.0	20.0	20.0
3	0.81%	0.69%	11.22%	5.97%	4.62%	5.94%	0.87%	9.0	60.0	60.0	60.0
9	0.60%	0.57%	5.45%	3.10%	2.29%	0.94%	0.63%	27.0	71.8	72.0	74.8
15	0.31%	0.24%	3.27%	2.37%	1.86%	0.24%	0.39%	45.0	71.9	71.4	73.3

Table 12 Performance of framing algorithms when $\mathbf{E}[X] = 4$ and X follows a uniform distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	1.22%	0.66%	7.27%	6.07%	5.34%	5.29%	1.93%	6.9	20.0	20.0	20.0
3	1.21%	0.93%	6.96%	3.88%	2.91%	2.50%	2.20%	20.7	60.0	60.0	60.0
9	0.37%	0.35%	2.63%	1.80%	1.46%	0.34%	1.01%	62.9	71.8	71.5	74.4
15	0.12%	0.09%	1.33%	0.95%	0.76%	0.09%	0.25%	72.0	71.9	71.8	72.7

Table 13 Performance of framing algorithms when $\mathbf{E}[X] = 8$ and X follows a uniform distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	1.38%	0.78%	3.53%	2.72%	2.47%	2.28%	3.49%	14.6	20.0	20.0	20.0
3	0.97%	0.79%	3.80%	2.36%	1.90%	0.95%	3.30%	44.7	60.0	60.0	60.0
9	0.16%	0.15%	1.13%	0.78%	0.63%	0.15%	0.88%	71.5	71.8	71.6	74.8
15	0.05%	0.04%	0.59%	0.43%	0.35%	0.04%	0.19%	71.8	71.8	71.5	73.0

Table 14 Performance of framing algorithms when $\mathbf{E}[X] = 2$ and X follows a Poisson distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	1.45%	0.51%	12.20%	11.02%	10.57%	10.41%	0.72%	5.2	20.0	20.0	20.0
3	1.55%	0.88%	11.53%	5.82%	4.83%	5.86%	0.93%	14.2	60.0	60.0	60.0
9	0.98%	0.66%	5.24%	3.08%	2.27%	1.08%	0.69%	58.7	71.9	71.6	74.9
15	0.34%	0.26%	3.36%	2.33%	1.85%	0.30%	0.32%	72.0	71.5	71.6	73.3

Table 15 Performance of framing algorithms when $\mathbf{E}[X] = 4$ and X follows a Poisson distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	1.13%	0.48%	7.32%	5.79%	5.56%	5.62%	1.81%	8.9	20.0	20.0	20.0
3	1.03%	0.62%	6.81%	3.49%	2.50%	2.23%	2.23%	27.5	60.0	60.0	60.0
9	0.21%	0.19%	2.61%	1.98%	1.65%	0.19%	1.17%	71.5	71.9	71.9	74.6
15	0.05%	0.04%	1.05%	0.82%	0.75%	0.04%	0.30%	71.9	71.5	71.6	73.1

Table 16 Performance of framing algorithms when $\mathbf{E}[X] = 8$ and X follows a Poisson distribution.

	Avg gap							Avg number of displayed products			
p	NEST	NEST+	SORT_1	SORT_2	SORT_3	TD	BU	NEST	NEST+	TD	BU
1	0.67%	0.29%	3.22%	2.33%	2.03%	1.76%	3.37%	14.7	20.0	20.0	20.0
3	0.37%	0.25%	3.38%	2.29%	1.83%	0.40%	3.51%	55.1	60.0	60.0	60.0
9	0.01%	0.01%	0.44%	0.38%	0.35%	0.01%	0.86%	71.5	71.8	71.9	74.4
15	0.00%	0.00%	0.07%	0.05%	0.05%	0.00%	0.15%	72.1	71.7	71.9	72.4

Table 17 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 2$ and X follows a uniform distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.14%	4.95%	0.16%	0.17%	7.34%	0.19%
3	0.11%	1.77%	0.14%	0.15%	3.06%	0.17%
9	0.06%	0.25%	0.08%	0.11%	0.60%	0.12%
15	0.03%	0.07%	0.04%	0.06%	0.18%	0.08%

Table 18 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 4$ and X follows a uniform distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.21%	2.62%	0.41%	0.26%	3.83%	0.47%
3	0.15%	0.83%	0.32%	0.21%	1.72%	0.41%
9	0.05%	0.11%	0.15%	0.10%	0.21%	0.24%
15	0.02%	0.03%	0.07%	0.05%	0.08%	0.15%

Table 19 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 8$ and X follows a uniform distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.27%	1.28%	0.69%	0.35%	1.87%	0.80%
3	0.16%	0.40%	0.49%	0.23%	0.85%	0.61%
9	0.04%	0.05%	0.20%	0.08%	0.10%	0.31%
15	0.01%	0.01%	0.09%	0.03%	0.04%	0.17%

Table 20 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 2$ and X follows a Poisson distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.16%	5.27%	0.16%	0.18%	7.99%	0.18%
3	0.12%	1.87%	0.13%	0.17%	3.36%	0.16%
9	0.06%	0.27%	0.07%	0.11%	0.64%	0.12%
15	0.03%	0.07%	0.04%	0.07%	0.18%	0.09%

Table 21 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 4$ and X follows a Poisson distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.12%	2.04%	0.42%	0.15%	3.27%	0.49%
3	0.08%	0.58%	0.33%	0.11%	1.10%	0.41%
9	0.03%	0.06%	0.17%	0.06%	0.12%	0.28%
15	0.01%	0.01%	0.08%	0.03%	0.03%	0.16%

Table 22 Performance of joint pricing and framing algorithms when $\mathbf{E}[X] = 8$ and X follows a Poisson distribution.

	Avg gap			Max gap		
p	NEST-P	NEST-P(m)	NEST-P(1)	NEST-P	NEST-P(m)	NEST-P(1)
1	0.06%	0.45%	0.74%	0.08%	0.73%	0.87%
3	0.02%	0.09%	0.54%	0.05%	0.21%	0.68%
9	0.00%	0.00%	0.22%	0.01%	0.01%	0.35%
15	0.00%	0.00%	0.10%	0.00%	0.00%	0.22%