Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

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 for pricing under framing effects. For profit maximization problem, at optimality products are sorted in descending order of value gap, which is defined as expected utility when the product is priced at cost; and markups are page-dependent, with higher markups associated with products on later pages, so that products in the first page are of the highest utility and have the lowest 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 on 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 a 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 he 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, 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 price framing problem is to determine both the framing and pricing of the products in order to maximize the expected revenues.

For the product framing problem, we capture additional effects that come into play after the consideration sets have been formed. We call these effects location preference. Location preference works as follows. First, given that a collection of pages enters into a consumer’s consideration set, products that are displayed higher on a page are more likely to be chosen than those displayed
lower on the same page due to the evaluation effect, all other factors being equal. Second, products that are listed in earlier pages are more likely to be chosen than products that are listed in later pages due to the attraction effect, all other factors being equal. With location preference, we require that the choice model be the MNL model. We capture the effect of location on choice by using location-dependent preference weights, which we will describe in greater detail in Section 8.

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 is more general than those previously proposed in two important ways: It allows for a general choice model and a more general framing structure.
- We prove that the product framing problem is NP-hard, even when there are just 2 pages and the choice model is the MNL model.
- We propose fast, easy-to-implement algorithms with worst-case performance guarantees. Our algorithms are significantly simpler than existing algorithms and offer strong performance bounds. The ease and simplicity of the algorithms mean that they can be personalized on-line for each arriving consumer. The algorithms also apply to an extended model when the product’s display location affects both attention (i.e., the formation of consideration sets) and valuation (i.e., the comparison of utilities).
- We prove new structural results for pricing under framing effects. We show that given a fixed placement of products, at optimality, each page is filled with products until all products are displayed. All products on the same page have the same page-level markup, which increases monotonically with respect to page indices. Products with higher value gaps are given higher priority, where the value gap is defined as expected utility when priced at cost. This implies that the optimal markup is higher for less attractive products. This last finding is contrary to the findings of Arbatskaya (2007), who argues that sellers lower on a list will charge lower prices; thus, consumers with lower search costs will search longer and obtain better deals.
We remark that our model, although it is tailored to e-commerce and virtual pages, can also be used to model brick-and-mortar retailers, with a suitable interpretation of what consumers are willing to look at. Some consumers, for example, would only look at the most prominent displays (see Chandon, Hutchinson, Bradlow and Young 2009, Corstjens and Corstjens 2012), while others may enter a store and look at some aisles or the rest of the store. Our location-preference model also applies in brick-and-mortar settings to signify the value of having products at eye level versus waist level, versus shoe level; and the value of end-of-aisle locations. Our pricing results also have implications for brick-and-mortar retailers. The most prominent displays should have the highest utility products at the lowest markups.

1.3. Relation to assortment planning

Our paper falls within the literature on assortment planning, which is currently a very active area of research. Assortment planning began with a stylized model introduced by van Ryzin and Mahajan (1999). Van Ryzin and Mahajan (1999) 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, Talluri and van Ryzin (2004) prove that an optimal assortment includes a certain number of products with the highest revenues.

The assortment-planning 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-planning problem can be solved by a linear program when the nest-dissimilarity parameters of the choice model are less than one, and a consumer always makes a purchase within the selected nest. Relaxing either of these assumptions renders the problem NP-hard.

The assortment-planning 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-planning 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 planning with location effects, but ignoring consideration sets. Davis, Gallego and Topaloglu (2013) model location effects by introducing location-dependent item weights to the MNL model. The resulting assortment-optimization problem reduces to a linear program with totally unimodular constraints.

Assortment planning 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 sequence is predetermined by all the products’ expected utilities, which are common knowledge.

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 much 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, and all products must be displayed even if doing so is suboptimal.
1.4. Relation to assortment pricing

Our work also falls within the area of 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 to model framing effects. Under the MNL revenue-(profit-) maximization model, we find that the constant price (markup) property still holds at the page level. We also show that the price is higher for less attractive products, which is contrary to the findings of Arbatskaya (2007).

2. Product-Framing Problem

Consider $n$ products. Product $i$ has revenue $r_i, i \in \mathcal{N} = \{1, \ldots, n\}$. The revenue can be the profit contribution net of costs in some applications. Products are organized into 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. Consumers who arrive at the system have consideration sets that are governed by a random variable $X$ taking values in a set $\mathcal{M}$ that consists of all the positive integers, or is of the form $\{1, \ldots, m\}$ for some finite positive integer $m$. Let $\lambda(x) = \mathbb{P}[X = x]$, and $\Lambda(x) = \mathbb{P}[X \geq x]$ for all $x \in \mathcal{M}$. A consumer who draws $X = x \in \mathcal{M}$ has consideration set $\{1, \ldots, x\}$. From this consideration set, the consumer purchases at most one product according to a general choice model.
The product-framing problem is to distribute the products among the pages to maximize the expected revenue that can be obtained from an arriving consumer. We assume that we do not know the number of pages that a consumer is willing to view when he 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 $x \in \mathcal{M}$ pages. By the law of large numbers, these frequencies converge to the probability distribution of $X$. We call this multi-page assortment-optimization problem, the product-framing problem. 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 will assume that consumers choose according to a general choice model that is independent of the consumer type $x \in \mathcal{M}$. Later, we will relax this assumption and show that under mild conditions we can still guarantee a constant ratio of the expected revenue relative to the upper bound.

The product-framing problem can be formulated in terms of decision variables $y_{ix} \in \{0,1\}, i \in \mathcal{N}, x \in \mathcal{M}$, where $y_{ix} = 1$ if item $i$ is displayed on page $x$ and is zero otherwise. Let $P(i,S)$ denote the purchase probability of item $i$ when the consideration set is $S \subseteq \mathcal{N}$, with $P(i,S) = 0$ if $i \notin S$. The formulation in terms of the variables $y_{ix}$ is given by

$$\text{OPT} = \max_{y_{ix}} \sum_{x \in \mathcal{M}} \lambda(x) \sum_{i \in \mathcal{N}} r_i P(i, \{k \in \mathcal{N} : \sum_{l=1}^{x} y_{kl} = 1\})$$

s.t. $\sum_{x \in \mathcal{M}} y_{ix} \leq 1, \forall i \in \mathcal{N}$

$$\sum_{i \in \mathcal{N}} y_{ix} \leq p, \forall x \in \mathcal{M}$$

$$y_{ix} \in \{0,1\}, \forall i \in \mathcal{N}, x \in \mathcal{M}.$$  (1)

We will show that problem (1) is NP-hard. Therefore, to derive performance bounds for our algorithms, we will first find an upper bound on (1), which can be easily computed.
3. Upper Bound on Optimal Revenue for Product Framing

Consider the following assortment-optimization problem, which constrains the number of products in an assortment to be at most \( c \).

\[
G(c) = \max_{S \subseteq N} \sum_{i \in S} r_i P(i, S) \tag{2}
\]

s.t. \(|S| \leq c\).

Let \( R(x) = G(x \cdot p) \) be the optimal expected revenue from consumers who see \( x \in \mathcal{M} \) pages. Let \( S(x) \subset \mathcal{N} \) be an optimal solution associated with the revenue \( R(x) \). If we had the luxury of knowing the number of pages \( x \in \mathcal{M} \) upon the arrival of a consumer, we would offer him assortment \( S(x) \), and would earn expected revenue

\[
E[R(X)] = \sum_{x \in \mathcal{M}} \lambda(x) R(x). \tag{3}
\]

The following result shows that this \( E[R(X)] \) is an upper bound on the expected value \( V^{OPT} \) of \( OPT \).

**Theorem 1.** \( E[R(X)] \geq V^{OPT} \).

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

\[
R(x) \geq \sum_{i \in \mathcal{N}} r_i P(i, \{k \in \mathcal{N} : \sum_{l=1}^{x} y^*_{kl} = 1\}) \Rightarrow \sum_{x \in \mathcal{M}} \lambda(x) R(x) \geq V^{OPT}.
\]

\[\square\]

4. Hardness of 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 \( w_1, w_2, ..., w_n \), determine whether there is a set \( S \subseteq \{1, 2, ..., n\} \) 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 2.** Problem (1) is NP-hard even when all consumers follow the same MNL model.

### 5. Assumptions for Analysis of Algorithms

Given that the product-framing problem is NP-hard, one of our goals will be to propose algorithms with guaranteed performance ratios relative to $\text{OPT}$. Towards this goal, we will make three innocuous assumptions:

**Assumption A1**

$$P(i, S) \geq P(i, T) \text{ for all } i \in S, \text{ and } S \subseteq T \subseteq N.$$  

**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 $\epsilon \in (0, 1]$.

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

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) $R(x)/x$ is decreasing in $x \in M$.

Assumption A2 states that we can approximately solve the capacity-constrained problem (2) within a constant approximation ratio. For the MNL and Nested Logit models, the capacity-constrained problem can be solved exactly in polynomial time ($\epsilon = 0$) (Gallego and Topaloglu 2014). The Mixed Multinomial Logit model with a constant number of mixtures can also be solved
within any given 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 will just assume that $\epsilon = 0$, but all of our algorithms and bounds can be easily extended to the case of $\epsilon > 0$ by replacing the revenue function $R(x)$ with an approximate value, and scaling the corresponding bound by $(1 - \epsilon)$.

Let $q(x) = E(X - x|X > x)$ for all $x \in M$. Assumption A3 is equivalent to $q(x) \leq q(0)$ for all $x \in M$. Assumption A3 implies that the additional number of pages that a consumer will see is no more than the expected number of pages he or she would like to see before the search.

6. Algorithms with constant Performance Bounds for Product Framing

We now propose algorithms for problem (1), which have guaranteed constant performance ratios relative to $OPT$ under Assumptions A1, A2, and A3. All of our performance guarantees in this section are 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.

Our first algorithms, called NEST, start by truncating the number of pages to an arbitrary integer $y$, and by selecting an optimal assortment, say $S(y)$, for the first $y$ pages. Then they select an optimal 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 pages have been determined. With a little bit abuse of notations, we let $\tilde{R}(S)$ denote the expected revenue when a consumer considers all of the products in the set $S$.

NEST($y$) Algorithms:

- Let $y$ be an integer and let $S(y)$ be the set of products with revenue $R(y)$. Without loss of generality, we assume that $|S(y)| > (y - 1)p$. In other words, a minimum of $y$ pages are required to hold all products in $S(y)$. If this fails, then we can select a smaller value of $y$ until this assumption holds.

- For $x = y - 1$ down to 1, let $S(x)$ be the set of products to be displayed on the first $x$ pages, and let $\tilde{R}(x) = R(S(x))$. Using Lemma 1, we can set $S(x)$ to be a subset of $S(x + 1)$ such that $|S(x)| = x \cdot p$ and

\[
\frac{\tilde{R}(x)}{x} \geq \frac{\tilde{R}(x + 1)}{x + 1}.
\]
This implies that $\tilde{R}(x)/x \geq \tilde{R}(y)/y = R(y)/y$ for all $x < y$.

- Use any heuristic to fill products into pages $x > y$, such that the total expected revenue of products in the first $x > y$ pages is at least $R(y)$. As a default, we can leave pages $x > y$ blank.

### 6.1. Constant lower bound

Notice that the revenue of $NEST(x)$ for pages $x < y$ is guaranteed to be at least $\tilde{R}(x) \geq xR(y)/y$, and the expected revenue for $x \geq y$ is at least $yR(y)/y$. Consequently,

$$V^{NEST}_{N} \geq \frac{R(y)}{y} \sum_{x=1}^{y-1} xN \lambda(x) + R(y)\Lambda(y)$$

$$\geq \frac{R(y)}{y} \sum_{x=1}^{y-1} xN \lambda(x) + R(y)\Lambda(y) \quad \text{(by (4))}$$

$$= \frac{R(y)}{y} \sum_{x=1}^{y-1} xN \lambda(x) + \frac{R(y)}{y} y\Lambda(y)$$

$$= \frac{R(y)}{y} E[\min(X,y)].$$

Let $NEST = NEST(y)$ be an algorithm in the above class where $y$ is selected to maximize the lower bound. More precisely, let $y$ be the largest integer in the set

$$\arg\max_{x \in \mathcal{M}} E[\min(X,x)]R(x)/x.$$

Then $V^{NEST} \geq \frac{R(y)}{y} E[\min(X,y)]$ for this choice of $y$. We will show that $NEST$ achieves at least $6/\pi^2 = 0.607921\ldots$ of the optimal expected revenue.

The idea of the proof is to minimize $\frac{R(y)}{y} E[\min(X,y)]$ over all functions $R(\cdot)$ satisfying Lemma 2, and over all distributions $X$ satisfying Assumption A3. For convenience we will scale $R(\cdot)$ without loss of generality so $E[R(X)] = 1$. This leads to a min max problem that can be formulated as follows:

$$\gamma = \min_{R, \Lambda} \max_{x \in \mathcal{M}} \frac{R(x)}{x} E[\min(X,x)],$$

subject to

$$1 = \Lambda(1) \geq \Lambda(2) \geq \cdots \geq 0.$$
The first constraint ensures that $\Lambda$ corresponds to a valid tail distribution. The second constraint ensures that $X$ has the NBUE property. The third and fourth constraint ensure that $R$ is increasing and satisfies Lemma 2. The fifth constraint normalizes $E[R(X)]$ to 1, and the last ensures that $R$ is non-negative.

To uncover the structure of an optimal solution to (5) we will first characterize the functions $R(\cdot)$ and $\Lambda(\cdot)$ in the worst case. In the process of establishing the bounds, we will not use special notation, say $\gamma^*, R^*(\cdot)$ or $\Lambda^*(\cdot)$ to denote the optimal solution to program (5). This comes at a small cost of ambiguity, but makes the exposition a bit cleaner.

Let $y$ be the largest integer in the set $\arg \max_{x \in M} \frac{R(x)E[\min(X,x)]}{x}$ and $\gamma = \frac{R(y)E[\min(X,y)]}{y}$. We next show the worst-case structure for $R(x)$ for all $x \in M$.

**Lemma 3.** $R(x) = \gamma \frac{x}{E[\min(X,x)]}$ for all $x \in M$.

The lemma implies that in the worst case, the maximum value of $E[\min(X,x)]R(x)/x$ is achieved by all points $x$ in the set $M$. In other words, the worst-case function $R$ and the worst-case distribution for $X$ are such that the value $E[\min(X,x)]R(x)/x$ is constant for all $x \in M$.

**Proof.** We first show that the function $\frac{R(x)}{x}E[\min(X,x)]$ cannot decrease in $x$. Suppose for a contradiction that there is a smallest $x$, $x > 1$, such that the function decreases strictly from $x - 1$ to $x$. That is, $\frac{R(x-1)}{x-1}E[\min(X,x-1)] > \frac{R(x)}{x}E[\min(X,x)]$. Then we can revise the function as follows. For points $y \in \{x, x + 1, x + 2, ..., m\}$ such that $R(y) = R(x)$, we increase their value $R(y)$, such that $R(y) \leftarrow (1 + \epsilon)R(x)$. Let us denote this set of $y$'s by $Y$. Note it might not be a singleton, since there might be consecutive $y$'s such that $R(y) = R(x)$. In this meantime, we scale
down the $R(\cdot)$ at all other points $y \in \mathcal{M} - \mathcal{Y}$, i.e., $R(y) \leftarrow (1 - \epsilon')R(y)$. We can properly chose the values of $\epsilon$ and $\epsilon'$ such that $E[R(X)] = 1$ is maintained. Let $z$ be the largest index in the set $\mathcal{Y}$, then the only two constraints that might get violated by this revision are $\frac{R(x-1)}{x} \geq R(x)$ and $R(z) \leq R(z+1)$. However, $\frac{R(x-1)}{x} E[\min(X, x-1)] > \frac{R(x)}{x} E[\min(X, x)]$ implies $\frac{R(x-1)}{x} > \frac{R(x)}{x}$, which means the decreasing unit reward condition will not be violated as long as $\epsilon$ and $\epsilon'$ are chosen small enough; moreover, $R(z) \leq R(z+1)$ will not be violated either since by the definition of set $\mathcal{Y}$ we know $R(z) < R(z+1)$ before the revision. So again, the increasing reward condition will not be violated as long as $\epsilon$ and $\epsilon'$ are chosen small enough. Thus the optimality condition meets contradiction, as we can strictly decrease the value of $\gamma$.

We now show that the function $\frac{R(x)}{x} E[\min(X, x)]$ cannot increase in $x$ either. Suppose there is a smallest $x$, $x > 1$ such that the function increases strictly from $x-1$ to $x$, or equivalently $\frac{R(x-1)}{x-1} E[\min(X, x-1)] < \frac{R(x)}{x} E[\min(X, x)]$. Then we can scale up the value of $R(\cdot)$ at all points $y < x$, and scale down $R(\cdot)$ at all points $y \geq x$, while maintaining the constraint $E[R(X)] = 1$. This adjustment is valid because the only violation might occur at pair $(x-1, x)$ for the constraint $R(x-1) \leq R(x)$. However, $\frac{R(x-1)}{x-1} E[\min(X, x-1)] < \frac{R(x)}{x} E[\min(X, x)]$ implies $R(x-1) < R(x)$, as $\frac{E[\min(X, x-1)]}{x-1} > \frac{E[\min(X, x)]}{x}$. Consequently, the condition that $R(\cdot)$ be increasing will not be violated as long as we properly choose the scaling parameter. This leads to a contradiction as we can strictly decrease the value of $\gamma$. □

For $R(x)$ of the form given by Lemma 3, we next show how $\gamma$ depends on the distribution of $X$, by using the fact that $E[R(X)] = 1$.

**Theorem 3.** Let $Y$ be a random variable, independent of $X$, with the same distribution as $X$. Then

$$1 = \frac{1}{\gamma} = \max_{X,Y} E \left[ \frac{X}{E[\min(X,Y)|X]} \right].$$

**Proof.** By Lemma 3, $R(x) = \gamma \frac{x}{E[\min(Y, x)]}$, where $Y$ is a random variable with the same distribution as the worst case $X$. Then

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

The result follows after dividing by $\gamma$. □
Our next goal is to show that \( E\left[ \frac{X}{E[\min(X,Y)|X]} \right] \leq \frac{\pi^2}{6} \) among all non-negative, NBUE distributions. In our analysis, we will allow for continuous distributions to obtain the bound. We will later show that there is a discrete distribution that in the limit achieves the bound.

**Lemma 4.** If \( X \) is non-negative with mean \( \mu \) and NBUE, and \( Z \) is exponential with the same mean, then \( X \leq_{icx} Z \) in the increasing convex ordering.

**Proof.** Let \( \bar{F}_X \) denote the CCDF of \( X \) and \( \bar{F}_Z \) denote the CCDF of \( Z \). Proving \( X \geq_{icx} Z \) 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 \( a \geq 0 \). Since \( X \) has NBUE property, we have \( E(X-t|X>t) \equiv \int_t^\infty \bar{F}_X(v)dv \leq \mu \) for all \( t \geq 0 \), indicating \( \bar{F}_X(t) \int_t^\infty \bar{F}_X(v)dv \geq \frac{1}{\mu} \).

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

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

Since \( \ln \int_0^\infty \bar{F}_X(v)dv = \ln E(X) = \ln \mu \), 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 1.** If \( Y \) is a non-negative, NBUE random variable with mean \( \mu \), and \( Z \) is exponential with the same mean, then \( E[\min(Y,x)] \geq E[\min(Z,x)] \) for all \( x \geq 0 \).

**Proof.** Applying the Lemma 4 to \( Y \) and \( Z \), we see that \( E[(Y-x)^+] \leq E[(Z-x)^+] \), so \( E[\min(Y,x)] = E[Y] - E[(Y-x)^+] \geq E[Z] - E[(Z-x)^+] = E[\min(Z,x)] \).

**Corollary 2.** For any independently and identically distributed \( X \) and \( Y \) that are NBUE and have mean \( \mu \), \( E[X/E[\min(X,Y)|X]] \leq E[W/E[\min(Z,W)|W]] \), where \( Z \) and \( W \) are independent exponentials with mean \( \mu \).
Proof. From Corollary 1, \( \frac{x}{E[\min(Y,x)]} \leq \frac{x}{E[\min(Z,x)]} = G(x) \). Since \( G(x) \) is increasing convex, and \( X \leq w \), it follows from the Lemma 4, that

\[
E[G(X)] \leq E[G(W)].
\]

□

We are now ready to present the bound of our algorithm.

**Theorem 4.** The expected revenue \( V^{NEST} \) is at least \( \frac{6}{\pi^2} V^{OPT} \).

**Proof.** By Lemma 3, \( R(x) = \gamma \frac{x}{E[\min(X,x)]} \), and by Theorem 3, and Corollary 2, \( 1/\gamma = E \left[ \frac{W}{E[\min(Z,W)\mid W]} \right] \), where \( W \) and \( Z \) are independent exponentially distributed random variables with the same mean, say \( \mu \). We will now show that \( E \left[ \frac{Z}{E[\min(Z,W)\mid W]} \right] = \frac{\pi^2}{6} \), which is independent of \( \mu \). This will imply that

\[
\frac{V^{NEST}}{V^{OPT}} \geq \gamma = \frac{6}{\pi^2},
\]

completing the proof.

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

\[
E \left[ \frac{Z}{\mu (1 - \exp(-Z/\mu))} \right] = \int_0^\infty \frac{z/\mu \exp(-z/\mu)}{1 - \exp(-z/\mu)} \frac{dz}{\mu} = \int_0^\infty \frac{ue^{-u}}{1 - e^{-u}} du = \sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6},
\]

where the first equality follows from the distribution of \( Z \), the second from the transformation \( u = z/\mu \), 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. □

Our results show that the bound for Problem (5) is tight for the exponential distribution. The reader may wonder whether there is a discrete distribution over the non-negative integers such that the bound for Problem (5) is also tight. The following corollary asserts that this is indeed the case.

**Corollary 3.** The performance ratio \( \frac{6}{\pi^2} \) with respect to the upper bound (3) is attained when \( X \) and \( Y \) have geometric distributions with mean \( 1/(1-p) \) as \( p \uparrow 1 \).
Proof. Let $X, Y$ be geometrically distributed with mean $\frac{1}{1-p}$. That is, $P[X = x] = P[Y = x] = p^{x-1}(1-p)$. Then we can write

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

For $p < 1$, we have

$$\sum_{y=1}^{\infty} \frac{p^{y-1}(1-p)^2 y}{1-p^y} = (1-p)^2 \sum_{y=1}^{\infty} \frac{y}{p} \frac{p^y}{1-p^y}$$

$$= (1-p)^2 \sum_{y=1}^{\infty} \frac{y}{p} \sum_{n=1}^{\infty} p^{yn}$$

$$= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \sum_{y=1}^{\infty} y(p^n)^{y-1}$$

$$= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \frac{d}{dp^n} \frac{1}{1-p^n}$$

$$= \frac{(1-p)^2}{p} \sum_{n=1}^{\infty} p^n \frac{1}{(1-p^n)^2}$$

$$= \sum_{n=1}^{\infty} p^{n-1} \frac{(1-p)^2}{(1-p^n)^2}.$$ 

The above series is an increasing function of $p$. It is maximized both locally and globally at $p = 1$. To find the limit at $p = 1$, we use two applications of L'Hospital's rule to obtain

$$\lim_{p \to 1} \sum_{n=1}^{\infty} p^{n-1} \frac{(1-p)^2}{(1-p^n)^2} = \lim_{n \to \infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \quad \Box$$

### 6.2. Improvement to the algorithms

We can refine the NEST Algorithms by extending the algorithms to pages beyond $y$. In particular, let $\tilde{R}(x)$ be the performance of the NEST Algorithms described above for all $x \leq y$. Set $x = y + 1$ and solve the $x$-page problem subject to the constraint that the first $y$ pages are composed of $S(y)$. Then select the products for page $x = y + 1$ by solving the problem for page $x$ among all products in the complement of $S(y)$. For higher values of $x$, continue the procedure, fixing the pages for all $z < x$ and selecting page $x$ among the complement of the products used in pages 1, $\ldots$, $x - 1$. Let $NEST^+(y)$ be an algorithm in this improved class. When $y$ is chosen as the largest integer in the set

$$\arg \max_{x \in M} V^{NEST^+(x)}$$
then we call \( NEST^+ (y) \) as \( NEST^+ \). Note that \( NEST^+ \) continues to have a worst-case performance bound of \( \frac{6}{\pi} \) as we have proved in Section 6.1.

**Corollary 4.** *If the capacitated assortment problem \( G(c) \) has nested solutions, i.e., \( S(1) \subseteq S(2) \subseteq \ldots \subseteq S(|N|) \), where \( S(c) \) is the optimal assortment for \( G(c) \), then \( NEST^+ (y) \) returns an optimal solution.*

### 7. 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 customers. This relaxation allows us to model *keen customers* who are more determined to buy, and *picky customers* who are more price- or quality-sensitive.

To this end, we allow the choice model to be type dependent. Accordingly, for a consumer of type \( x \), who intends to view \( x \) pages, we let \( P_x(i, S) \) be the purchase probability of item \( i \) 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

\[
    r(x, S) = \sum_{i \in S} r_i P_x(i, S)
\]

(7)

to be the expected revenue from presenting assortment \( S \) to a consumer of type \( x \). We will still use \( R(x) \) to denote the optimal revenue from the capacitated assortment problem for consumers with type \( x \), namely,

\[
    R(x) = \max_{S \subseteq N} r(x, S)
\]

s.t. \( |S| \leq x \cdot p \).

It is easy to verify that in this extended model Theorem 1 is still valid, i.e., \( E[R(X)] \) 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 1.

We make the following assumptions:

**Assumption B1** \( r(x, S) \leq r(y, S) \) for all \( x \leq y \) and \( S \subseteq N \) with \( |S| \leq x \cdot p \)
Assumption B2 $R(x)/x$ is decreasing in $x$.

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

Assumption B4 is stronger than Assumption A3 as it is sufficient for Assumption A3 (Shaked and Shanthikuma 2007). It is equivalent to $h(x) \equiv \frac{\lambda(x)}{\Lambda(x)}$ increasing in $x \in \mathcal{M}$. Assumption B4 implies that the probability that a consumer will view the next page is decreasing in the number of pages he or 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 10.2.2. 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.

7.1. TRUNCATE (TRUNC) Algorithms

This algorithm simply optimizes for the set of items to be included in the first $x$ pages, for a well-chosen $x$. In other words, it truncates the number of pages to exactly $x$. It does not try to optimize for the placement of these items within the $x$ pages. The idea is to cater only to consumers who will view at least $x$ pages.

Algorithm:

1. Choose $y = \max \arg \max_{x \in \mathcal{M}} R(x)P[X \geq x]$. Let $S(y)$ be the set of items with expected revenue $R(y)$.

2. Use any heuristic to fill in the first $y$ pages such that the set of all items in the first $y$ pages is $S(y)$. 
3. Use any heuristic to fill in pages $x > y$ such that the total expected revenue of items in the first $x > y$ pages is at least $R(y)$. As a default, we can leave pages $x > y$ blank. This directly follows Assumption B1.

Let TRUNC be an algorithm in the above class. We next show that TRUNC achieves at least $1/3$ of the optimal expected revenue. The main idea of the proof is as follows. Clearly $R(y)\Lambda(y)$ is a lower bound on the expected revenue $V^{TRUNC}$ of TRUNC. We will minimize the quantity $R(y)\Lambda(y)$ over all IFR distributions of $X$ and all increasing functions $R(\cdot)$ satisfying Assumption (B2). We will scale $R(\cdot)$ without loss of generality so the upper bound $E[R(X)]$ is normalized to 1. We will then show that the smallest value of $R(y)\Lambda(y)$ is at least $1/3$. Doing this is equivalent to proving the same lower bound on the following optimization problem:

$$r = \min_{R, \Lambda} \max_{x \in M} R(x)\Lambda(x) \quad (8)$$

subject to:

$$1 = \Lambda(1) \geq \Lambda(2) \geq \cdots \geq 0$$

$$\Lambda(x + 1)\Lambda(x - 1) \leq \Lambda(x)^2, \quad x \geq 2$$

$$R(1) \leq R(2) \leq \cdots$$

$$R(x) \geq \frac{x}{x+1}R(x+1), \quad x \geq 1,$$

$$\sum_{x \in M} (\Lambda(x) - \Lambda(x+1))R(x) = 1$$

$$R(x) \geq 0 \quad x \in M.$$  

All the constraints follow the same logic as those in optimization problem (5). However, note that the third constraint that $R$ be increasing is just a necessary condition of Assumption B1.

We are now ready to present our main result for the TRUNC algorithm. Please refer to the Appendix for the detailed analysis.

**Theorem 5.** Let $R$ and $\Lambda$ be an optimal solution of (8). Then in the worst case $X$ is geometric with mean $E[X] = y$, and

$$r = \frac{1}{3 - 2/y} \geq \frac{1}{3}. \quad (9)$$
Notice the this bound is sharper than $1/3$ when $y$ is small. We have $r = 1$ if $y = E[X] = 1$ as expected because then people only see page one and TRUNC is optimal. The result is $r \geq 1/2$ when $y = E[X] = 2$, and decreases slowly to $1/3$ as $y$ increases to infinity. Thus, the $1/3$ bound is attained in the limit in the unrealistic case that people see in expectation an infinite number of pages, and there are an infinite number of products.

As a corollary, we have the main result for this section.

**Corollary 5.** Under Assumptions $B1$, $B2$, $B3$, and $B4$, $r \geq 1/3$, and consequently $V_{TRUNC} \geq 1/3V_{OPT}$.

### 8. Product Framing with Location Preferences

In online retail, 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 that is displayed earlier, even if the consumer has determined his consideration set. We model this phenomenon by introducing location-dependent preference weights for all products. We use $\nu_{ixq}$ to denote the preference weight of product $i$ when this product is displayed at location $q$ on page $x$. Without loss of generality, we assume that there are as many locations as there are products so that we can offer all products at once. If the number of possible locations is smaller than the number of products, then we can define additional locations with $\nu_{ixq} = 0$ for all $i \in \mathcal{N}$ for each additional location $(x,q)$. In this case, using one of these additional locations for a product is equivalent to not displaying the product at all. To capture the product-framing decisions, we use $y = \{y_{ixq} : i \in \mathcal{N}, x \in \mathcal{M}, q \in \mathcal{P}\} \in \{0,1\}^{n \times m \times p}$, where $y_{ixq} = 1$ if we offer product $i$ in location $q$ of page $x$; otherwise $y_{ixq} = 0$. As each page can hold $p$ products, the set of available locations of each page $\mathcal{P} \equiv \{1,2,...,p\}$. If the product offer decisions are given by $y$, then we obtain an expected revenue of $R_x(y) = \frac{\sum_{i \in \mathcal{N}} \sum_{l=1}^{\mathcal{L}_x} \sum_{q \in \mathcal{P}} y_{ilq} \nu_{ilq} r_{ix} + \sum_{i \in \mathcal{N}} \sum_{l=1}^{\mathcal{L}_x} \sum_{q \in \mathcal{P}} y_{ilq} \nu_{ilq} r_{ix}}{1 + \sum_{i \in \mathcal{N}} \sum_{l=1}^{\mathcal{L}_x} \sum_{q \in \mathcal{P}} y_{ilq} \nu_{ilq}}$ from consumers who only view the first $x$ pages.
Recall that we can obtain the upper bound of the problem without location preference by 
\[ E(R(X)) \equiv \sum_{x \in M} \lambda(x) R(x), \]
where \( R(x) \equiv G(x \cdot p) \) is the highest expected revenue from consumer who is willing to view \( x \) pages. Under the MNL choice model, for the problem with location preference, the problem \( G(x \cdot p) \) is still polynomially solvable. This is because the problem is formulated as

\[
\max_y \quad \frac{\sum_{i \in N} \sum_{l=1}^{x} \sum_{q \in P} y_{ilq} \nu_{ilq} r_i}{1 + \sum_{i \in N} \sum_{l=1}^{x} \sum_{q \in P} y_{ilq} \nu_{ilq}} \\
\text{s.t.} \quad \sum_{i \in X_l} y_{ilq} \leq 1 \quad \forall l = 1, ..., x, \quad q \in P; \\
\sum_{l=1}^{x} \sum_{q \in P} y_{ilq} \leq 1 \quad \forall i \in N; \\
y_{ilq} \in \{0, 1\} \quad \forall i \in N, \quad l = 1, ..., x, \quad q \in P;
\]
where the first set of constraints ensures 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 5.** Assume that \( \nu_{ixq} \leq \nu_{ix'q} \) if \( (x - 1) \times p + q > (x' - 1) \times p + q' \) for all products \( i \in N \), i.e., the preference weight will decrease if the item is displayed further in the rear. Then \( \frac{R(x)}{x} \leq \frac{R(x')}{x'} \) for any \( x, x' \in M \) and \( x > x' \).

**Proof.** It is easy to see that the value of \( R(x) \) is a root of the following equation:

\[
R(x) = \max_y \sum_{i \in N} \sum_{l=1}^{x} \sum_{q \in P} (r_i - R(x)) \nu_{ilq} y_{ilq}
\]

Any suboptimal assortment yields the left hand side bigger than the right hand side. Under the optimal solution yielding \( R(x) \), we can pick the \( x' \times p \) products that give the highest \( (r_i - R(x)) \nu_{ilq} \). Let us denote this set of products by \( S \). Then it must be true that

\[
\sum_{i \in S} \sum_{l=1}^{x} \sum_{q \in P} (r_i - R(x)) \nu_{ilq} y_{ilq} \geq x' \frac{R(x)}{x}
\]
If any item \( i \in S \) is displayed in some page later than \( x' \), for example at position \( lq \), then there must exist one product \( i' \) at position \( l'q' \) such that \( l' \leq x' \) and \( i' \notin S \). We replace product \( i' \) by product \( i \), then \( \nu_{i'l'q'} \geq \nu_{ilq} \) by assumption. Under the updated configuration \( y' \) it must be

\[
\sum_{i \in S} \sum_{l=1}^{x'} \sum_{q \in \mathcal{P}} (r_i - R(x))\nu_{ilq}y_{ilq} \geq \frac{x'R(x)}{x},
\]

Since we are dealing with more restrictive cardinality constraint, i.e., \( x' < x \), we have \( R(x') \leq R(x) \), therefore

\[
R(x') \geq \sum_{i \in S} \sum_{l=1}^{x'} \sum_{q \in \mathcal{P}} (r_i - R(x'))\nu_{ilq}y_{ilq} \geq \frac{x'R(x)}{x},
\]

which implies that

\[
\frac{R(x')}{x'} \geq \frac{R(x)}{x}.
\]

With Lemma 5 and the increasing-failure-rate property, we can show that the \( \frac{6}{x^2} \) performance bound still holds.

**Theorem 6.** Assume that \( \nu_{ixq} \leq \nu_{ix'q'} \) if \( (x - 1) \times p + q > (x' - 1) \times p + q' \) for all products \( i \in \mathcal{N} \). Then all bounds proved in Sections 6 continue to hold for the model with location preference.

**9. Price Framing Problem**

In practice, the retailer may care not only about how to select and display the products, but also how to price them to maximize expected revenues. In this section we consider the problem faced by a retailer who is jointly framing and pricing all products. The framing policy will still determine consumers’ consideration sets, i.e., each customer will consider products on the first \( x \) pages with probability \( \lambda(x) \), \( x \in \mathcal{M} \). However, prices will influence consumers’ valuation of the products. More specifically, under a given consideration set, we assume that consumers make choices according to the MNL model, with utility following a linear form:

\[
u_{i} = a_i - \beta r_i + \epsilon_i.
\]

This formulation is commonly used in Economics, Marketing and Psychology (Berry, Levinsohn and Pakes 1995; Fader and Hardie 1996; Shugan 1980). Here, \( a_i \) is the price-independent quality of item \( i \), \( \beta > 0 \) is the
price sensitivity parameter, and \( r_i \) is the price of product \( i \). Without loss of generality, the i.i.d. random perturbation terms \( \epsilon_i \)'s follow a standard Gumbel distribution and the expected utility of the outside alternative is zero, i.e., \( E(u_0) = 0 \).

Before rushing into the joint optimization problem, let us first consider the case where the products are framed and the only issue is to find optimal prices. More precisely, we assume that an assignment of products to pages is given, where \( y_{ix} \in \{0,1\} \), and \( y_{ix} = 1 \) if and only if item \( i \) is displayed on page \( x \). Naturally \( \sum_x y_{ix} = 1 \) and \( \sum_i y_{ix} \leq 1 \). Without loss of generality, we first look into revenue maximization problem by assuming cost \( c = 0 \). The decision variables then become the price vector \( r = (r_1, ..., r_n) \in \mathbb{R}_+^n \), where \( r_i \) denotes the price of item \( i \). Later we will see the result can be easily extended to the profit maximization case with cost \( c \neq 0 \).

Let \( S_x = \{i : y_{ix} = 1\} \) be the set of products displayed on page \( x \). Then \( S_x = \bigcup_{l=1}^x S_l \) is the consideration set of a consumer who views \( x \) pages. Assuming that the sets \( S_x, \forall x \in \mathcal{M} \), are fixed, we can express the pricing problem as follows:

\[
R^{\text{pricing}} = \max_r \ R(r|S) \equiv \sum_{x \in \mathcal{M}} \lambda(x) R_x(r|S_x). \tag{10}
\]

Here, \( R_x(r|S_x) = \sum_{i \in S_x} r_i P(i,S_x) \) is the expected profit from consumer \( x \), and \( P(i,S_x) \) is the probability that consumer \( x \) purchases item \( i \):

\[
P(i,S_x) = \begin{cases} 
\frac{\exp(a_i - \beta r_i)}{1 + \sum_{k \in S_x} \exp(a_k - \beta r_k)}, & i \in S_x, \\
0, & \text{otherwise},
\end{cases}
\]

which is naturally derived from the MNL model. We prove the following structural results:

**Theorem 7.** Assume that the assignment of products onto \( m \) pages is given. To maximize expected revenue, all products on the same page have a common price. That is, there are page-dependent parameters \( \theta = (\theta_1, ..., \theta_m) \) such that \( r_i = r_k = \theta_x \) if both products \( i \) and \( k \) are displayed on page \( x \). Moreover, the page-dependent parameters monotonically increase with the page indices, i.e., \( \theta_1 < \theta_2 < ... < \theta_m \).
This result extends the classical optimal assortment-pricing structure under the MNL model. It is known that the optimal prices under the MNL profit-optimization problem have a constant markup (see for example, Anderson, de Palma and Thisse (1992), Hopp and Xu (2005) and Gallego and Stefanescu (2011)). We are considering the revenue-maximization problem, and therefore, we set the costs to zero, treating them as sunk costs. We obtain an analogous result that the prices should be equal for products that are displayed on the same page. We can easily extend the result to the profit maximization problem by treating \( a_i = a_i - \beta c_i \), then we will see the markups should be page-dependent.

Therefore, we can reduce the decision variables from \( n \) prices, one for each item, to \( m \) prices, one for each page. The resulting problem in terms of \( \theta_x, x = 1, \ldots, m \), is given by

\[
R^{\text{pricing}} = \max_{\theta} R(\theta|S) \equiv \sum_{x \in M} \lambda(x) R_x(\theta|S_x) = \sum_{x \in M} \lambda(x) \frac{\sum_{l=1}^{x} \theta_l \sum_{i \in S_l} \exp (a_i - \beta \theta_l)}{1 + \sum_{l=1}^{x} \sum_{i \in S_l} \exp (a_i - \beta \theta_l)}
\]

The first order condition yields the system of equations

\[
\frac{\partial R(\theta|S)}{\partial \theta_x} = \sum_{l=x}^{m} \lambda(l) P(S_x, S_l)(1 - \beta \theta_x + \beta R_l(\theta|S_l)) = 0 \quad \forall x \in M,
\]

which are equivalent to

\[
\theta_x = \frac{1}{\beta} + \frac{\sum_{l=x}^{m} \lambda(l) P(S_x, S_l) R_l(\theta|S_l)}{\sum_{l=x}^{m} \lambda(l) P(S_x, S_l)} = \frac{1}{\beta} + \frac{\sum_{l=x}^{m} \lambda(l) / V(S_l) R_l(\theta|S_l)}{\sum_{l=x}^{m} \lambda(l) / V(S_l)}.
\]

In other words, the (common) price for products listed on page \( t \) is a constant, plus the weighted average of the expected revenue from the consumers who will view, and therefore will consider, these products.

In general, the expected revenue \( R(\theta|S) \) may not be jointly concave in the \( \theta \) vector.

**Example 1.** Two products are to be displayed on two pages. Suppose that \( P(X = 1) = 56\% \) and \( P(X = 2) = 44\% \). Product one has quality \( a_1 = 4 \) and product two has \( a_2 = 2 \). The price sensitivity is equal to one. The graph below shows how the expected revenue changes with respect to the price. As we can see, the expected revenue is not jointly concave in the price vector.
With the above optimal pricing structure, we now look into the problem of how to display the products. A related question is whether all products should be displayed. By the presumed utility structure, we see that products are differentiated by their quality parameters $a_i$, $i = 1, ..., n$. Without loss of generality, we will assume that the products are ordered in decreasing quality, so that $a_1 \geq a_2 \geq ... \geq a_n$.

**Theorem 8.** Assume that products differ in their quality parameters $a_i$, $i = 1, ..., n$ such that $a_1 \geq a_2 \geq ... \geq a_n$. At optimality, each page will be filled with products until all products are displayed. The products are displayed in the order of their indices, so that higher quality products are displayed earlier.

In essence, Theorems 7 and 8 tell us that higher-quality products should be displayed first and given lower prices. This result might appear unintuitive, and therefore requires some explanation. The prices are of the form $r_i = \theta_{x(i)}$, where $x(i)$ is the index of the page for product $i$, with
Thus products seen by fewer people have a higher price, with the lowest price enjoyed by products in page one, which are seen by everybody. Interestingly, the price \( \theta_1 \), levied on the highest quality products that appear in the first page is higher than the optimal price, that would result if \( P(X \leq 1) = 1 \); the price \( \theta_m \), levied on the lowest quality products that appear in the last page is lower than the optimal price, that would result if \( P(X \geq m) = 1 \). In essence, consumers are penalized for not being willing to see all products and are compensated for being willing to see more products. Another way to understand the result is that lower-quality products are charged higher prices and serve to steer consumers to the higher-quality products. This result is in sharp contrast with the result for pricing in an oligopoly market, where several firms are deciding prices, see Arbatskaya (2007).

**Corollary 6.** Under optimal pricing and frame, the expected revenue \( R_x(r|S_x) \) for a consumer who views \( x \) pages is increasing and concave with respect to \( x \).

This result is intuitive as the objective function is increasing in the number of products, and consumers who look at more pages also look at more products.

The optimal structure can also be extended to the profit-maximization problem. Interestingly, products now should be ordered by their value gaps, defined as the quality minus price sensitivity times price, i.e., the expected utility when products are priced at their costs, which is first introduced by Gallego, Li and Beltran (2016).

**Corollary 7.** For the profit-maximization problem, at optimality, products are displayed in decreasing order of their value gaps, i.e., \( \tilde{v}_i = a_i - \beta c_i \), and they are priced with page-level constant markups, i.e, \( r_i - c_i = \theta_{x(i)} \), where the page-level markups should monotonically increase with the page indices. Therefore the products are displayed in decreasing order of their expected net utility as well.

This result can be easily seen by replacing \( a_i \) with the value gap \( a_i - \beta c_i \). For details, see Gallego, Li and Beltran (2016).
10. Computational Experiments

In this section, we numerically test the performance of the following framing heuristics:

- The NEST algorithm introduced in Section 6. Recall that NEST(y) leaves pages x > y empty.
- The enhanced algorithm NEST+ introduced in Section 6.2, which is designed to improve practical performance.
- A heuristic, SORT$_1$, which sorts and displays products in increasing order of price.
- A heuristic, SORT$_2$, which sorts and displays products in decreasing order of attractiveness.
- A greedy heuristic, BOTTOM-UP (BU), which starts with the first page (x = 1) and sequentially fills in products that would maximize the expected revenue for type $x$ customers, $x = 2, \ldots, m$, such that this assortment includes the assortment for type-$x - 1$ customers.
- A greedy heuristic, TOP-DOWN (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 type-$m - k$ customers, from the assortment for type $m - k + 1$ customers.

10.1. Experimental Setup

We proceed by describing the instances being tested in our experiments. In all of our test problems, we assume that consumer choice is governed by the Multinomial Logit Model. The MNL associates the attractiveness $\{v_i : i \in \mathcal{N}\}$ with the products. If the set of products $S(x)$ is displayed in the first $x$ pages, then conditional on consumer type $x$’s arrival, he will buy product $i \in S(x)$ with probability

$$v_i \over 1 + \sum_{i' \in S(x)} v_{i'}.$$ 

By convention, $v_i = e^{a_i - \beta r_i}$, where $a_i$ is the product quality. In the tests, we independently draw every $r_i$ from a uniform distribution over [50, 100]. 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 $i$.

For all test cases of the framing algorithms, we use $n = 300$ and $m = 20$. We test three different distributions of $X$: geometric, uniform and Poisson. We also differentiate the simulation scenarios.
by the expectation of $X$: Small ($E(X) = 2$), Median ($E(X) = 4$), and Large ($E(X) = 8$). We also vary the number $p$ of products in each page.

For each test case, we simulate 1,000 replicates and report the average gaps between the heuristics and the upper bound.

### 10.2. Experimental Result

#### 10.2.1. Results of Framing Algorithms

Refer to Tables 1 to 9. In all scenarios and according to our metrics, $SORT_2$ outperforms $SORT_1$; TD, same as BU dominates $SORT_1$ and $SORT_2$; and unsurprisingly $NEST+$ outperforms $NEST$. $NEST+$ dominates all other heuristics $SORT_1$, $SORT_2$, TD and BU, with the average optimality gap of just 0.99% in the worst case, compared to 35.18% for $SORT_2$, 3.03% for TD and 3.83% for BU. The optimality gaps for $SORT_1$ and $SORT_2$ are relatively uniform across different scenarios, whereas the optimality gaps for $NEST$ and $NEST+$ tend to be larger for the geometric and exponential distributions, which we have shown to be worst-case distributions for these algorithms. When the page capacity and expected number of pages viewed by a consumer increase, $NEST+$, TD and BU tend to give same performance as they become optimal.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>$SORT_1$</th>
<th>$SORT_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.22%</td>
<td>0.76%</td>
<td>25.12%</td>
<td>12.71%</td>
<td>1.06%</td>
<td>3.83%</td>
<td>2.1</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>7.43%</td>
<td>0.26%</td>
<td>31.49%</td>
<td>23.84%</td>
<td>0.43%</td>
<td>0.55%</td>
<td>6.0</td>
<td>60.0</td>
<td>60.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>12.23%</td>
<td>0.01%</td>
<td>37.52%</td>
<td>32.60%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>9.0</td>
<td>83.0</td>
<td>83.0</td>
<td>83.0</td>
</tr>
<tr>
<td>15</td>
<td>8.02%</td>
<td>0.00%</td>
<td>38.73%</td>
<td>34.50%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>15.0</td>
<td>83.2</td>
<td>83.2</td>
<td>83.2</td>
</tr>
</tbody>
</table>
### Table 2  Performance of NEST when $E[X] = 4$ and $X$ follows a geometric distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.46%</td>
<td>0.99%</td>
<td>29.39%</td>
<td>20.12%</td>
<td>2.01%</td>
<td>1.68%</td>
<td>4.5</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>9.63%</td>
<td>0.16%</td>
<td>35.15%</td>
<td>29.20%</td>
<td>0.60%</td>
<td>0.24%</td>
<td>9.0</td>
<td>60.0</td>
<td>60.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>7.34%</td>
<td>0.01%</td>
<td>37.92%</td>
<td>33.77%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>18.0</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
</tr>
<tr>
<td>15</td>
<td>11.07%</td>
<td>0.00%</td>
<td>37.12%</td>
<td>33.60%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>18.0</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
</tr>
</tbody>
</table>

### Table 3  Performance of NEST when $E[X] = 8$ and $X$ follows a geometric distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.49%</td>
<td>0.56%</td>
<td>33.28%</td>
<td>26.17%</td>
<td>2.44%</td>
<td>0.74%</td>
<td>7.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>9.81%</td>
<td>0.08%</td>
<td>37.32%</td>
<td>32.34%</td>
<td>0.56%</td>
<td>0.11%</td>
<td>13.2</td>
<td>60.0</td>
<td>60.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>5.42%</td>
<td>0.00%</td>
<td>36.26%</td>
<td>32.87%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>27.0</td>
<td>83.2</td>
<td>83.2</td>
<td>83.2</td>
</tr>
<tr>
<td>15</td>
<td>5.38%</td>
<td>0.00%</td>
<td>33.20%</td>
<td>30.58%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>30.0</td>
<td>83.0</td>
<td>83.0</td>
<td>83.0</td>
</tr>
</tbody>
</table>

### Table 4  Performance of NEST when $E[X] = 2$ and $X$ follows a uniform distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.79%</td>
<td>0.79%</td>
<td>24.43%</td>
<td>12.13%</td>
<td>0.88%</td>
<td>3.34%</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>6.24%</td>
<td>0.17%</td>
<td>32.03%</td>
<td>24.77%</td>
<td>0.53%</td>
<td>0.40%</td>
<td>6.0</td>
<td>9.0</td>
<td>9.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>2.66%</td>
<td>0.01%</td>
<td>38.20%</td>
<td>33.45%</td>
<td>0.02%</td>
<td>0.02%</td>
<td>18.0</td>
<td>27.0</td>
<td>27.0</td>
<td>83.2</td>
</tr>
<tr>
<td>15</td>
<td>10.01%</td>
<td>0.00%</td>
<td>39.32%</td>
<td>35.09%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>15.0</td>
<td>45.0</td>
<td>45.0</td>
<td>83.3</td>
</tr>
</tbody>
</table>

**10.2.2. Picky customers**  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,
Table 5  Performance of NEST when $E[X] = 4$ and $X$ follows a uniform distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.90%</td>
<td>0.64%</td>
<td>29.42%</td>
<td>20.48%</td>
<td>2.26%</td>
<td>1.17%</td>
<td>5.0</td>
<td>7.0</td>
<td>7.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>4.87%</td>
<td>0.10%</td>
<td>36.25%</td>
<td>30.69%</td>
<td>0.67%</td>
<td>0.15%</td>
<td>12.0</td>
<td>21.0</td>
<td>21.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>8.97%</td>
<td>0.01%</td>
<td>38.91%</td>
<td>34.79%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>18.0</td>
<td>63.0</td>
<td>63.0</td>
<td>82.9</td>
</tr>
<tr>
<td>15</td>
<td>4.85%</td>
<td>0.00%</td>
<td>37.70%</td>
<td>34.24%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>30.0</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
</tr>
</tbody>
</table>

Table 6  Performance of NEST when $E[X] = 8$ and $X$ follows a uniform distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.26%</td>
<td>0.32%</td>
<td>34.08%</td>
<td>27.63%</td>
<td>2.75%</td>
<td>0.42%</td>
<td>8.0</td>
<td>15.0</td>
<td>15.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>6.14%</td>
<td>0.05%</td>
<td>38.36%</td>
<td>33.72%</td>
<td>0.60%</td>
<td>0.06%</td>
<td>18.0</td>
<td>45.0</td>
<td>45.0</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>3.69%</td>
<td>0.00%</td>
<td>36.50%</td>
<td>33.31%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>35.4</td>
<td>83.0</td>
<td>83.0</td>
<td>83.0</td>
</tr>
<tr>
<td>15</td>
<td>2.24%</td>
<td>0.00%</td>
<td>32.63%</td>
<td>30.25%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>45.0</td>
<td>83.0</td>
<td>83.0</td>
<td>83.0</td>
</tr>
</tbody>
</table>

Table 7  Performance of NEST when $E[X] = 2$ and $X$ follows a Poisson distribution.

<table>
<thead>
<tr>
<th>$p$</th>
<th>NEST</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>TD</th>
<th>BU</th>
<th>NEST</th>
<th>NEST+</th>
<th>TD</th>
<th>BU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.11%</td>
<td>0.49%</td>
<td>24.84%</td>
<td>12.24%</td>
<td>0.94%</td>
<td>3.53%</td>
<td>2.0</td>
<td>17.1</td>
<td>17.5</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>6.19%</td>
<td>0.18%</td>
<td>31.91%</td>
<td>24.48%</td>
<td>0.49%</td>
<td>0.45%</td>
<td>6.0</td>
<td>48.4</td>
<td>48.4</td>
<td>60.0</td>
</tr>
<tr>
<td>9</td>
<td>14.09%</td>
<td>0.02%</td>
<td>38.02%</td>
<td>33.21%</td>
<td>0.02%</td>
<td>0.02%</td>
<td>9.0</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
</tr>
<tr>
<td>15</td>
<td>9.51%</td>
<td>0.01%</td>
<td>39.16%</td>
<td>34.96%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>15.0</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
</tr>
</tbody>
</table>

in the linear case, we set $v_0(x) = 1 + (x - 1)\beta$, for page index $x \in \{1, 2, ..., m\}$. In the exponential case,
we set $v_0(x) = e^{(x-1)\beta}$, for page index $x \in \{1, 2, ..., m\}$. When computing $\text{NEST}(y)$, we set $v_0 = v_0(y)$.

We explore the range of values $[0.2, 2]$ for $\beta$ in the linear case and $[0.1, 1]$ in the exponential case.

Since $\text{NEST}+$ inherits the structure of both $\text{NEST}$ and $\text{TRUNC}$, and is optimized for performance, we have not added separate computational experiments for $\text{TRUNC}$.

Tables 10 and 11 display the results in the linear and exponential case, respectively. The results show that our leading algorithm, namely $\text{NEST}+$ continues to dominate all heuristics by substantial amounts in nearly all cases, except for a small difference in one case, where $\beta$ is very small. Thus, 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 customer type. Thus, our model is a good starting point for an investigation of framing decisions involving heterogeneous customers.
Table 10  Performance of NEST and other heuristics. $E[X] = 4, p = 3$, $X$ follows a Poisson distribution; $v_0(i) = 1 + (i - 1)\beta$, where $i$ is the page index.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>BU</th>
<th>TD</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>BU</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.02%</td>
<td>32.40%</td>
<td>25.93%</td>
<td>0.74%</td>
<td>0.03%</td>
<td>0.10%</td>
<td>34.28%</td>
<td>28.85%</td>
<td>1.70%</td>
<td>0.14%</td>
</tr>
<tr>
<td>0.4</td>
<td>0.01%</td>
<td>29.65%</td>
<td>22.13%</td>
<td>0.60%</td>
<td>0.01%</td>
<td>0.04%</td>
<td>32.18%</td>
<td>24.50%</td>
<td>1.57%</td>
<td>0.07%</td>
</tr>
<tr>
<td>0.6</td>
<td>0.00%</td>
<td>27.44%</td>
<td>19.03%</td>
<td>0.57%</td>
<td>0.00%</td>
<td>0.02%</td>
<td>31.46%</td>
<td>21.40%</td>
<td>1.35%</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.8</td>
<td>0.00%</td>
<td>26.05%</td>
<td>16.76%</td>
<td>0.51%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>29.60%</td>
<td>19.32%</td>
<td>1.18%</td>
<td>0.00%</td>
</tr>
<tr>
<td>1</td>
<td>0.00%</td>
<td>24.81%</td>
<td>14.78%</td>
<td>0.45%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>28.70%</td>
<td>17.01%</td>
<td>1.11%</td>
<td>0.00%</td>
</tr>
<tr>
<td>1.2</td>
<td>0.00%</td>
<td>24.05%</td>
<td>13.54%</td>
<td>0.39%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>27.74%</td>
<td>15.84%</td>
<td>0.91%</td>
<td>0.00%</td>
</tr>
<tr>
<td>1.4</td>
<td>0.00%</td>
<td>22.84%</td>
<td>12.11%</td>
<td>0.35%</td>
<td>0.00%</td>
<td>0.02%</td>
<td>27.08%</td>
<td>15.24%</td>
<td>0.93%</td>
<td>0.02%</td>
</tr>
<tr>
<td>1.6</td>
<td>0.00%</td>
<td>22.61%</td>
<td>11.35%</td>
<td>0.29%</td>
<td>0.00%</td>
<td>0.06%</td>
<td>27.84%</td>
<td>13.48%</td>
<td>0.70%</td>
<td>0.06%</td>
</tr>
<tr>
<td>1.8</td>
<td>0.01%</td>
<td>21.63%</td>
<td>10.48%</td>
<td>0.26%</td>
<td>0.01%</td>
<td>0.11%</td>
<td>27.07%</td>
<td>13.35%</td>
<td>0.62%</td>
<td>0.11%</td>
</tr>
<tr>
<td>2</td>
<td>0.02%</td>
<td>21.52%</td>
<td>9.92%</td>
<td>0.26%</td>
<td>0.02%</td>
<td>0.08%</td>
<td>27.85%</td>
<td>14.01%</td>
<td>0.66%</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

11. 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 item, 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.
Table 11  Performance of NEST and other heuristics. $E[X] = 4; p = 3; X$ follows a Poisson distribution; $v_0(i) = e^{(i-1)\beta}$, where $i$ is the page index.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>BU</th>
<th>TD</th>
<th>NEST+</th>
<th>SORT$_1$</th>
<th>SORT$_2$</th>
<th>BU</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.02%</td>
<td>33.77%</td>
<td>27.79%</td>
<td>0.61%</td>
<td>0.03%</td>
<td>0.09%</td>
<td>36.16%</td>
<td>29.81%</td>
<td>1.78%</td>
<td>0.09%</td>
</tr>
<tr>
<td>0.2</td>
<td>0.02%</td>
<td>30.95%</td>
<td>24.06%</td>
<td>0.73%</td>
<td>0.49%</td>
<td>0.10%</td>
<td>34.73%</td>
<td>26.15%</td>
<td>1.73%</td>
<td>1.27%</td>
</tr>
<tr>
<td>0.3</td>
<td>0.01%</td>
<td>28.17%</td>
<td>19.98%</td>
<td>0.76%</td>
<td>2.30%</td>
<td>0.08%</td>
<td>30.48%</td>
<td>21.91%</td>
<td>1.46%</td>
<td>4.55%</td>
</tr>
<tr>
<td>0.4</td>
<td>0.03%</td>
<td>25.99%</td>
<td>16.69%</td>
<td>0.72%</td>
<td>2.37%</td>
<td>0.10%</td>
<td>30.66%</td>
<td>19.13%</td>
<td>1.37%</td>
<td>4.71%</td>
</tr>
<tr>
<td>0.5</td>
<td>0.09%</td>
<td>24.22%</td>
<td>14.08%</td>
<td>0.69%</td>
<td>2.03%</td>
<td>0.23%</td>
<td>29.65%</td>
<td>15.82%</td>
<td>1.49%</td>
<td>4.66%</td>
</tr>
<tr>
<td>0.6</td>
<td>0.17%</td>
<td>23.37%</td>
<td>12.19%</td>
<td>0.59%</td>
<td>1.79%</td>
<td>0.38%</td>
<td>29.12%</td>
<td>13.60%</td>
<td>1.47%</td>
<td>4.48%</td>
</tr>
<tr>
<td>0.7</td>
<td>0.25%</td>
<td>22.53%</td>
<td>10.72%</td>
<td>0.58%</td>
<td>1.74%</td>
<td>0.53%</td>
<td>27.61%</td>
<td>12.74%</td>
<td>1.31%</td>
<td>4.29%</td>
</tr>
<tr>
<td>0.8</td>
<td>0.27%</td>
<td>21.82%</td>
<td>9.81%</td>
<td>0.60%</td>
<td>1.53%</td>
<td>0.61%</td>
<td>27.22%</td>
<td>13.05%</td>
<td>1.35%</td>
<td>3.83%</td>
</tr>
<tr>
<td>0.9</td>
<td>0.31%</td>
<td>21.46%</td>
<td>9.09%</td>
<td>0.65%</td>
<td>1.47%</td>
<td>0.76%</td>
<td>28.61%</td>
<td>11.12%</td>
<td>1.24%</td>
<td>4.29%</td>
</tr>
<tr>
<td>1</td>
<td>0.40%</td>
<td>21.76%</td>
<td>8.53%</td>
<td>0.60%</td>
<td>1.51%</td>
<td>1.07%</td>
<td>28.15%</td>
<td>10.81%</td>
<td>1.36%</td>
<td>3.94%</td>
</tr>
</tbody>
</table>

12. Appendix

12.1. Proof in Section 4

Proof of Theorem 2.

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 prices 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 $\epsilon$ is some small value which we will define shortly, and $M$ is determined by $\epsilon$ via

\[
2d \cdot M + \epsilon \sum_{i=1}^{d} w_i = 2.
\]
• For \( i = d + 1, d + 2, \ldots, 2d \), we set \( r_i = 34 \) and \( \nu_i = M \).

• For \( i = 2d + 1, 2d + 2, \ldots, 3d \), we set \( r_i = 59 \) and \( \nu_i = 2\delta \) where

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

• For \( i = 3d + 1, 3d + 2, \ldots, n \), we set \( r_i = 59 \) and \( \nu_i = \delta \).

Given the special case constructed above, we argue that for an optimal frame, it is critical to decide which of the first \( 2d \) products with price 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 that are 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 the revenues of all products are at most 59, the expected total revenue should 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 price 59 and attractiveness \( \delta \) by, if any, a spare product with price 59 and attractiveness \( 2\delta \). Furthermore, whenever there is a product with price 59 and attractiveness \( \delta \) on the first page and a product with price 59 and attractiveness \( 2\delta \) on the second page, it is better to greedily swap the two products. Thus, products with price 59 and attractiveness \( 2\delta \) should all be put on the first page.

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

\[
\approx \frac{59 \cdot d \cdot 2\delta + 59 \cdot 2p \cdot \delta}{1 + d \cdot 2\delta + 2p \cdot \delta} \approx 39.33
\]
for customers who view two pages. Thus, when we put those products with price 34 back into the solution, the expected revenue for customers who view two pages should 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, the optimal solution must put some of the products with price 34 on the first page if any, all products with price 59 and attractiveness 2δ on the first page, and fill in all other spots using products with price 59 and attractiveness δ.

Let $S \subseteq \{1, 2, ..., 2d\}$ denote the set of products with price 34 that are put in the first page. We set $\lambda(1) = \lambda(2) = 0.5$. The expected revenue under decision $S$ is

$$R(S) = \lambda(1) \frac{34 \sum_{i \in S} \nu_i + 59 \cdot 2\delta \cdot d + 59 \cdot \delta \cdot (p - |S| - d)}{1 + \sum_{i \in S} \nu_i + 2\delta \cdot d + \delta \cdot (p - |S| - d)} + \lambda(2) \frac{34 \sum_{i \in S} \nu_i + 59 \cdot 2\delta \cdot d + 59 \cdot \delta \cdot (2p - |S| - d)}{1 + \sum_{i \in S} \nu_i + 2\delta \cdot d + \delta \cdot (2p - |S| - d)}$$

$$= 0.5 \frac{59(1 - (|S| - d)\delta) + 34 \sum_{i \in S} \nu_i + 0.5 \frac{59(2 - (|S| - d)\delta) + 34 \sum_{i \in S} \nu_i}{3 - (|S| - d)\delta + \sum_{i \in S} \nu_i}}.$$  

(11)

Now consider the following function defined for $l = 1, 2, ..., 2d$ and $\theta \in \mathbb{R}$.

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

$$= 0.5 \frac{59(1 - \frac{l - d}{200d\theta}) + 34 \left(\frac{l}{d} + \theta\right)}{2 - \frac{l - d}{200d\theta} + \frac{l}{d} + \theta} + 0.5 \frac{59(2 - \frac{l - d}{200d\theta}) + 34 \left(\frac{l}{d} + \theta\right)}{3 - \frac{l - d}{200d\theta} + \frac{l}{d} + \theta}.$$  

By this definition, we have $R(S) = f(|S|, \sum_{i \in S} \nu_i - |S|/d)$ and $\lim_{\epsilon \to 0} R(S) = f(|S|, 0)$. Furthermore, we can prove the following properties for $f(\cdot, \cdot)$:

- $f(l, 0)$ is quasi-concave in $l$ for $l \geq 0$. When $d \geq 1$ and $l$ is relaxed to a non-negative continuous variable, the only solution for $\frac{\partial f(l, 0)}{\partial l} = 0$ is

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

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 $l \geq 0$. 


• When \( l \) can only take non-negative integral values, the maximizer for \( f(l, 0) \) is \( l = d \). We can deduce that

\[
f(d, 0) - f(d - 1, 0) = \frac{-49 + 9600d + 5000d^2}{2(1 - 200d + 600d^2)(1 - 200d + 800d^2)} > \frac{1}{200d^2}, \quad \forall d \geq 1,
\]

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

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.

• When \( l = d \) and \( \theta \in [-1, 1] \), the unique maximizer for \( f(d, \theta) \) is \( \theta = 0 \) as shown by the following calculation.

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

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

When \( \theta \in [-1, 1] \),

\[
\frac{\partial f(d, \theta)}{\partial \theta} = 0 \implies \theta = 0.
\]

It is easy to check that \( \theta = 0 \) is a maximizer for \( f(d, \theta) \).

Thus,

\[
f(d, \theta) < f(d, 0), \quad \forall \theta \neq 0, \theta \in [-1, 1]. \quad (12)
\]

• When \( l \neq d, l \in \{0, 1, 2, \ldots, 2d\} \) and \( \theta \in [-0.5, 0.5] \), we can deduce that

\[
f(l, \theta) = 0.5 \frac{59(1 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{2 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 0.5 \frac{59(2 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{3 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta}
\]

\[
\leq 0.5 \frac{59(1 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{2 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 0.5 \cdot 34|\theta| + 0.5 \frac{59(2 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{3 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 0.5 \cdot 34|\theta|
\]

(because \( 2 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta \geq 1 \) and \( 3 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta \geq 1 \))

\[
= 0.5 \frac{59(1 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{2 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 0.5 \frac{59(2 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{3 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 34|\theta|
\]

\[
\leq 0.5 \frac{59(1 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{2 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 0.5 \frac{59(2 - \frac{l - d}{200d^2}) + 34(\frac{l}{d} + \theta)}{3 - \frac{l - d}{200d^2} + \frac{l}{d} + \theta} + 34|\theta|
\]

(since \( \theta \in [-0.5, 0.5] \))
\[
\begin{align*}
\leq & 0.5 \left( \frac{59}{2} - \frac{l - d}{200d^2} + \frac{34l}{d} \right) (1 + 2|\theta|) + 0.5 \left( \frac{59}{3} - \frac{l - d}{200d^2} + \frac{1}{d} \right) (1 + 2|\theta|) + 34|\theta| \\
= & f(l, 0) + (2f(l, 0) + 34)|\theta| \\
\leq & f(l, 0) + (2f(d, 0) + 34)|\theta|.
\end{align*}
\]

- Combining these results, for \( l \neq d, l \in \{0, 1, \ldots, 2d\} \), if

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

we must always have \(|\theta| < 0.5\) and

\[
f(l, \theta) \leq f(l, 0) + (2f(d, 0) + 34)|\theta| \leq f(l, 0) + \frac{1}{200d^2} < f(d, 0),
\]

where the last inequality follows from the bound in the second bullet point.

In our model, we set

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

Then we can bound the total expected revenue \( R(S) \) using \( f(d, 0) \).

\[
R(S) = f(|S|, \sum_{i \in S} \nu_i - |S|/d)
\]

\[
= f(|S|, |S|M - |S|/d + \sum_{i \in S \cap \{1, 2, \ldots, d\}} \epsilon w_i)
\]

\[
= f(|S|, |S| \frac{2 - \epsilon \sum_{i=1}^{d} w_i}{2d} - |S|/d + \epsilon \sum_{i \in S \cap \{1, 2, \ldots, d\}} w_i)
\]

\[
= f(|S|, \epsilon \left[ -\frac{|S| \sum_{i=1}^{d} w_i}{2d} + \sum_{i \in S \cap \{1, 2, \ldots, d\}} w_i \right]).
\]

Since

\[
\left| \epsilon \left[ -\frac{|S| \sum_{i=1}^{d} w_i}{2d} + \sum_{i \in S \cap \{1, 2, \ldots, d\}} w_i \right] \right|
\]

\[
\leq \frac{1}{2} \sum_{i=1}^{d} w_i \cdot \frac{1}{2} \frac{f(d, 0) + 34}{200d^2} \left[ -\frac{|S| \sum_{i=1}^{d} w_i}{2d} + \sum_{i \in S \cap \{1, 2, \ldots, d\}} w_i \right]
\]

\[
\leq \frac{1}{2} \sum_{i=1}^{d} w_i \cdot \frac{1}{2} \frac{f(d, 0) + 34}{200d^2} \left[ \sum_{i=1}^{d} w_i + \sum_{i \in S \cap \{1, 2, \ldots, d\}} w_i \right]
\]

\[
\leq \frac{1}{2} \frac{f(d, 0) + 34}{200d^2},
\]

\[
\epsilon \leq \frac{1}{2} \frac{f(d, 0) + 34}{200d^2}.
\]
we must have, according to (12) and (13),

\[
R(S) = f(|S|, \sum_{i \in S} \nu_i - |S|/d) \begin{cases} 
= f(d, 0), & \text{if } |S| = d, \sum_{i \in S} \nu_i = 1 \\
< f(d, 0), & \text{otherwise.}
\end{cases}
\] (14)

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. 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.

The above equivalence shows that the ranking problem reduces to the 2-partition problem as follows. If we can solve the ranking problem, then we can find the optimal value. If the optimal value 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.

First, suppose the 2-PARTITION problem has a solution \( T \subset \{1, 2, ..., d\} \) 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 (14), \( S \) is an optimal solution to our model which 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 (14), we must have \( |S^*| = d \) and \( \sum_{i \in S^*} \nu_i = 1 \), which gives

\[
\sum_{i \in S} \nu_i = 1
\]
\[|S|M + \sum_{i \in S \cap \{1,2,\ldots,d\}} \epsilon w_i = 1\]
\[d - \frac{2 - \epsilon \sum_{i=1}^{d} w_i}{2d} + \sum_{i \in S \cap \{1,2,\ldots,d\}} \epsilon w_i = 1\]
\[\sum_{i \in S \cap \{1,2,\ldots,d\}} w_i = \frac{1}{2} \sum_{i=1}^{d} w_i.\]

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

\[\square\]

12.2. Proof in Section 7

To uncover the structure an optimal solution to (8) we will first establish some elementary results concerning the function \(g(x) = x \Lambda(x)\) which turns out to play an important role in the analysis.

**Lemma 6.** If \(X\) has an IFR distribution, then \(g(x)\) is unimodal, and the largest maximizer, say \(y\), is the smallest integer \(x\) such that \((x+1)h(x) > 1\). Moreover, if \(X\) is geometric with \(h(x) = \theta\) for all \(x \in \mathcal{M}\), then \(1/(y+1) < \theta \leq 1/y\).

**Proof of Lemma 6**

A little algebra shows that \(g(x) \leq g(x+1)\) if and only if \(h(x) \leq 1/(x+1)\). Since \(h(x)\) is increasing and \(1/(x+1)\) is decreasing, there is a smallest \(x\), say \(y\), such that \(h(y) > 1/(y+1)\). Then, \(h(x) \leq 1/(x+1)\) for all \(x < y\) implies that \(g(1) \leq \ldots \leq g(y)\). On the other hand, \(h(x) > 1/(x+1)\) for all \(x \geq y\), implies that \(g(y) > g(y+1) > \ldots\). In the geometric case, \(h(x) = \theta\) for all \(x\) implies that at \(\theta = h(y-1) \leq 1/y\) and \(\theta = h(y) > 1/(y+1)\), together imply that \(1/(y+1) < \theta \leq 1/y\) as claimed.

\[\square\]

**Lemma 7.** Let \(R\) and \(\Lambda\) be an optimal solution of (8), and let \(y = \max_{x \in \mathcal{M}} \arg \max_{x \in \mathcal{M}} R(x)\Lambda(x)\). Then

1. \(R(x) = x \frac{R(w)}{y}\) for all \(x \geq y\); and
2. \(h(x) = h(y)\) for all \(x \in \mathcal{M}, \ x \geq y\).
The intuition behind the above result is that if the above properties are not satisfied, then we can find another set of values for $R(\cdot)$ and $\Lambda(\cdot)$ satisfying the constraints but generating a lower objective value in program (8). The result also hints at a geometric or truncated geometric distribution for values of $x \geq y$.

Proof of Lemma 7.

Proof. Consider $x > y$. By Assumption B2, $R(x)/x \leq R(y)/y$, so $R(x) \leq xR(y)/y$ for all $x > y$. If any of the $R(x), x > y$ values is not at its upper bound, then we can rescale the values of $R(x), x > y$, and reduce $R(x), x \leq y$, while maintaining the constraint $E[R(X)] = 1$. But this impossible because it would contradict the optimality of Problem (8).

We next verify that $h(x) = h(y)$ for all $x \in \{y, \ldots, m\}$. This implies that $\Lambda(x+1)\Lambda(x-1) = \Lambda(x)^2$ for all $x = y, \ldots, m-1$. If not, there is a largest $z$ 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 has the effect of increasing $E[R(X)]$. To maintain $E[R(X)] = 1$ we would need to scale down all the $R$'s and in the process reduce $R(y)\Lambda(y)$, again contradicting the optimality of Problem (8).

□

We will now show that $y = \max \arg \max_{x \in \mathcal{M}} R(x)\Lambda(x)$ is also the largest maximizer of $g(x)$ in $\mathcal{M}$.

Lemma 8. Let $R$ and $\Lambda$ be an optimal solution of (8), then $y = \max \arg \max_{x \in \mathcal{M}} R(x)\Lambda(x)$ is the largest maximizer of $g(x)$ in $\mathcal{M}$.

Proof of Lemma 8.

Proof. By Lemma 7, $R(x)/x = R(y)/y$ for all $x > y$. Since $y$ is the largest maximizer of $R(x)\Lambda(x)$ it follows that $R(x)\Lambda(x) < R(y)\Lambda(y)$ for all $x > y$. Multiplying both sides by $x/R(x) = y/R(y)$, we obtain $x\Lambda(x) < y\Lambda(y)$ for all $x > y$. For all $x < y$, $R(x)\Lambda(x) \leq R(y)\Lambda(y)$ and $x/R(x) \leq y/R(y)$, multiplying the inequalities, we obtain $x\Lambda(x) \leq y\Lambda(y)$. Therefore, $y$ is the largest maximizer of $g(x)$ over $x \in \mathcal{M}$. □

Let $r = R(y)\Lambda(y)$. We next show the structure of $R(x)$ for all $x \in \{1, \ldots, y\}$.

Lemma 9. $R(x) = r/\Lambda(x)$ for all $x \in \{1, \ldots, y\}$.
Proof of Lemma 9.

**Proof.** From \( r = R(y)\Lambda(y) \), it follows that \( R(x) \leq r/\Lambda(x) \) for all \( x \in \{1, \ldots, y\} \). We will show that \( R(x) = r/\Lambda(x) \) for all \( x \in \{1, \ldots, y\} \). Suppose for a contradiction that this is false, and let \( z \) be the smallest \( x \in \{1, \ldots, y\} \) such that \( R(x) < r/\Lambda(x) \). We first argue that \( z > 1 \), for if \( z = 1 \), then we could increase \( R(1) \) a bit, without disrupting \( R(x)/x \) decreasing in \( x \). This would then allow us to scale down \( R(x), x > 1 \) and reduce \( r = R(y)\Lambda(y) \), contradicting the optimality of Problem (3). Consequently, \( z > 1 \). We now argue that \( R(z)/z = R(z - 1)/(z - 1) \). Otherwise, we could increase \( R(z) \) and decrease \( R(x), x > z \), and again get a contradiction. Then from \( r = R(z - 1)\Lambda(z - 1) > R(z)\Lambda(z) = R(z - 1)z/(z - 1)\Lambda(z) \), we obtain \( g(z - 1) > g(z) < g(y) \), but this contradicts the unimodality of \( g(x) \). Consequently, \( R(x) = r/\Lambda(x) \) for all \( x \in \{1, \ldots, y\} \). \( \Box \)

This implies that the \( R(x)\Lambda(x) = r \) for all \( x \in \{1, \ldots, y\} \), so the maximizers of \( R(x)\Lambda(x) \) are consecutive. It also implies that \( R(x)\lambda(x) = rh(x) \) for all \( x \in \{1, \ldots, y\} \).

**Proof of Theorem 5.**

**Proof.** We will work with the condition \( 1 = E[R(X)] = \sum_{x \in \mathcal{M}} R(x)\lambda(x) \) to show that in the worst case the distribution of \( X \) is geometric with mean \( E[X] = y \), and from this conclude that \( r \geq 1/(3 - 2/y) \geq 1/3 \). Let \( h(y) = \theta \) for some \( \theta > 0 \) and recall from Lemma 7 that \( h(x) = \theta \) for all \( x \in \mathcal{M}, x \geq y \). This implies that \( X \) has tail probabilities \( \Lambda(x) = (1 - \theta)^{x-1} \) for all \( x \in \mathcal{M}, x \geq y \). For any such \( X \), we have \( E[X|X \geq y] = E[X|X > y - 1] \leq y - 1 + E[X] = y - 1 + 1/\theta \) with the upper bound attained by the distribution with tail probabilities \( \Lambda(x) = (1 - \theta)^{x-1} \) for all \( x \geq y \). Finally notice that \( R(x) = r/\Lambda(x), x < y \), implies that \( R(x)\lambda(x) = rh(x) \leq r\theta \). Consequently,

\[
1 = E[R(X)] = \sum_{x=1}^{y-1} R(x)\lambda(x) + \sum_{x>y-1} R(x)\lambda(x)
= r \sum_{x=1}^{y-1} h(x) + \frac{r}{y} \sum_{x>y-1} x \frac{\lambda(x)}{\Lambda(y)}
= r \sum_{x=1}^{y-1} h(x) + \frac{r}{y} E[X|X > y - 1]
\leq r \left( (y - 1)\theta + \frac{1}{y} (y - 1 + 1/\theta) \right)
\]
where the bound is attained by the geometric distribution with hazard rate \( h(x) = \theta \) and mean \( E[X] = 1/\theta \). Because Lemma 6, it follows that \( 1/(y+1) < \theta \leq 1/y \). Since the expression is brackets is convex in \( \theta \), the maximum is at an extreme point and it is easy to see that it is maximized when \( \theta = 1/y \), or equivalently when \( E[X] = y \). Evaluating the last equation at \( \theta = 1/y \), we see that
\[
1 \leq r(3 - 2/y), \text{ or equivalently } r \geq 1/(3 - 2/y) \geq 1/3.
\]
\( \square \)

12.3. Proof in Section 9

Proof of Theorem 7.

Proof. Given the sets \( S_x, \forall x = 1, ..., m \) fixed, we maximize solely over prices. That is, we want to compute
\[
\max_r R(r|S) = \sum_{x=1}^{m} \lambda(x) R_x(r|S_x) = \sum_{x=1}^{m} \lambda(x) \sum_{i \in S_x} r_i P(i,S_x),
\]
where \( R_x(r|S_x) \) is the expected revenue from consumer who has consideration set \( S_x \) and \( P(i,S_x) = \frac{\exp(a_i - \beta r_i)}{1 + \sum_{k \in S_x} \exp(a_k - \beta r_k)} \) is the probability of choosing item \( i \) if \( i \) is in consumer \( x \)'s consideration set. Taking partial derivative of \( P(i,S_x) \) with respect to product \( i \)'s and \( k \)'s prices \( r_i \) and \( r_k \) respectively, we have the following formulas:
\[
\frac{\partial P(i,S_x)}{\partial r_i} = \beta P(i,S_x)(P(i,S_x) - 1), \quad \text{and}
\]
\[
\frac{\partial P(i,S_x)}{\partial r_k} = \beta P(i,S_x) P(k,S_x).
\]
Taking the first order derivative of the expected revenue \( R_x(r) \) with respect to \( r_i \), we obtain
\[
\frac{\partial R_x(r|S_x)}{\partial r_i} = P(i,S_x) + r_i \frac{\partial P(i,S_x)}{\partial r_i} + \sum_{k \neq i} r_k \frac{\partial P(k,S_x)}{\partial r_i}
\]
\[
= \beta P(i,S_x) \left\{ \frac{1}{\beta} + \sum_{k \in S_x} r_k P(k,S_x) - r_i \right\}
\]
\[
= \beta P(i,S_x) \left\{ \frac{1}{\beta} + R_x(r) - r_i \right\}.
\]
Let \( x(i) \equiv \{x : i \in S_x\} \). That is, \( x(i) \) denotes the page where item \( i \) is displayed. Taking partial derivative of the total expected revenue with respect to \( r_i \), we obtain

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

Equivalently,

\[
\sum_{l=x(i)}^m \lambda(l) P(i, S_l) \{ \frac{1}{\beta} + R_l(r|S_l) \} = \sum_{l=x(i)}^m \lambda(l) P(i, S_l) r_i. \tag{16}
\]

Notice that equation (16) is satisfied either when \( P(i, S_l) = 0 \), \( \forall l = 1, ..., m \), i.e., when \( r_i = +\infty \) meaning item \( i \) is priced out of the market; or when

\[
R_l(r|S_l) - r_i = 0,
\]

\[
r_i = \frac{\sum_{l=x(i)}^m \lambda(l) P(i, S_l) \{ \frac{1}{\beta} + R_l(r|S_l) \}}{\sum_{l=x(i)}^m \lambda(l) P(i, S_l)}. \tag{17}
\]

It is easy to check that in the above equation (17), the right hand side is invariant for all \( i \) and \( k \) such that \( x(i) = x(k) \), i.e, invariant for all products that are displayed on the same page. Thus, for every finitely priced item \( i \), there should 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 price, notice that equation (17) tells us that the price is a weighted average of \( \frac{1}{\beta} + R_l(r|S_l) \), where \( R_l(r|S_l) \), the expected revenue from consumer \( l \), must be nondecreasing. To see, suppose that there is a drop at \( l \). This could only be because we set prices too low for products in page \( l \). Then we would just raise their prices and therefore increase the expected revenue consumer \( l \). Moreover, we argue the expected revenue from the later market can also be increased due to the unimodularity of \( R_l(\cdot) \) function with respect to all \( r \)’s (See Gallego, Li and Beltran 2016). And the saddle point will shift right as more products are considered. According to Theorem (8), indeed all products should be introduced, so \( R_l(r|S_l) \) must be monotonically increasing in \( l \). Thus, the page-level price must be monotonically increasing in \( l \).

To prove the optimal prices for all products are finite, we want to show two facts: (1) the optimal price for all pages, i.e., \( \theta_1, ..., \theta_m \) are finite; and (2) supposing that at optimality, some products are priced at infinity, then the item offering cannot be optimal.
The first fact is equivalent to having $\theta_m < +\infty$, since we know that the page-level price $\theta_x$ is monotonically increasing in $x$. The first-order condition given by page $m$ implies $\theta_m = \frac{1}{\beta} + R_m(\theta|S_m)$, which proves the finiteness of $\theta_m$.

Consider the second fact that we need to prove. Notice that having some prices equal to infinity is equivalent to pricing the corresponding products out of the market, or not displaying them at all. Equivalently, there is some page $x$ displaying fewer than $p$ products. Therefore we can introduce one more item without increasing the total attractiveness of page $x$. This can be achieved by increasing products’ prices on page $x$. In this situation, the expected revenue from consumer segment $x' = 1 \ldots x - 1$ will not be affected, and the expected revenue from segment $x$ and larger will strictly increase. That is, $R_{x'}(r|S_{x'})$, $x' = x \ldots m$, will become larger. Thus, the total expected revenue will strictly increase when we fill all pages with finitely priced products. \hfill □

**Proof of Theorem 8.**

**Proof.** Suppose that the page configuration $S_x$ is fixed for all $x = 1, \ldots, m$. Given the quality vector $a = (a_1, \ldots, a_n)$, we want to optimize with respect to prices. That is, we want to solve

$$R(a) = \max_r \ R(r|S) \equiv \sum_{x=1}^m \lambda(x)R_x(r|S_x). \tag{18}$$

Now, suppose we can change the vector $a$. We wish to determine how the optimal revenue $R(a)$ will change. Let us take the partial derivative of $R(a)$ with respect to $a_i$.

$$\frac{\partial R(a)}{\partial a_i} = \sum_{l=x(i)}^{m} \lambda(l) \frac{\partial R_l(r(a)|S_l)}{\partial a_i}$$

$$= \sum_{l=x(i)}^{m} \lambda(l) \{r_i(a) - R_l(r(a)|S_l)\}P(i,S_l), \tag{19}$$

where $r(a)$ is the optimal price vector under quality vector $a$, which must satisfy the first-order condition given by (17), i.e, $\sum_{l=x(i)}^{m} \lambda(l)P(i,S_l)r_i(a) = \sum_{l=x(i)}^{m} \lambda(l)P(i,S_l)\{\frac{1}{\beta} + R_l(r(a)|S_l)\}$. Plug this relationship into (19), we obtain

$$\frac{\partial R(a)}{\partial a_i} = \frac{1}{\beta} \sum_{l=x(i)}^{m} \lambda(l)P(i,S_l). \tag{20}$$
From the above relationship, we can draw several conclusions. First, the optimal revenue $R(a)$ is increasing in $a_i$ for all $i = 1, \ldots, n$. Second, suppose there are two item $i$ and $k$ with the same quality $a_i = a_k$, but $x(i) < x(k)$, in other words, $i$ is displayed before $k$ so it can be viewed by more consumer. Equation (20) tells us $\frac{\partial R(a)}{\partial a_i} > \frac{\partial R(a)}{\partial a_k}$, so if we can either increase $a_i$ or $a_k$, it is always better to prioritize $a_i$. This indicates that products should be displayed in decreasing order of their quality.

To prove that all products should be displayed, notice that eliminating item $i$ is equivalent to letting its quality $a_i$ go to negative infinity. However, since $\frac{\partial R(a)}{\partial a_i} > 0$ and it is always better to put products with higher quality ahead of products with lower quality, we can conclude that at optimality, all products should be displayed; and they are displayed in the order of their indices. Each page is filled up with products until its capacity is saturated. □

References


