Management Science

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

Offline Assortment Optimization in the Presence of an Online Channel

Daria Dzyabura, Srikanth Jagabathula

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 64, No. 6, June 2018, pp. 2767–2786
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Offline Assortment Optimization in the Presence of an Online Channel

Daria Dzyabura,[a] Srikanth Jagabathula[a]

[a] Stern School of Business, New York University, New York, New York 10012

Contact: ddzyabur@stern.nyu.edu (DD); sjagabat@stern.nyu.edu (SJ)

Received: July 27, 2015
Revised: March 16, 2016; August 18, 2016
Accepted: September 5, 2016
Published Online in Articles in Advance: June 14, 2017

https://doi.org/10.1287/mnsc.2016.2708

Abstract. Firms are increasingly selling through both offline and online channels, allowing customers to experience the touch and feel of product attributes before purchasing those products. Consequently, the selection of products offered offline affects the demand in both channels. We address how firms should select an optimal offline assortment to maximize profits across both channels; we call this the *showcase decision* problem. We incorporate the impact of physical evaluation on preferences into the consumer demand model. Under this model, we show that the decision problem is NP-hard. Analytically, we derive optimal results for special cases and near-optimal approximations for general cases. Empirically, we use conjoint analysis to identify changes in consumer preferences resulting from physically evaluating products. For this application, we demonstrate gains in expected revenue of up to 40% due to accounting for the impact of offline assortment on the online sales.

History: Accepted by Vishal Gaur, operations management.
Funding: S. Jagabathula's research was supported in part by the National Science Foundation [Grant CMMI 1454310].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2016.2708.

Keywords: assortment optimization • omnichannel • choice modeling • conjoint analysis

1. Introduction

A growing number of firms are selling to their customers through both online and offline (or brick-and-mortar) channels. Selling through multiple channels allows the firm to reach customers who have differing channel preferences for purchasing. In addition, the firm can offer a wide selection of its products (at lower inventory costs) through its online channel and showcase its product line through its offline channel. Despite the growing significance of the online channel, maintaining offline channels is essential for firms because customers visit offline stores to physically inspect the products and gain tactile (or "touch-and-feel") information before making purchases. Examples include furniture purchases (from firms such as Crate & Barrel, West Elm, etc.) and apparel purchases[1] (from firms such as Bonobos, MM.LaFleur, etc.). Therefore, the selection of products that a customer is exposed to in the offline channel impacts her purchase behavior, and the firm faces the key operational problem of optimizing its offline selection with the objective of maximizing overall sales or profits.

Existing work in operations and marketing provides guidance on how firms should optimize their offerings, but most such work focuses on single-channel settings. While several such proposals exist, at the core, they rely on *restricting* customer choices to trade off losing profits by not offering low-profit products for gaining profits from switches to higher-profit products. These proposals, however, do not extend to multichannel settings in which the offline offer set does not necessarily *restrict* choices as (some) customers may purchase from the (the typically larger) assortment offered online. Furthermore, they do not account for the fact that the assortment may *change* the product the customer will purchase because of the touch-and-feel information provided by the offline channel. For example, consider a customer looking to purchase a messenger bag with a laptop compartment and the store offers a blue bag *without* the laptop compartment. In the absence of a store visit, the customer would have purchased the black bag online, but after the store visit, the customer realizes that she prefers blue to black and purchases the blue bag with a laptop compartment online. In other words, the offline channel is not only a sales channel but also an information channel. Ignoring such interactions between channels will result in suboptimal decisions.

Motivated by the above considerations, this paper studies a firm's *showcase decision*—that of determining the subset of products from the online channel to offer in the offline channel to maximize aggregate sales or profits across both channels. We focus on the following setup. A firm is selling products through an online and an offline channel. The products are differentiated but close substitutes—a single consumer purchases at most one product and the rest of the market does not offer perfect substitutes. The products are generally infrequently purchased, or comprise a large variety, so that customers can benefit from a visit to the offline store to examine the products physically. The customer is utility maximizing and purchases her maximum utility product if its utility is greater than her no-purchase utility. Products are multiattribute, and a product's utility is the sum of its attribute partworths, which capture consumers' preferences for each attribute. Product categories such as furniture, apparel and accessories, consumer electronics, etc., satisfy these assumptions. The selection of products offered online is large and fixed. The profit associated with each product is exogenously specified and decomposes into the sum of the profits from the constituent attributes. The objective of the firm is to choose a selection of products from the online assortment to offer offline to maximize the expected sales or profits from both channels.

In the context of the above setup, this work makes three key contributions: (a) a novel utility-based model to capture the interactions between the online and offline channels, (b) analytical results on the structure of the optimal offer sets, and (c) a scalable integer programming (IP)-based optimization algorithm to solve the firm's showcase decision. We also validate our modeling assumptions and our methods using real-world preference data on messenger bags.

The modeling contribution of this work is to extend the standard utility model to capture the impact of physical evaluation through *changing* partworths. The standard utility-based models suppose that the utility obtained from a product decomposes into the sum of the partworths (or valuations) of the attributes comprising the product. They assume that customers arrive at the partworths by evaluating the products, so they are fixed and known to the customers. However, when products are sold online, customers may find it difficult to evaluate some attributes based only on their online descriptions or pictures; for instance, it may be difficult ascertain how large the large size is, how bright the blue color is, etc. For such attributes, the consumer may learn her preferences by physically evaluating products with those attributes. We capture this learning by allowing the partworth of an attribute to *change* upon physical evaluation. Thus, for each attribute, the customer has an online (or preevaluation) and an offline (or postevaluation) partworth, and the difference between them quantifies the information gained from physical examination of that attribute.

Under the above model, we derive analytical results for the structure of the optimal solution to obtain insights on how the firm's decisions differ from the single-channel setting when there is also an online channel. We consider two separate settings: (a) all customers visit both channels and (b) only a portion of customers visits both channels (the *online* segment), while the others purchase only what is available in the offline store (the *offline* segment). In the former case, which we term the *pure showcase* setting, the offline channel acts only as an information channel, while in the latter case, which we term the *general showcase* setting, it acts as both an information channel and a sales channel. We make the standard assumption that when given the attribute partworths, the customers make choices according to a multinomial logit (MNL) model. In the pure showcase setting, we show that to maximize sales (Theorem 3.1), the firm must offer the attributes that are undervalued (more attractive after physical evaluation) and hide the attributes that are overvalued (less attractive after physical evaluation) by the customers. On the other hand, to maximize profits (Theorem 3.2), the firm must offer the most profitable undervalued attributes and the least profitable overvalued attributes. In doing this, the firm provides information to its customers, resulting in an increase (decrease) in the attractiveness of the most (least) profitable attributes; this shifts the demand to the most profitable attributes. In the general showcase setting, we show that to maximize sales (Theorem 3.5), the firm must offer all the undervalued attributes and hide only a subset of the overvalued attributes. Precisely which overvalued attributes are hidden depends on their corresponding magnitudes of attractiveness. Intuitively, the overvalued attributes that drive a large amount of offline sales should be offered but the rest hidden. In contrast, when there is no online channel, the firm must offer all the attributes to maximize sales and offer the most profitable attributes to maximize profits.

Our second set of results addresses computation. We show that the pure showcase sales maximization problem can be efficiently solved. The pure showcase profit maximization and the general showcase sales (and, hence, profit) maximization problems are NP-hard to solve (Theorems 3.3 and 3.6). However, we show that the pure showcase profit maximization problem and a natural relaxation of the general showcase profit maximization problem admit fully polynomial time approximation schemes (FPTASs); see Theorems 3.4 and 3.7. Applying the ideas used to construct the FPTAS, we propose an IP-based heuristic to determine the profit maximizing subset. Using a simulation study, we show that (a) our IP-based heuristic scales to large, practical-sized problems, and (b) the solutions obtained from the IP-based heuristic provide significantly higher profits and sales when compared to the solutions from the standard revenue-ordered (RO) or greedy heuristics, described in detail in Section 4 (see Tables 1 and 2).

Finally, to illustrate the value of our methodology, we analyzed customer preference data on messenger bags. The data were obtained from a conjoint study and demonstrate that customers' valuations of many attributes change significantly when they evaluate the products in an offline as opposed to an online channel (see Table 3). Using the proposed IP-based heuristic, we then computed optimal sales/revenue maximizing assortments for various sizes of the offline segment. We found that significant gains in sales/revenues (up to 40% in our study) were attained. We also gained the following broad insight into the structure of the sales maximizing assortment when there is a constraint on the size of the offered assortment: it is optimal to offer a mix of "popular" and "informative" products; the popular products have high utilities and drive sales in the offline channel, whereas the informative products expose customers to undervalued attributes and drive sales in the online channel.

1.1. Related Work

This research builds on existing research in marketing, operations, and information systems. We categorize relevant existing research into work pertaining to single- and multiple-channel settings.

In the single-channel setting, our work is related to work on product line optimization in marketing and work on assortment optimization in operations. Product line optimization focuses on a canonical manufacturer selecting a set of products to offer with the objective of maximizing consumer welfare, market share, or profit. In this literature, products are represented in a multiattribute space, and the optimal product line is constructed directly from the attribute levels, given their corresponding partworths. The problem has been shown to be NP-hard to solve in general. Several researchers have proposed heuristic solutions: Kohli and Krishnamurti (1987, 1989) and Kohli and Sukumar (1990) proposed dynamic programming (DP)-based greedy heuristics, Balakrishnan and Jacob (1996) and Fruchter et al. (2006) proposed genetic-algorithm-based heuristics, Dobson and Kalish (1988, 1993) and Green and Krieger (1985) proposed a priori selecting some candidate products from a large number of feasible products and then selecting a product line from only this candidate set. Belloni et al. (2008) presented a comparison of the performance of different heuristics for product line design and found that greedy heuristics perform well.

The work on assortment optimization in operations management (OM) focuses on a canonical retailer selecting a subset of products to offer to maximize expected profits (for a review, see Kök and Fisher 2007). Because the viewpoint is that of a retailer, the problem is typically solved in product space rather than in attribute space. The focus of this body of work has been on deriving either exact or approximate optimization algorithms under various choice models: multinomial logit (Talluri and Van Ryzin 2004, Rusmevichientong et al. 2010, Davis et al. 2013), nested logit (Gallego and Topaloglu 2014, Alptekinoğlu and Grasas 2014, Davis et al. 2014, Feldman and Topaloglu 2015), d-level nested logit (Li et al. 2015), mixed logit (Rusmevichientong et al. 2014), and the locational choice model (Gaur and Honhon 2006, Alptekinoğlu and Corbett 2010, Ulu et al. 2012). Jagabathula and Rusmevichientong (2017) jointly optimize assortment and prices; Ghoniem and Maddah (2015) jointly optimize assortment, prices, and inventories; and Ghoniem et al. (2016) focus on the problem of jointly optimizing the assortments and prices of a firm selling products belonging to complementary categories.

We contribute to both of the above bodies of work by focusing on two channels rather than a single channel. Unlike most work in OM, we focus on the attribute space rather than the product space. Similarly, unlike most work in marketing, we focus on optimization and computational issues.

In the multiple-channel setting, our work is related to the literature on marketing and information systems that focuses on the interaction between online and offline channels. Most of this work has focused on the setting in which a customer has the option to buy the *exact same* product online and offline and on determining firms' pricing decisions and equilibrium price outcomes (Brynjolfsson et al. 2009, Forman et al. 2009, Mehra et al. 2013). We contribute to this literature by focusing on multiattribute products that are sold by a single firm and studying assortment rather than pricing decisions.

2. Model

This section provides the precise details of the decision problem and the corresponding consumer choice model we consider. Our objective is to solve a firm's *showcase decision* of determining the assortment of products to "showcase" in the offline channel to maximize the profit from both offline and online channels. We focus on a firm selling products through an online and an offline channel. The products are reasonably expensive, infrequently purchased, and require a high-involvement buying decision; consequently, customers can benefit from an offline store visit. Furniture, messenger bags, apparel, etc., are good examples. The products are close substitutes but are differentiated along K prespecified attributes. We assume that attribute k takes values from the set $\mathcal{L}_k := \{0, 1, \ldots, L_k - 1\}$ of L_k discrete levels. We let $\mathcal{X}$ denote the set of all feasible products, with each product represented by a length K feature vector $\mathbf{x} \in \mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \times \mathcal{L}_K$, where $x_k \in \mathcal{L}_k$ denotes the level of attribute k in the product. For example, suppose products are described by two attributes, size and color, with size being either small or large and color being either blue or black. Then, $K = 2$ and $L_1 = L_2 = 2$, with the vectors $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$ denoting small blue, small black, large blue, and large black products, respectively.

2.1. Omnichannel Selling

The objective of the firm is to determine the selection of products from the online channel to showcase in the offline store to maximize the total profit across *both* the online and the offline channels. We distinguish two settings: the *pure showcase* and the *general showcase*. In a pure showcase setting, the firm does not carry inventory in offline stores and sells products only through the online channel; examples include bulky or customized purchases such as furniture or kitchen cabinetry, or apparel from firms that sell online but showcase their products in offline showrooms (such as Bonobos). In the general showcase setting, by contrast, the firm carries inventory and sells through *both* the channels; examples include most firms that sell products through multiple channels. Mathematically, the pure showcase decision is a special case of the general showcase decision. However, we study them separately because the (simpler) pure showcase problem is more tractable, while encompassing a wide range of practically important applications.

2.2. Customer Choice Model

The showcase decision affects the purchase behavior of *only* the customers who visit the offline store. Among these customers, we distinguish two types: the *offline type* and the *online type*. The offline type chooses from only the selection of products offered offline, whereas the online type chooses from the entire offered selection (online or offline) and is willing to purchase from either channel. Note that both types may purchase from either channel, depending on their channel purchase preferences and their preferred products. We do not distinguish purchases from different channels and focus on maximizing the combined profit/sales from both channels. We let α and $1-\alpha$, for some $\alpha \in [0,1]$, denote the sizes of the offline and online segments, respectively. For the pure showcase setting described above, all the customers are required to choose from the selection of products online; therefore, the customers are composed entirely of the online type ($\alpha = 0$). The general showcase setting corresponds to $\alpha \in (0,1]$. The case with $\alpha = 1$, in which none of the customers purchase from the online channel, is similar to the single-channel assortment optimization that has been studied in the operations management literature. However, as discussed in greater detail below, existing results do not apply because of the presence of product features.

To model the purchase behavior of customers, we start with the standard multiattribute utility model (see Green and Rao 1971, Green and Srinivasan 1990). The customer's utility for product $\mathbf{x} \in \mathscr{X}$ is equal to the sum of the utility partworths of the features present in the product: $U(\mathbf{x}) = \sum_{k=1}^{K} \tilde{u}_k(x_k) + \beta_{\text{price}} \cdot \pi_{\mathbf{x}} + \varepsilon_{\mathbf{x}}$, where $\pi_{\mathbf{x}}$ is the product price, β_{price} is the price coefficient, $\tilde{u}_k(x_k) = \sum_{l \in \mathscr{L}_k} \tilde{w}_{kl} \cdot \mathbb{1}_l[x_k]$ is the utility obtained from attribute k, $\tilde{w}_{kl}$ is the utility partworth assigned to level l of attribute k, and $\mathbb{1}_l[x_k]$ is the indicator variable taking value 1 if $x_k = l$ and 0 otherwise. The term $\varepsilon_{\mathbf{x}}$ is the error term that captures any unexplained variance. Customers are utility maximizing, so they purchase the product with the maximum utility from any choice set.

We extend the standard model to capture how exposure to a (subset) of products in the offline channel impacts the purchase behavior of customers. To model this impact, we suppose that customers associate different utility partworths with each feature, depending on whether they were exposed to the feature in the offline channel or not. Particularly, with respect to the feature corresponding to attribute k and level l, the customer associates utility partworth $\tilde{w}_{kl}^{\text{off}}$ if she was exposed to feature (k,l) in the offline store and $\tilde{w}_{kl}^{\text{on}}$ if she was *not* exposed. The difference between the online and offline partworths for an attribute-level may be interpreted as being caused by the information gained by the customer from "touching and feeling" the particular attribute-level in the offline store. For example, a customer may think she likes the "large" size but change her mind upon physical inspection. Then, for (k,l) representing the large size, we have that $\tilde{w}_{kl}^{\text{on}} > 0$ but $\tilde{w}_{kl}^{\text{off}} < 0$, indicating that after physical inspection, the customer incorporates the information that she dislikes the large size in her purchases, be they online or offline.

To be precise, we let $\mathbf{S} = (S_1, S_2, \ldots, S_K)$ denote the collection of attribute levels that are offered in the offline store, with $S_k \subseteq \mathscr{L}_k$ denoting the subset of levels for attribute k that are offered. Because the firm selects the offline assortment from the online assortment, we suppose that the firm offers the universe $\mathscr{X}$ of feasible products online. Now, consider a customer who has visited the offline store. For any product $\mathbf{x} \in \mathscr{X}$, the customer assigns the utility

$$U_{\mathbf{S}}(\mathbf{x}) = \begin{cases} \displaystyle\sum_{k=1}^{K} \tilde{u}_k^{\text{off}}(x_k) + \beta_{\text{price}} \cdot \pi_{\mathbf{x}} + \varepsilon_{\mathbf{x}} & \text{if } \mathbf{x} \text{ offered offline,} \\ \displaystyle\sum_{k=1}^{K} \tilde{u}_k^{\text{on}}(x_k, S_k) + \beta_{\text{price}} \cdot \pi_{\mathbf{x}} + \varepsilon_{\mathbf{x}} & \text{otherwise,} \end{cases} \tag{1}$$

where

$$\tilde{u}_k^{\text{on}}(x_k, S_k) = \sum_{l \in S_k} \tilde{w}_{kl}^{\text{off}} \cdot \mathbb{1}_l[x_k] + \sum_{l \notin S_k} \tilde{w}_{kl}^{\text{on}} \cdot \mathbb{1}_l[x_k] \quad \text{and}$$
$$\tilde{u}_k^{\text{off}}(x_k) = \sum_{l \in \mathscr{L}_k} \tilde{w}_{kl}^{\text{off}} \cdot \mathbb{1}_l[x_k].$$

We use the notation $\tilde{u}_k^{\text{off}}(x_k)$ to emphasize that the utility of a product, when the customer is evaluating offline, is independent of what else is on offer. On the other hand, the utility partworth $\tilde{u}_k^{\text{on}}(x_k, S_k)$ used for a product that is offered online, but not offline, depends on whether x_k is offered offline (as part of S_k) or not. Our notation is consistent: indeed, $\tilde{u}_k^{\text{on}}(x_k, S_k) = \tilde{u}_k^{\text{off}}(x_k)$ whenever x_k is offered offline, i.e., $x_k \in S_k$. We note that according to our model assumptions, consumers use the (same) offline price coefficient *both* online and offline. This is because we are focusing only on consumers who visit the offline channel, where they are exposed to the price attribute.

A key aspect of our model is that physical exposure to product A affects the utility of product B if B shares attributes with A, even in the absence of physical exposure to B. For example, exposure to a large black bag affects the utility of a large blue bag, even if the blue bag was not physically evaluated. Because the utilities and, hence, the purchase probabilities of products in the online channel are affected by features the customer was exposed to in the offline channel, the profit and revenue from *both* channels are affected by the offline assortment; this interaction makes the assortment problem challenging.

We assume that the price of a product can be decomposed into the sum of the prices of its constituent attributes. Letting $\tilde{\pi}_{kl} \geq 0$ denote the price associated with level l of attribute k, the price $\pi_{\mathbf{x}}$ of product $\mathbf{x}$ is equal to $\sum_{k=1}^{K} \sum_{l=1}^{L_k} \tilde{\pi}_{kl} \cdot \mathbb{1}_l[x_k]$. This price structure arises when the firm adopts a "hedonic price model," expressing product price in terms of included attributes, to obtain a simple pricing scheme for an exponentially large configurable product space. It is commonly used in literature (Cohen et al. 2016, Randall et al. 1998, Rodríguez and Aydın 2011) and practice as "optional product pricing" or "feature-based pricing" (e.g., \$10 extra for the black color, \$20 extra for the large size, etc.) for pricing configurable products such as computers, furniture, cars, etc. With this assumption, the utility expressions above simplify as follows. Letting w_{kl}^{c} denote $\tilde{w}_{kl}^{\mathsf{c}} + \beta_{\text{price}} \tilde{\pi}_{kl}$, for $\mathsf{c} \in \{\text{on}, \text{off}\}$, and defining

$$u_k^{\text{on}}(x_k, S_k) = \sum_{l \in S_k} w_{kl}^{\text{off}} \cdot \mathbb{1}_l[x_k] + \sum_{l \notin S_k} w_{kl}^{\text{on}} \cdot \mathbb{1}_l[x_k] \quad \text{and}$$

$$u_k^{\text{off}}(x_k) = \sum_{l \in \mathscr{L}_k} w_{kl}^{\text{off}} \cdot \mathbb{1}_l[x_k],$$

we obtain that $U_S(\mathbf{x}) = \sum_{k=1}^{K} u_k^{\text{off}}(\mathbf{x}) + \varepsilon_{\mathbf{x}}$ if $\mathbf{x}$ is offered offline, and $\sum_{k=1}^{K} u_k^{\text{on}}(\mathbf{x}) + \varepsilon_{\mathbf{x}}$ otherwise.

We make the standard logit assumption that the idiosyncratic terms $\varepsilon_{\mathbf{x}}$ are independent and identically distributed standard Gumbel random variables for $\mathbf{x} \in \mathscr{X}$. Furthermore, suppose that $M \subseteq \mathscr{X}$ is the assortment offered offline, and the universe $\mathscr{X}$ is offered online. Then, the probability that an online-type customer chooses product $\mathbf{x}$ from the selection $\mathscr{X}$ is given by

$$P_{\mathbf{x}}(M) = \frac{\exp(\sum_{k=1}^{K} u_k^{\text{on}}(x_k, S_k^M))}{1 + \sum_{\mathbf{y} \in \mathscr{X}} \exp(\sum_{k=1}^{K} u_k^{\text{on}}(y_k, S_k^M))}, \tag{2}$$

where $\mathbf{S}^M = (S_1^M, S_2^M, \ldots, S_K^M)$ denotes the set of features that the products in M are composed of, i.e., $l \in S_k^M$ if and only if $x_k = l$ for some $\mathbf{x} \in M$. Note that we have made the standard assumption that the mean utility of the no-purchase option is 0. On the other hand, the probability that an offline-type customer purchases product $\mathbf{x} \in M$ is given by

$$Q_{\mathbf{x}}(M) = \frac{\exp(\sum_{k=1}^{K} u_k^{\text{off}}(x_k))}{1 + \sum_{\mathbf{y} \in M} \exp(\sum_{k=1}^{K} u_k^{\text{off}}(y_k))}.$$

We make the following remarks. The choice probability $Q_{\mathbf{x}}(M)$ is similar to classical choice probability expression under the MNL choice model in which the customer is restricted to choose from the subset M. The choice probability expression $P_{\mathbf{x}}(M)$ for the online segment differs from $Q_{\mathbf{x}}(M)$ in two key ways: (a) the utility partworths depend on whether a customer is exposed to a feature offline or not, and (b) the customer chooses from the entire collection $\mathscr{X}$ of products offered online and offline. Because of these distinctions of choice probability expressions, the optimization problems we consider are different in structure from the classical assortment optimization problems studied in the literature. Finally, for the above expressions for choice probabilities to be valid, we require only that the union of the sets of products offered online and offline be equal to $\mathscr{X}$, and not necessarily that the online assortment be equal to $\mathscr{X}$. However, because online assortments tend to be larger than offline assortments, we make the assumption that the online assortment is equal to $\mathscr{X}$.

2.3. Firm's Showcase Decision

In the context of the above model, we consider the following decision problems:

$$\max_{M \subseteq \mathscr{X}} \sum_{\mathbf{x} \in \mathscr{X}} P_{\mathbf{x}}(M), \tag{Pure Showcase Sales Max}$$

$$\max_{M \subseteq \mathscr{X}} \sum_{\mathbf{x} \in \mathscr{X}} p_{\mathbf{x}} P_{\mathbf{x}}(M), \tag{Pure Showcase Profit Max}$$

and

$$\max_{M \subseteq \mathscr{X}} \alpha \sum_{\mathbf{x} \in M} p_{\mathbf{x}} Q_{\mathbf{x}}(M) + (1 - \alpha) \sum_{\mathbf{x} \in \mathscr{X}} p_{\mathbf{x}} P_{\mathbf{x}}(M), \tag{General Showcase Profit Max}$$

where $p_{\mathbf{x}}$ is the net profit obtained from the sale of product $\mathbf{x}$. Problem (Pure Showcase Sales Max)[2] is the simplest nontrivial decision problem we consider. When the firm is selling through a single channel, it is always optimal (for maximizing sales) to offer all the products when there is no capacity constraint. As we show below, this simple structure no longer holds in the presence of two channels. For profit maximization problems, we assume that the profit from a product can be decomposed into the sum of the profits from its constituent attributes. Letting $r_{kl} \geq 0$ denote the profit margin associated with level l of attribute k, the net profit $p_{\mathbf{x}}$ obtained from selling product $\mathbf{x}$ is equal to $\sum_{k=1}^{K}\sum_{l=1}^{L_k} r_{kl} \cdot \mathbb{1}_l[x_k]$. Because product prices are assumed to decompose into sums of attribute prices, the profits also decompose if we assume that the product cost decomposes into the sum of individual attribute costs. Such cost structure is commonly assumed in literature (Belloni et al. 2008) and may be backed out from the product's "bill of materials." We show that the profit maximization decision problems are NP-hard to solve in general but admit fully polynomial time approximation schemes.

3. The Product Showcase Decision

We now discuss our results for solving the showcase decision problems introduced above. For the development below, we make the common assumption (e.g., see Kohli and Krishnamurti 1989) that the product universe $\mathscr{X}$ is full factorial, i.e., $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_K$, so that every *nonprice* feature combination is feasible. With the full-factorial assumption, the problem (Pure Showcase Sales Max) becomes tractable, but as we show below both problems (Pure Showcase Profit Max) and (General Showcase Profit Max) remain NP-complete. In fact, even the sales maximization problem (with $p_{\mathbf{x}}$ set to 1 for all $\mathbf{x}$ in (General Showcase Profit Max)) under the general showcase setting is NP-complete. However, there may be cases in which the full-factorial assumption is not reasonable. The optimization problems become significantly harder if we allow for arbitrary constraints on feature combinations, even in the single-channel setting.

3.1. Pure Showcase Decision

In the pure showcase setting, all customers who visit the offline store are assumed to also visit the online store before making a purchase. Because all customers choose from the same selection $\mathscr{X}$ of products, only the *offered attributes*, and not the offered products, impact the choice probabilities and, consequently, the profits. Therefore, the choice probability expression in (2) may be simplified to

$$P_{\mathbf{x}}(\mathbf{S}) = \frac{\exp(\sum_{k=1}^{K} u_k^{\text{on}}(x_k, S_k))}{1 + \sum_{\mathbf{y}\in\mathscr{X}} \exp(\sum_{k=1}^{K} u_k^{\text{on}}(y_k, S_k))}, \tag{3}$$

where as above, $\mathbf{S} = (S_1, \ldots, S_K)$, with $S_k \in \mathscr{L}_k$ denoting the subset of levels for attribute k that are offered in the offline store, and we use the notation $P_{\mathbf{x}}(\mathbf{S})$, instead of the more general $P_{\mathbf{x}}(M)$. The decision problem now reduces to determining the optimal subset of attribute levels to offer in the store. Once the optimal vector of attribute-level sets $\mathbf{S}$ is determined, the assortment M of products to offer in the store is given by the Cartesian product $S_1 \times S_2 \times \cdots \times S_K$. Of course, multiple assortments[3] achieve $\mathbf{S}$, and the eventual decision may be driven by other considerations not modeled in our work, such as shelf space constraints, variety requirements, etc.

With the assumption that the profit from product $p_{\mathbf{x}}$ from product $\mathbf{x}$ decomposes as $p_{\mathbf{x}} = \sum_{k=1}^{K}\sum_{l=1}^{L_k} r_{kl} \cdot \mathbb{1}_l[x_k]$, the expected profit from offering the attribute levels $\mathbf{S}$ in the offline store equals

$$R^{\text{pure}}(\mathbf{S}) = \sum_{\mathbf{x}\in\mathscr{X}} p_{\mathbf{x}} P_{\mathbf{x}}(\mathbf{S}) = \frac{\sum_{\mathbf{x}\in\mathscr{X}}(\sum_{k,l} r_{kl}\mathbb{1}_l[x_k])\exp(\sum_{k=1}^{K} u_k^{\text{on}}(x_k, S_k))}{1 + \sum_{\mathbf{y}\in\mathscr{X}} \exp(\sum_{k=1}^{K} u_k^{\text{on}}(y_k, S_k))}.$$

The denominator of the choice probability can be simplified by noting that

$$\sum_{\mathbf{y}\in\mathscr{X}} \exp\left(\sum_{k=1}^{K} u_k^{\text{on}}(y_k, S_k)\right) = \sum_{\mathbf{y}\in\mathscr{X}} \prod_{k=1}^{K}\left(\sum_{l\in S_k} \mathbb{1}_l[y_k] e^{w_{kl}^{\text{off}}} + \sum_{l\notin S_k} \mathbb{1}_l[y_k] e^{w_{kl}^{\text{on}}}\right) = \prod_{k=1}^{K}\left(\sum_{l\in S_k} e^{w_{kl}^{\text{off}}} + \sum_{l\notin S_k} e^{w_{kl}^{\text{on}}}\right),$$

where the last equality follows from interchanging the sum and the product operators.

Using a similar simplification of the numerator, we obtain the following expression for the expected pure showcase profit:

Lemma 3.1 (Simplified Pure Showcase Profit). *Suppose we offer the collection of attribute levels represented by* $\mathbf{S} = (S_1, \ldots, S_K)$ *in the offline store. Then, the pure showcase expected profit function can be simplified as*

$$R^{\text{pure}}(\mathbf{S}) = \frac{\sum_{k=1}^{K} R_k(S_k)}{1 + 1/D(\mathbf{S})}, \quad \text{where}$$

$$R_k(S_k) = \frac{b_k + \sum_{l\in S_k} r_{kl}\delta_{kl}}{D_k(S_k)}, \; D_k(S_k) = d_k + \sum_{l\in S_k} \delta_{kl}, \; \text{and}$$

$$D(\mathbf{S}) = \prod_{k=1}^{K} D_k(S_k), \tag{4}$$

with $b_k := \sum_{l\in\mathscr{L}_k} r_{kl} e^{w_{kl}^{\text{on}}}$, $d_k := \sum_{l\in\mathscr{L}_k} e^{w_{kl}^{\text{on}}}$, *and* $\delta_{kl} := e^{w_{kl}^{\text{off}}} - e^{w_{kl}^{\text{on}}}$.

The lemma is proved in Online Appendix A, Section A.1.

3.1.1. Sales Maximization. The expected sales function is obtained by setting $p_{\mathbf{x}} = 1$ for all $\mathbf{x} \in \mathscr{X}$ in the expression for $R^{\text{pure}}(\mathbf{S})$. Because $p_{\mathbf{x}} = \sum_{k,l} r_{kl}\mathbb{1}_l[x_k]$ and $\sum_{l\in\mathscr{L}_k} \mathbb{1}_l[x_k] = 1$ for all k, setting $r_{kl} = 1/K$ for all k, l, yields $p_{\mathbf{x}} = 1$ for all $\mathbf{x} \in \mathscr{X}$. Setting $r_{kl} = 1/K$ for all k, l in the expression in (4) results in the expression $(1 + 1/D(\mathbf{S}))^{-1}$ for the expected sales from offering $\mathbf{S}$ in the offline store. The sales maximization problem now reduces to

$$\max_{\mathbf{S}\in 2^{\mathcal{L}_1}\times\cdots\times 2^{\mathcal{L}_K}} D(\mathbf{S}) = \max_{\mathbf{S}\in 2^{\mathcal{L}_1}\times\cdots\times 2^{\mathcal{L}_K}} \prod_{k=1}^{K} D_k(S_k) = \prod_{k=1}^{K}\left(\max_{S_k\in 2^{\mathcal{L}_k}}\left[d_k+\sum_{l\in S_k}\delta_{kl}\right]\right),$$

where 2^S denotes the power set $\{S'\subseteq S\colon S'\neq\varnothing\}$ for any set S, and the last equality follows because the optimization problem is separable in k.

It is immediately seen that an optimal solution $\mathbf{S}^*=(S_1^*,\ldots,S_k^*)$ of the above optimization problem is such that $S_k^*=\{l\in\mathcal{L}_k\colon \delta_{kl}\geq 0\}$ if $\{l\in\mathcal{L}_k\colon \delta_{kl}\geq 0\}\neq\varnothing$ and $S_k^*=\{l_k^*\}$, where $l_k^*\in\mathcal{L}_k$ achieves the maximum, i.e., $\delta_{k,l_k^*}=\max_{l\in\mathcal{L}_k}\delta_{kl}$. To simplify notation, let $\mathcal{L}_k^+$ denote the set $\{l\in\mathcal{L}_k\colon \delta_{kl}\geq 0\}$, and let $\mathcal{L}_k^-$ denote the set $\{l\in\mathcal{L}_k\colon \delta_{kl}\}<0$. Because $\delta_{kl}=e^{w_{kl}^{\text{off}}}-e^{w_{kl}^{\text{on}}}$, $\mathcal{L}_k^+$ comprises the set of undervalued attributes (for which $w_{kl}^{\text{off}}\geq w_{kl}^{\text{on}}$) and $\mathcal{L}_k^-$ comprises the set of overvalued attributes.

Our argument above shows that the sales maximizing subset of attribute levels has the following intuitive structure: include undervalued attribute levels and *exclude* overvalued attribute levels; if, for a particular attribute k, all the levels are overvalued, then offer the least overvalued level. We summarize this result as the following theorem:

Theorem 3.1 (Pure Showcase Sales Max Solution). *The optimal solution to the problem* (Pure Showcase Sales Max) *is* $\mathbf{S}^*=(S_1^*,\ldots,S_k^*)$ *such that*

$$S_k^*=\begin{cases}\{l\in\mathcal{L}_k\colon \delta_{kl}\geq 0\} & \text{if } \mathcal{L}_k^+\neq\varnothing,\\ \{l_k^*\} & \text{otherwise},\end{cases}$$

where $\delta_{kl}=e^{w_{kl}^{\text{off}}}-e^{w_{kl}^{\text{on}}}$, $\mathcal{L}_k^+=\{l\in\mathcal{L}_k\colon \delta_{kl}\geq 0\}$, *and* $l_k^*\in\mathcal{L}_k$ *is such that* $\delta_{kl_k^*}=\max_{l\in\mathcal{L}_k}\delta_{kl}$.

3.1.2. Profit Maximization. The structure of the profit-maximizing subset of attribute levels is more complex. Exploiting the profit expression in (4), we establish the following result:

Theorem 3.2 (Pure Showcase Profit Max Solution Structure). *Any optimal solution* $\mathbf{S}^*=(S_1^*,\ldots,S_k^*)$ *to the problem* (Pure Showcase Profit Max) *must satisfy*

$$\{l\in\mathcal{L}_k^+\colon r_{kl}>t_k^*\}\subseteq S_k^{*+}\subseteq\{l\in\mathcal{L}_k^+\colon r_{kl}\geq t_k^*\},$$
$$\{l\in\mathcal{L}_k^+\colon r_{kl}<t_k^*\}\subseteq S_k^{*-}\subseteq\{l\in\mathcal{L}_k^+\colon r_{kl}\leq t_k^*\},$$
$$\text{where } t_k^*:=R_k(S_k^*)-R(\mathbf{S}^*)/D(\mathbf{S}^*),$$

where S_k^+ *denotes* $S_k\cap\mathcal{L}_k^+$ *and* S_k^- *denotes* $S_k\cap\mathcal{L}_k^-$ *for any subset* $S_k\subseteq\mathcal{L}_k$.

To understand this result, suppose that the optimal solution to (Pure Showcase Profit Max) is unique. Then, Theorem 3.2 establishes that for each attribute k, the optimal set of levels consists of a profit-ordered (PO) subset of *undervalued* attribute levels and a reverse profit-ordered (RPO) subset of *overvalued* attribute levels. We call a subset of levels a PO subset if it consists of the top m most profitable levels for some m, and an RPO subset if it consists of the bottom m least profitable levels for some m. Because offering undervalued levels increases their attractiveness and offering overvalued levels decreases their attractiveness, our result provides the following intuitive suggestion: increase the attractiveness of the most profitable levels and decrease the attractiveness of the least profitable levels.

It is instructive to contrast the result of Theorem 3.2 with that of the classical single-channel setting, but when the universe $\mathcal{X}=\mathcal{L}_1\times\cdots\times\mathcal{L}_K$, consisting of all possible feature combinations. It is known that the profit maximizing assortment M^* satisfies $M^*=\{\mathbf{x}\in\mathcal{X}\colon p_{\mathbf{x}}\geq Z^*\}$, for some Z^* (Talluri and Van Ryzin 2004). To understand how this result differs from the result of Theorem 3.2, consider the following example. Suppose a firm is selling horizontally differentiated products, concretely, shirts that differ only in color (so that $K=1$). There are two types of colors: "base" colors such as black, blue, etc., that customers are familiar with and "fashion" colors such as orange, pink, etc., that are newly introduced. Suppose that base colors are overvalued and fashion colors are undervalued by the customers. Then, in the absence of an online channel, it is optimal for the firm to offer the most profitable base and fashion colors. In contrast, in the presence of an online channel, it is optimal for the firm to offer the most profitable fashion colors and the *least* profitable base colors. By doing this, the firm is providing information to the customer that the profitable fashion colors are being undervalued and the least profitable base colors are being overvalued, shifting the demand to more profitable products. This distinction makes it clear that (a) in the pure showcase setting, the offline channel acts as an "information" channel, as opposed to the single-channel setting in which the offline channel acts as a sales channel, and (b) an algorithm that can find the best subset for the single-channel case will necessarily be suboptimal for the pure showcase problem in general.

A consequence of Theorem 3.2 is that (Pure Showcase Profit Max) reduces to

$$\max_{\mathbf{S}\in Z_1\times\cdots\times Z_K}\frac{\sum_{k=1}^{K}R_k(S_k)}{1+\exp(-\sum_{k=1}^{K}\log D_k(S_k))},\tag{5}$$

where Z_k is the collection of all possible subsets S of $\mathcal{L}_k$ such that S^+ is a PO subset and S^- is an RPO subset:

$$Z_k=\{S\subseteq\mathcal{L}_k\colon S^+=\{l\in\mathcal{L}_k^+\colon \delta_{kl}\geq t^+\},\ S^-=\{l\in\mathcal{L}_k^-\colon \delta_{kl}<t^-\}\text{ for some } t^+,t^-\}.$$

Unlike for the case of sales maximization, however, Theorem 3.2 does not yield an efficient algorithm to determine the optimal solution. In particular, for each k, we must search over all possible combinations of PO subsets of $\mathcal{L}_k^+$ and RPO subsets of $\mathcal{L}_k^-$. Because there is a total of at most $|\mathcal{L}_k|^2$ such combinations for each k, in the worst case, brute force search requires searching over $O(\prod_{k=1}^K |\mathcal{L}_k|^2)$, which scales exponentially in K. In fact, solving (Pure Showcase Profit Max) is NP-hard:

Theorem 3.3 (Hardness of (Pure Showcase Profit Max)). *The following decision problem is NP-complete: for any $Q \geq 0$, is there a subset $\mathbf{S} = (S_1, \ldots, S_K)$ such that $R^{\text{pure}}(\mathbf{S}) \geq Q$?*

The theorem is proved in Online Appendix A, Section A.1. The reduction is from the NP-complete *partition* problem (Garey and Johnson 1979). The proof focuses on the special case when each attribute k has only two levels and one of the levels, say, the first one, has zero discrepancy: $w_{k1}^{\text{off}} = w_{k1}^{\text{on}}$. The decision problem then reduces to whether to offer the second level in each attribute or not. A brute forces search has $O(2^K)$ complexity. We then obtain a reduction from the partition problem to this special case.

Despite the fact that the problem is NP-hard in the worst case, we show below that the optimization problem admits a fully polynomial time approximation scheme. An algorithm is formally defined to be an ε-approximation algorithm of a maximization problem if for each problem instance and tolerance parameter $0 < \varepsilon \leq 1$, the algorithm produces a solution with objective value R such that $R^* \geq R \geq (1-\varepsilon)R^*$, where R^* is the objective value of an optimal solution. An ε-approximation algorithm is called an FPTAS if for any fixed ε, the running time of the algorithm is bounded above by a polynomial in the size of the input and $1/\varepsilon$.

In constructing the FPTAS, we use the ideas developed for the construction of an FPTAS for the classical knapsack problem (Lawler 1979). These ideas have been used in the existing literature to construct either an FPTAS or a polynomial time approximation scheme (PTAS)[4] for assortment optimization problems (see Rusmevichientong et al. 2009, Désir and Goyal 2014). However, whereas the existing body of work has considered objective functions that can be expressed as the sum of ratios of functions that are *linear* in the decision variables, our setting results in objective functions that are sums of ratios of functions that are *nonlinear* in the decision variables. As a result, constructing an FPTAS requires a treatment different from the existing work, as presented below.

We use the following general procedure to solve (5). We guess the values of the numerator and the denominator of the objective function at the optimal solution and find a solution $\mathbf{S}$ that approximately achieves the guessed values. We show below that for the given values of the numerator and denominator of the objective function, the solution that achieves them approximately can be found by solving a dynamic program in time that is polynomial in the input size and $1/\varepsilon$. Because we do not know the optimal numerator and denominator values, we search over an ε-grid of the region of possible values. We show that the number of possible grid points we need to search over is polynomial in the input size and $1/\varepsilon$. Putting everything together results in the desired FPTAS.

The most challenging step of the above procedure is to find the solution $\mathbf{S} = (S_1, \ldots, S_k)$ that approximately achieves the guessed values of the numerator and denominator. In particular, let q denote our guess of the optimal value of the numerator, and let t denote our guess of $\sum_k \log D_k(S_k)$ at the optimal solution. Our goal is to find an $\mathbf{S}$ such that

$$\sum_k R_k(S_k) \geq q \quad \text{and} \quad \sum_k \log D_k(S_k) \geq t. \tag{6}$$

We solve the above problem approximately. In particular, for a given $\varepsilon > 0$, we find $\hat{\mathbf{S}}$ (if it exists) such that

$$\sum_k R_k(\hat{S}_k) \geq q \quad \text{and} \quad \sum_k \log D_k(\hat{S}_k) \geq (1-2\varepsilon)t. \tag{7}$$

To find $\hat{\mathbf{S}}$ we use the following DP formulation. We discretize $\log D_k(S_k)$s as follows. Let $j_{S,k} := \lfloor \log D_k(S)/(\varepsilon t/K) \rfloor$ for any $S \in Z_k$, where $\lfloor x \rfloor$ is the floor function defined as the integer such that $\lfloor x \rfloor \leq x \leq \lfloor x \rfloor + 1$, for any $x \in \mathbb{R}$. Furthermore, let $\rho := \lfloor K/\varepsilon \rfloor - K$. Now define the DP value function:

$$V(k, \omega) = \max_{\mathbf{S} \in Z_1 \times \cdots \times Z_k} \sum_{k'=1}^{k} R_{k'}(S_{k'}) \quad \text{subject to} \sum_{1 \leq k' \leq k} j_{S,k'} \geq \omega.$$

Our goal is to compute $V(K, \rho)$, for which we use the following DP recursion:

$$V(k, \omega) = \begin{cases} 0 & \text{if } k = 0,\ \omega \leq 0, \\ -\infty & \text{if } k = 0,\ \omega > 0, \\ \max_{S \in Z_k} [R_k(S) + V(k-1, \omega - j_{S,k})] & \\ & \text{otherwise.} \end{cases} \tag{8}$$

We carry out the above DP for integers k and ω such that $0 \leq k \leq K$ and $0 \leq \omega \leq \rho$. Each iteration requires a search over $O(|Z_k|) = O(|\mathcal{L}_k|^2)$ elements. Therefore, the running time of the DP scales as $O(K\rho \max_k |Z_k|) = O(K^2 \max_k |\mathcal{L}_k|^2/\varepsilon)$ because $\rho = O(K/\varepsilon)$. We first show that the DP above indeed obtains the desired approximation $\hat{\mathbf{S}}$.

Lemma 3.2 (Pure Showcase DP Approximation). *For given t and q, if there exists an $\mathbf{S}$ such that*

$$\sum_k R_k(S_k) \geq q \quad \textit{and} \quad \sum_k \log D_k(S_k) \geq t,$$

then the DP (8) *terminates with* $\hat{\mathbf{S}}$ *such that*

$$\sum_k R_k(\hat{S}_k) \geq q \quad \text{and} \quad \sum_k \log D_k(\hat{S}_k) \geq (1-2\varepsilon)t.$$

The lemma is proved in Online Appendix A, Section A.1. The precise algorithm is summarized below.

Algorithm 1 (FPTAS for (Pure Showcase Profit Max))
Input. Tolerance parameter $\varepsilon > 0$ and problem inputs $R_k(S)$ and $D_k(S)$ for all $S \in Z_k$ and $k = 1, \dots, K$.
Step 1. Define $\tau_{\min} := e^{-\sum_k \log D_k(\mathscr{L}_k^+)}$ and $\tau_{\max} := e^{-\sum_k \log D_k(\mathscr{L}_k^-)}$. Create the ε grid $\mathscr{T}$ of the interval $[\tau_{\min}, \tau_{\max}]$ such that $\mathscr{T} := \{\tau_{\min}(1+\varepsilon)^i\colon i = 0, 1, \dots, I\}$ with $I = \log(\tau_{\max}/\tau_{\min})/\log(1+\varepsilon)$.
Step 2. For $\tau \in \mathscr{T}$ do
define $t = -\log \tau$:
determine $\hat{S}_\tau$ by solving the DP with inputs t and $\varepsilon/|\log \tau_{\min}|$.
Output. Subset $\hat{S}$ that maximizes the profit from the collection $\{\hat{S}_\tau\colon \tau \in \mathscr{T}\}$.

We now show that the above scheme produces an ε approximation of the optimal solution with computational complexity that is polynomial in $1/\varepsilon$ and K.

Theorem 3.4 (FPTAS for (Pure Showcase Profit Max)). *Algorithm* 1 *produces a* $1-8\varepsilon$ *optimal solution in* $O(K^2 \max_k |\mathscr{L}_k|^2 |\log \tau_{\min}| \log(\tau_{\max}/\tau_{\min})/(\varepsilon \log(1+\varepsilon)))$ *running time, where* $\tau_{\min} := e^{-\sum_k \log D_k(\mathscr{L}_k^+)}$ *and* $\tau_{\max} := e^{-\sum_k \log D_k(\mathscr{L}_k^-)}$.

3.2. General Showcase Decision

We now consider the more general setting in which the population comprises both online- and offline-type customers. Recall that both segments visit the offline store. The offline segment of customers chooses from the assortment offered offline, whereas the online segment chooses from the assortments offered both online and offline. Let α and $1-\alpha$ denote the sizes of the offline and online segments of customers, respectively, for some $\alpha \in (0,1)$. Our goal, then, is to find the profit maximizing assortment of products to carry in the store, i.e., to solve problem (General Showcase Profit Max), restated here:

$$\max_{M \subseteq \mathscr{X}} R^{\text{general}}(M) = \alpha \sum_{\mathbf{x} \in M} p_{\mathbf{x}} Q_{\mathbf{x}}(M) + (1-\alpha) \sum_{\mathbf{x} \in \mathscr{X}} p_{\mathbf{x}} P_{\mathbf{x}}(M).$$

3.2.1. Sales Maximization. The sales maximization problem is obtained by setting $p_{\mathbf{x}} = 1$ for all $\mathbf{x} \in \mathscr{X}$ in the optimization problem above. The sales maximizing assortment has the following structure:

Theorem 3.5 (General Showcase Sales Max Solution Structure). *The optimal solution* M^* *to the optimization problem* $\max_{M \subseteq \mathscr{X}} \alpha \sum_{\mathbf{x} \in M} Q_{\mathbf{x}}(M) + (1-\alpha) \sum_{\mathbf{x} \in \mathscr{X}} P_{\mathbf{x}}(M)$ *is of the form* $M^* = S_1^* \times \cdots \times S_K^*$, *with* $\mathscr{L}_k^+ \subseteq S_k$ *for all* k.

We note the following implications of Theorem 3.5. First, the sales maximizing assortment can be obtained by searching through the attribute space, as opposed to the space of subsets of products. Second, it is optimal to offer *all* the undervalued attribute levels. This is because offering an undervalued attribute level in the offline store increases its attractiveness, resulting in higher sales from both online and offline customers.

On the other hand, offering an overvalued attribute *decreases* the attractiveness of the attribute level and, hence, decreases sales from the online-type customers, but *increases* the offered selection and, hence, increases sales from the offline type. Therefore, the optimal offering of overvalued attribute levels should balance sales from both customer types. By contrast, if the market consists of only the online-type customers (reducing the problem to pure showcase sales maximization), then the offline channel becomes a pure "information" channel (as above), and it is optimal to "hide" all the overvalued attribute levels. If, on the other hand, the market consists of only the offline-type customers (reducing the problem to the classic single-channel sales optimization), then the offline channel becomes a pure "sales" channel (by restricting the selection of products), and it is optimal to offer all the overvalued attribute levels. When there is a mix of offline- and online-type customers, the offline channel becomes *both* an information and a sales channel, and the overvalued attribute levels should be chosen to balance providing (or hiding) information and driving sales from the offline channel.

Unfortunately, determining the optimal offering of overvalued attribute levels is computationally challenging. Specifically, unlike the pure showcase setting, we show that even sales maximization is NP-hard to solve:

Theorem 3.6 (Hardness of (General Showcase Profit Max)). *The following decision problem is NP-complete*: *for any* $Q \geq 0$, *is there a subset* $M \subseteq \mathscr{X}$ *such that*

$$\alpha \sum_{\mathbf{x} \in M} Q_{\mathbf{x}}(M) + (1-\alpha) \sum_{\mathbf{x} \in \mathscr{X}} P_{\mathbf{x}}(M) \geq Q?$$

The theorem is proved in Online Appendix A, Section A.2. The reduction is from the partition problem.

3.2.2. Profit Maximization. The profit maximization problem is more challenging. Because the offline segment of customers is impacted by particular product configurations, as opposed to only the attributes, the problem cannot be solved in the attribute space. In fact, even the $\alpha = 1$ case is in general hard to solve. When $\alpha = 1$, the problem reduces to finding the profit maximizing assortment for a single channel under the MNL model. It is known that the single-channel profit maximizing assortment is one of the profit-ordered assortments (Talluri and Van Ryzin 2004, Rusmevichientong et al. 2010), but because the universe of products is exponentially (in K) large, finding the best subset is computationally challenging, except when the optimal subset has polynomial (in K) size (Gallego et al. 2016).

A key source of hardness with (General Showcase Profit Max) is that we need to search over subsets $M \subseteq \mathscr{X}$ of products as opposed to subsets of attribute levels; indeed, for a given subset $\mathbf{S}$ of attribute levels, we obtain different profits from different sets M that achieve $\mathbf{S}$. Therefore, to make the problem tractable, we restrict the search to the collection of assortments $\{S_1 \times \cdots \times S_K : S_k \subseteq \mathscr{L}_k \text{ for } 1 \le k \le K\}$. This restriction[5] can be justified by noting that customers tend to extrapolate attribute combinations based on what they have been exposed to. For instance, if an offline-type customer is exposed to a large red bag and a small blue bag, then she may infer the availability of a large blue bag and choose from the subset $S_1 \times \cdots \times S_k$ when the attribute levels $(S_1, \ldots, S_K)$ are offered in the store, even though not all products in $S_1 \times \cdots \times S_k$ are offered.

With the above restriction, we arrive at an optimization problem that can be solved in the attribute space and admits an FPTAS. In particular, the profit function can now be simplified as follows:

Lemma 3.3 (Restricted Showcase Profit).

$$R^{\text{general}}(\mathbf{S}) = \sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_{k=1}^{K} R_k^{\mathsf{c}}(S_k)}{1 + 1/D^{\mathsf{c}}(\mathbf{S})}, \quad \textit{where}$$

$$R_k^{\mathsf{c}}(S_k) = \frac{b_k^{\mathsf{c}} + \sum_{l \in S_k} r_{kl} \delta_{kl}^{\mathsf{c}}}{D_k^{\mathsf{c}}(S_k)}, \; D_k^{\mathsf{c}}(S_k) = d_k^{\mathsf{c}} + \sum_{l \in S_k} \delta_{kl}^{\mathsf{c}}, \; \textit{and}$$

$$D^{\mathsf{c}}(\mathbf{S}) = \prod_{k=1}^{K} D_k^{\mathsf{c}}(S_k), \tag{9}$$

with $b_k^{\text{on}} := \sum_{l \in \mathscr{L}_k} r_{kl} e^{w_{kl}^{\text{on}}}$, $d_k^{\text{on}} := \sum_{l \in \mathscr{L}_k} e^{w_{kl}^{\text{on}}}$, *and* $\delta_{kl}^{\text{on}} := e^{w_{kl}^{\text{off}}} - e^{w_{kl}^{\text{on}}}$ *for all* k. *Furthermore,* $b_k^{\text{off}}, d_k^{\text{off}} = 0$ *and* $\delta_{kl}^{\text{off}} := e^{w_{kl}^{\text{off}}}$ *for all* k.

The lemma is proved in Online Appendix A, Section A.2. The decision problem now reduces to

$$\max_{\mathbf{S} \in 2^{\mathscr{L}_1} \times \cdots \times 2^{\mathscr{L}_K}} \sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_k R_k^{\mathsf{c}}(S_k)}{1 + \exp(-\sum_{k=1}^{K} \log D_k^{\mathsf{c}}(S_k))}. \tag{Restricted Showcase Profit Max}$$

Because sales maximization is a special case of (Restricted Showcase Profit Max), it follows from Theorem 3.6 that the above problem is NP-hard. Therefore, we extend the ideas from the pure showcase setting to obtain an FPTAS. As above, we first consider the following maximization problem:

$$\max_{\mathbf{S} \in 2^{\mathscr{L}_1} \times \cdots \times 2^{\mathscr{L}_K}} \sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_k R_k^{\mathsf{c}}(S_k)}{1 + e^{-t_{\mathsf{c}}}}$$
$$\text{s.t.} \sum_k \log D_k^{\mathsf{c}}(S_k) \ge t_{\mathsf{c}}, \quad \mathsf{c} \in \{\text{on}, \text{off}\}.$$

We solve the above optimization problem by approximately satisfying the constraints through a DP formulation. For any $\varepsilon > 0$, define $j_{S,k}^{\mathsf{c}} := \lfloor \log D_k^{\mathsf{c}}(S)/(\varepsilon t_{\mathsf{c}}/K) \rfloor$ for $\mathsf{c} \in \{\text{on}, \text{off}\}$. Furthermore, let $\rho := \lfloor K/\varepsilon \rfloor - K$. Now define the DP value function:

$$V(k, \omega_{\text{on}}, \omega_{\text{off}}) = \max_{\mathbf{S} \in 2^{\mathscr{L}_1} \times \cdots \times 2^{\mathscr{L}_K}} \sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_{1 \le k' \le k} R_{k'}^{\mathsf{c}}(S_k)}{1 + e^{-t_{\mathsf{c}}}}$$
$$\text{s.t.} \sum_{1 \le k' \le k} j_{S_k,k}^{\mathsf{c}} \ge \omega_{\mathsf{c}}, \quad \mathsf{c} \in \{\text{on}, \text{off}\}.$$

Our goal is to compute $V(K, \rho, \rho)$, for which we use the following DP recursion:

$$V(k, \omega_{\text{on}}, \omega_{\text{off}}) = \begin{cases} 0 & \text{if } k = 0, \omega_{\text{on}}, \omega_{\text{off}} \le 0, \\ -\infty & \text{if } k = 0, \omega_{\text{on}} > 0 \text{ or } \omega_{\text{off}} > 0, \\ \max_{S \in 2^{\mathscr{L}_k}} \left[\sum_{\mathsf{c}} \nu_{\mathsf{c},k}(S) + V(k-1, \omega_{\text{on}} - j_{S,k}^{\text{on}}, \omega_{\text{off}} - j_{S,k}^{\text{off}}) \right] & \text{otherwise,} \end{cases}$$

where we used $\nu_{\mathsf{c},k}(S) := \alpha^{\mathsf{c}} R_k^{\mathsf{c}}(S)/(1 + e^{-t_{\mathsf{c}}})$ for compactness of notation. We carry out the above DP for integers k and ω_{c} such that $0 \le k \le K$ and $0 \le \omega_{\mathsf{c}} \le \rho$ for $\mathsf{c} \in \{\text{on}, \text{off}\}$. Each iteration requires a search over $O(|2^{\mathscr{L}_k}|) = O(2^{|\mathscr{L}_k|})$ elements. Therefore, the running time of the DP is $O(K\rho^2 \max_k 2^{|\mathscr{L}_k|}) = O(K^3 \max_k 2^{|\mathscr{L}_k|}/\varepsilon^2)$ because $\rho = O(K/\varepsilon)$. We first show that the DP above solves the desired optimization problem approximately.

Lemma 3.4 (Restricted Showcase DP Approximation). *For given* t_{on}, t_{off}, *and* q, *if there exists an* $\mathbf{S}$ *such that*

$$\sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_k R_k^{\mathsf{c}}(S_k)}{1 + e^{-t_{\mathsf{c}}}} \ge q \quad \textit{and}$$

$$\sum_{k=1}^{K} \log D_k^{\mathsf{c}}(S_k) \ge t_{\mathsf{c}}, \quad \mathsf{c} \in \{\text{on}, \text{off}\},$$

then the DP described above terminates with solution $\hat{\mathbf{S}}$ *such that*

$$\sum_{\mathsf{c} \in \{\text{on}, \text{off}\}} \alpha^{\mathsf{c}} \frac{\sum_k R_k^{\mathsf{c}}(\hat{S}_k)}{1 + e^{-t_{\mathsf{c}}}} \ge q \quad \textit{and}$$

$$\sum_{k=1}^{K} \log D_k^{\mathsf{c}}(\hat{S}_k) \ge (1 - 2\varepsilon) t_{\mathsf{c}}, \quad \mathsf{c} \in \{\text{on}, \text{off}\}.$$

The lemma is proved in Online Appendix A, Section A.2. The precise algorithm is summarized below.

Algorithm 2 (FPTAS for (Restricted Showcase Profit Max))

Input. Tolerance parameter $\varepsilon > 0$ and problem inputs $R_k^{\mathsf{c}}(S)$ and $D_k^{\mathsf{c}}(S)$ for all $S \in 2^{\mathscr{L}_k}$, $1 \le k \le K$, and $\mathsf{c} \in \{\text{on}, \text{off}\}$.

Step 1. Define $\tau_{\text{on,min}} := e^{-\sum_k \log D_k^{\text{on}}(\mathscr{L}_k^+)}$ and $\tau_{\text{on,max}} := e^{-\sum_k \log D_k^{\text{on}}(\mathscr{L}_k^-)}$. Similarly, $\tau_{\text{off,min}} := e^{-\sum_k \log D_k^{\text{off}}(\mathscr{L}_k)}$ and $\tau_{\text{on,max}} := e^{-\sum_k \min_{l \in \mathscr{L}_k} \log \delta_{kl}^{\text{off}}}$. Create the ε grid $\mathscr{T}_{\mathsf{c}}$ of the

interval $[\tau_{c,\min}, \tau_{c,\max}]$ such that $\mathcal{T}_c := \{\tau_{c,\min}(1+\varepsilon)^j : j = 0, 1, \ldots, J_c\}$ with $J_c = \log(\tau_{c,\max}/\tau_{c,\min})/\log(1+\varepsilon)$, for $c \in \{\text{on}, \text{off}\}$.
Step 2. For $\boldsymbol{\tau} = (\tau_{\text{on}}, \tau_{\text{off}}) \in \mathcal{T}_{\text{on}} \times \mathcal{T}_{\text{off}}$ do
define $t_c = -\log \tau_c$, $c \in \{\text{on}, \text{off}\}$:
determine $\hat{\mathbf{S}}_\tau$ by solving the DP with inputs t_{on}, t_{off}, and $\varepsilon/|\log \tau^*|$, where $\tau^* = \min\{\tau_{\text{on},\min}, \tau_{\text{off},\min}\}$.
Output. Solution $\hat{\mathbf{S}}$ that maximizes the expected profit from the collection $\{\hat{\mathbf{S}}_\tau : \boldsymbol{\tau} \in \mathcal{T}_{\text{on}} \times \mathcal{T}_{\text{off}}\}$.

We now show that the algorithm above is indeed an FPTAS.

Theorem 3.7 (FPTAS for (Restricted Showcase Profit Max)). *Algorithm 2 produces a* $1 - 8\varepsilon$ *optimal solution with a running time of* $O(K^3 \max_k 2^{|\mathcal{L}_k|} \log^2 \tau^* \cdot \log(\tau_{\text{on},\max}/\tau_{\text{on},\min}) \log(\tau_{\text{off},\max}/\tau_{\text{off},\min})/(\varepsilon^2 \log^2(1+\varepsilon)))$, *where* $\tau_{\text{on},\min} := e^{-\sum_k \log D_k^{\text{on}}(\mathcal{L}_k^+)}$, $\tau_{\text{on},\max} := e^{-\sum_k \log D_k^{\text{on}}(\mathcal{L}_k^-)}$, $\tau_{\text{off},\min} := e^{-\sum_k \log D_k^{\text{off}}(\mathcal{L}_k)}$, $\tau_{\text{on},\max} := e^{-\sum_k \min_{l \in \mathcal{L}_k} \log \delta_{kl}^{\text{off}}}$, *and* $\tau^* = \min\{\tau_{\text{on},\min}, \tau_{\text{off},\min}\}$.

The theorem is proved in Online Appendix A, Section A.2.

3.3. Integer-Programming-Based Heuristic

Building on the ideas in the construction of the FPTAS, we now propose an IP-based heuristic to approximately find the profit maximizing assortment of products to offer in the store. Existing work has used IP formulations to solve assortment problems (Bront et al. 2009, Subramanian and Sherali 2010) in the product space when the objective function can be expressed as a ratio of linear functions in decision variables. Instead, we solve the problem in the attribute space, and the objective function is a ratio of a linear to a nonlinear functions in the decision variables; hence, we use the ideas from our FPTAS to convert the nonlinear integer program (NLIP) into a collection of mixed integer linear programs (MILPs). We demonstrate the performance of the heuristic on synthetic data in Section 4 and on real-world data in Section 5.

Using the simplification of the profit function from Lemma 3.3, our objective is to solve the following decision problem:

$$\max_{\mathbf{S} \in 2^{\mathcal{L}_1} \times \cdots \times 2^{\mathcal{L}_K}} \sum_{c \in \{\text{on}, \text{off}\}} \alpha^c \frac{\sum_{k=1}^K R_k^c(S_k)}{1 + \exp(-\log D_k^c(S_k))}. \tag{10}$$

We first formulate this optimization problem as an NLIP. To do so, we encode subset $\mathbf{S}$ using binary vectors $\mathbf{z}_k$, for $1 \le k \le K$, defined as $z_{k,S} = 1$ if and only if $S = S_k$ and $z_{k,S} = 0$ otherwise. In other words, the binary vector $\mathbf{z}_k$ is of length $2^{|\mathcal{L}_k|} - 1$, with each component associated with a nonempty subset of $\mathcal{L}_k$. It is clear from the definition that the encoding is a bijection from $\{0,1\}^{2^{|\mathcal{L}_k|}-1}$ to $2^{\mathcal{L}_k}$. With this binary encoding of $\mathbf{S}$, we can formulate the optimization problem in (10) as the following NLIP:

$$\begin{aligned} \max_{\mathbf{z}_k, 1 \le k \le K} \quad & \sum_{c \in \{\text{on}, \text{off}\}} \alpha^c \frac{\sum_{k=1}^K \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} R_k^c(S)}{1 + \exp(-\sum_{k=1}^K \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} \log D_k^c(S))} \\ \text{subject to} \quad & \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} = 1, \quad k = 1, 2, \ldots, K, \\ & z_{k,S} \in \{0,1\}, \quad k = 1, 2, \ldots, K;\ S \in 2^{\mathcal{L}_k}. \end{aligned}$$

Because the above formulation has a nonlinear objective function, we solve it by reducing it to a collection of MILPs at various grid points. To obtain the reduction, we use the ideas from the FPTAS described above. Let $\tau_{c,\min}$ and $\tau_{c,\max}$ denote the minimum and maximum values of $\exp(-\sum_{k=1}^K \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} \log D_k^c(S))$, respectively, as $\mathbf{z}_k$ varies over all the binary vectors in $\{0,1\}^{2^{|\mathcal{L}_k|}-1}$ such that $\sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} = 1$ for all $1 \le k \le K$ and $c \in \{\text{on}, \text{off}\}$. Then, for a given $\varepsilon > 0$, we consider the ε-grid $\mathcal{T}_{\text{on}} \times \mathcal{T}_{\text{off}}$, where $\mathcal{T}_c = \{\tau_{c,\min}(1+\varepsilon)^j : j = 0, 1, \ldots, J_c\}$ with J_c chosen such that $\tau_{c,\min}(1+\varepsilon)^{J_c - 1} \le \tau_{c,\max} \le \tau_{c,\min}(1+\varepsilon)^{J_c}$. For each grid point $\boldsymbol{\tau} = (\tau_{\text{on}}, \tau_{\text{off}}) \in \mathcal{T}_{\text{on}} \times \mathcal{T}_{\text{off}}$, we solve the following MILP:

$$\begin{aligned} \max_{\mathbf{z}_k, 1 \le k \le K} \quad & \sum_{c \in \{\text{on}, \text{off}\}} \alpha^c / 1 + \tau_c \sum_{k=1}^K \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} R_k^c(S) \\ \text{subject to} \quad & \sum_{k=1}^K \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} \log D_k^c(S) \ge -\log \tau_c, \quad c \in \{\text{on}, \text{off}\} \\ & \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} = 1, \quad k = 1, 2, \ldots, K, \\ & z_{k,S} \in \{0,1\}, \quad k = 1, 2, \ldots, K;\ S \in 2^{\mathcal{L}_k}. \end{aligned}$$

Let $\mathbf{z}_{k,\tau}$, $1 \le k \le K$, denote the optimal solution obtained from solving the above MILP for grid point $\boldsymbol{\tau}$. Let $S_{\tau,k}$ denote the subset of attribute levels such that $z_{k,\tau,S} = 1$ for $S = S_{\tau,k}$, and let $\mathbf{S}_\tau$ denote $(S_{\tau,1}, \ldots, S_{\tau,K})$. We then output the subset $\mathbf{S}_{\tau^*}$ from the collection $\{\mathbf{S}_\tau : \boldsymbol{\tau} \in \mathcal{T}_{\text{on}} \times \mathcal{T}_{\text{off}}\}$ that maximizes the expected profit $R^{\text{general}}(\mathbf{S}_\tau)$, computed from the expression in (9). Instead of the value of ε, we may also specify the number of grid points J_c, in which case we can back out the value of ε as equal to $(\tau_{c,\max}/\tau_{c,\min})^{1/J_c} - 1$.

Note that for each grid point $\boldsymbol{\tau}$, we can find the solution $\mathbf{S}_\tau$ by solving the DP in Section 3.2 instead of solving the MILP above. Solving the MILP may be faster because of any structures that the commercial IP solvers exploit.

4. Numerical Study

We carried out two computational experiments using synthetic data. The first study is designed to assess the optimality gap of the IP-based heuristic on smaller problem instances. On problem instances with four attributes and two levels per attribute, the study shows that the IP-based heuristic obtains profits within 6.45%

to 0.11% of the optimal profit, on average, as the number of grid points is increased from $J_{on} = J_{off} = 2$ to $J_{on} = J_{off} = 32$. In all of these instances, the average running time was < 0.5 seconds. The second study is aimed at demonstrating the practical effectiveness of our IP-based heuristic. The study achieves two objectives: it demonstrates that (a) the computational time of the IP-based heuristic scales to large, practical-sized problems, and (b) the solutions obtained from the IP-based heuristic provide significantly higher profits and sales when compared to the solutions from standard revenue-ordered and greedy heuristics. Our results demonstrate that, on average, the IP-based heuristic runs in less than three minutes for problem instances with 100 attributes and 10 levels per attribute and provides 32% more profit than the best single-channel solution that ignores the presence of the online channel and 28% more profit than the best standard heuristic.

The broad simulation setup we used is as follows: (a) generate a random instance of the ground-truth model class; (b) determine the approximate profit/sales maximizing offline assortment using the IP-based heuristic, the RO heuristic, the greedy heuristic, and the single-channel heuristic that ignores the impact of the online channel; (c) compare the "true" profits/sales, as computed using the ground-truth model, from the different solutions. We repeated the above sequence of steps for a large number of instances and various parameter combinations to cover the spectrum. For smaller problem instances, we also determined the optimal profit through exhaustive search and determined the optimality gaps.

4.1. Ground-Truth Models Generated

We considered problem instances of difference sizes by varying the number of attributes K and the number of levels L per attribute. For each combination (K, L) of parameters, we randomly generated 100 model instances, with each instance generated as follows: (a) For each attribute-level combination (k, l), such that $1 \le k \le K$ and $1 \le l \le L$, sample w_{kl}^{off} uniformly at random from $[-4, 1]$. (b) Given w_{kl}^{off}, set $w_{kl}^{on} = w_{kl}^{off}$ with probability $1 - \rho = 0.4$ and sample w_{kl}^{on} uniformly at random from $[-4, w_{kl}^{off})$ with probability $\rho/2$ and $(w_{kl}^{off}, 1]$ with the remaining probability $\rho/2$. (c) Sample the profit r_{kl} for the attribute-level combination (k, l) uniformly at random from the interval $[100/K, 150/K]$, ensuring that product profits fall in the interval $[100, 150]$. (d) Sample the size of the offline segment α^{off} uniformly at random from $[\alpha_{min}, \alpha_{max}]$ and set the size of the online segment $\alpha^{on} = 1 - \alpha^{off}$. We set $\alpha_{min} = \alpha_{max} = 0.2$ for the first study and $[\alpha_{min}, \alpha_{max}] = [0, 0.5]$ for the second one.

The above generative model is designed to reflect the situation when the online and offline partworths are the same for some attribute levels; for instance, the partworths may be the same for attributes the customers are familiar with, such as the color blue (see Table 3). The parameter ρ captures the fraction of attribute levels for which the online and offline partworths differ. By construction, fractions of about $\rho/2$ attribute levels are overvalued and undervalued, respectively.

Table 1. Average Optimality Gaps and Running Times (in Seconds) of the IP-Based Heuristic as a Function of the Number of Grid Points; Our IP-Based Heuristic Obtains Near-Optimal Solutions Quickly

No. of grid points	Optimality gap	Run time (s)
2	6.45	0.01
4	2.97	0.01
8	0.92	0.04
16	0.19	0.12
32	0.11	0.43

4.2. Optimality Gaps for Smaller Problem Instances

Table 1 reports the optimality gaps of the IP-based heuristic proposed in Section 3.3 above when $K = 4$ and $L = 2$. We varied the number of grid points from $J_{on} = J_{off} = 2$ to $J_{on} = J_{off} = 32$ in powers of two, which ensured that the grids were nested. We fixed $\alpha^{off} = 20\%$, reflecting that 80% of the customers who visit the offline store also visit the online store. For each of the 100 model instances, we computed the optimal solution through exhaustive search and then the optimality gaps of the IP-based heuristic by varying the number of grid points. Table 1 reports the optimality gaps, averaged over the 100 random instances.

As expected, we observe that the optimality gap shrinks but the running time increases, on average, as the number of grid points increases. For the instances in our study, we observe that the gap decreases to within < 0.2% of the optimal revenue with 16 grid points. The IP-based heuristic runs in < 0.5 seconds,[6] even when the number of grid points is 32.

4.3. Scaling of the IP-Based Heuristic to Larger Problem Instances

For the second study, we considered eight larger problem instances by varying K over $\{10, 20, 50, 100\}$ and L over $\{5, 10\}$. Because the universe of the products is exponentially large (for $K = 100$ and $L = 10$, the universe consists of 10^{100} products), we focused on the problem of finding the profit maximizing assortment of size at most $C = 50$. Exhaustive search is no longer computationally feasible, so we assessed the performance of the IP-based heuristic by comparing its profit to that obtained by three benchmark methods: (a) the *offline heuristic*, (b) the *revenue-ordered* heuristic, and (c) the *greedy* heuristic, described next.

4.3.1. Benchmark Methods. The offline heuristic ignores the impact of the online channel and performs single-channel assortment optimization. Under the MNL model, this problem can be solved in $O(NC)$ time, where C is the maximum subset size, and N is the number of products in the universe. Because N is exponentially large, we implemented the following heuristic: find the most profitable subset from among the subsets of the form $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ for $1 \le m \le C$, where $\mathbf{x}_1, \ldots, \mathbf{x}_C$ are the C most profitable products such that $p_{\mathbf{x}_1} \ge p_{\mathbf{x}_2} \ge \cdots \ge p_{\mathbf{x}_C}$. This heuristic returns the optimal solution when the profit maximizing subset without the capacity constraint has size at most C because it is known that the unconstrained profit maximizing subset comprises the m most profitable products, for some m. To find the C most profitable products, we used the recent algorithm proposed by Gallego et al. (2016); details are provided in Online Appendix B.

The RO heuristic finds the profit maximizing subset from among the N subsets, each comprising the m most profitable products for m ranging from 1 to N. Because N is exponentially large, we search only over $m = 1, 2, \ldots, C$. The key difference from the offline heuristic is that while the offline heuristic picks the subset M that maximizes $R^{\text{off}}(M)$, the RO heuristic picks the subset M that maximizes the profit $R(M)$ from *both* online and offline channels.

The greedy heuristic is another general-purpose heuristic commonly applied to assortment optimization problems (Jagabathula 2014). While the existing heuristics typically operate in the product space, we implemented a natural variant that operates in the attribute space. In each iteration, we add the feature that results in the maximum increase in the profit. We stop if the capacity is reached or the profit no longer increases; details are provided in Online Appendix B.

4.3.2. IP-Based Heuristic. For the IP-based heuristic, we chose the number of grid points to be $J_{\text{on}} = J_{\text{off}} = 5$, so that we solve a total of $J_{\text{on}} \times J_{\text{off}} = 25$ IPs for each instance. We enforced the cardinality constraint by adding the linear constraint $\sum_{k=1}^{K} \sum_{S \in 2^{\mathcal{L}_k}} z_{k,S} \log|S| \le \log C$ to the IP described above.

4.3.3. Results and Discussion. The results from our simulation study are presented in Table 2. The table reports two metrics: (a) the profits extracted by the RO, greedy, and IP-based heuristics relative to the profit extracted by the offline heuristic, and (b) the computational times, in seconds, of each of the heuristics. Each row corresponds to one of the eight model types, represented by the tuple (K, L). The "Relative performance" columns report the average profit from each heuristic, relative to that from the offline heuristic, with the average computed over the 100 random instances for each model type: $(1/100)\sum_{t=1}^{100} R(M_t^{\text{method}})/R(M_t^{\text{offline}})$, where method = RO, greedy, IP and M_t^{method} and M_t^{offline} denote the solutions found by the particular method and the offline heuristics, respectively, for problem instance t. Higher values are better, and values above 1 indicate the improvements in profits from accounting for the presence of the online channel. The last three columns report the average computational times, in seconds, averaged over the 100 random instances for each model type. We draw the following key conclusions:

1. *IP-based heuristic scales well*. The computational times of the IP-based heuristic scale well to large, practical-sized problem instances. Even when we stress test our method by applying it to large instances with 100 attributes and 10 levels in each attribute (making the product universe consist of 10^{100} products), the IP-based heuristic provides good quality solutions within three minutes, on average.

2. *IP-based heuristic extracts the most profit*. The IP-based heuristic vastly outperforms all the other heuristics: 32% and 28% more profit on average extracted than the RO and greedy heuristics, respectively. This shows that because of the problem structure, relying on general-purpose heuristics can leave a lot of money on the table.

The above results establish the value of the IP heuristic: it scales to large problem sizes and extracts higher profits than existing benchmarks. Finally, we note that the IP-based heuristic can also be used to solve the single-channel assortment problem by setting $\alpha^{\text{on}} = 0$. Recent work (Gallego et al. 2016) has reduced the single-channel assortment problem (in the feature space) to the K-shortest path problem in a directed acyclic graph, which can be solved efficiently using Yen's (1971) algorithm. But existing techniques do not extend to the setting with cardinality constraints. The IP-based heuristic, on the other hand, can readily accommodate (linear) constraints.

Table 2. The Performance of the IP-Based Heuristic on Large Problem Istances Rlative to Standard Benchmarks

	Relative performance			Computation times		
(K, L)	RO	Greedy	IP	RO	Greedy	IP
(10, 5)	1.00	1.07	1.22	0.07	0.07	0.30
(20, 5)	1.00	1.01	1.25	0.21	0.24	0.51
(50, 5)	1.00	1.01	1.28	1.17	1.45	1.05
(100, 5)	1.00	1.01	1.27	4.58	5.65	1.93
(10, 10)	1.00	1.11	1.41	0.09	0.18	7.79
(20, 10)	1.00	1.01	1.45	0.30	0.66	15.93
(50, 10)	1.00	1.01	1.42	1.74	3.82	54.47
(100, 10)	1.00	1.00	1.27	6.79	14.66	131.50

Notes. The "Relative performance" columns report the average ratio of the profit from each method to that from the offline heuristic, averaged over 100 problem instances. The "Computation times" columns report the average computation times, in seconds. The IP heuristic scales to large problem instances and extracts significantly larger profit than standard heuristics.

5. Timbuk2 Case Study

This section describes a case study we conducted to illustrate how our techniques apply to a real-world application and quantify the value of our methodologies. Particularly, the study demonstrates that the utility partworths of the *same* individual can change significantly after physical evaluation. The case study focuses on messenger bags from Timbuk2, a San Francisco-based company that sells customized messenger bags through its online store and also showcases some of the bags in self-owned or third-party (such as Recreational Equipment Inc., or REI) brick-and-mortar retail stores. The key findings are that (a) the differences between online and offline partworths are statistically significant for six of the nine included product attributes, with the magnitude of some of these differences being "large"; (b) the gain in sales and revenue from accounting for channel interactions can be substantial (up to 40% in our case study); and (c) the single-channel optimal assortment, which ignores channel interactions, is substantially different from the optimal assortment that accounts for channel interactions.

For our analysis, we conducted a conjoint study to collect preference data on messenger bags. Using the collected data, we estimated participants' online and offline partworths to validate our modeling assumption that online and offline partworths differ. Then, using the estimated partworths, we computed the sales and revenue maximizing subsets[7] using the offline heuristic (which ignores channel interactions) and the IP heuristic. By comparing the sales/revenues from the resulting assortments, we show that the gains from accounting for the channel interactions are substantial.

5.1. Details of the Conjoint Study

Conjoint analysis is widely used by practitioners for quantitative preference measurement. In a typical conjoint study, participants are shown a set of products and asked to provide evaluations by either rating, ranking, or choosing products. These evaluations are then used to back out individual-level attribute partworths by fitting utility or choice models to the responses. The measured preferences are used by firms for demand predictions, product design decisions (Kohli and Krishnamurti 1989), and assortment decisions in a single channel (Dobson and Kalish 1988, 1993).

Conjoint studies are typically conducted either online (in which participants evaluate descriptions of products on a computer) or offline (in which participants evaluate physical prototypes). However, because our goal is to measure the differences between online and offline partworths, we asked each participant to complete an online task, followed by an offline task.

5.1.1. Product and Attributes. We chose Timbuk2 messenger bags for our study for the following reasons: (a) they vary on several attributes, some of which are touch-and-feel attributes for which we expect differences between online and offline partworths; (b) they are in the right price range (expensive enough for participants to take the decision seriously but cheap enough for the participants to be interested in purchasing them); (c) they are infrequently purchased, so many participants may be unfamiliar with at least some of the attributes and lack well-formed preferences; (d) they are configurable through Timbuk2's website, allowing us to purchase bags to create a balanced orthogonal design, required to efficiently estimate the attribute partworths; and (e) they are physically small enough to simplify the logistics of carrying out the study in a behavioral lab. Figure 1 shows the included attributes of the bag.

5.1.2. Study Design. Based on the six attributes described above, there is a total of $4^2 \times 3 \times 2^3 = 384$ (four levels of exterior design, four levels of price, three levels of interior compartments, and two levels each of size, strap pad, and water bottle pocket) feasible feature combinations. We used the "D-optimal" design criterion (Kuhfeld et al. 1994) to select a subset of 20 bags from the above universe to be included in our study. Our design has a D-efficiency metric of 0.97, which is considered sufficiently high for reliable estimation. The configurations of the 20 bags that were included in the study are presented in Table EC.1 in Online Appendix C.

5.1.3. Participant Tasks. Each participant was asked to complete two ratings-based tasks in sequence: an *online task* followed by an *offline task*. In the online task, the participants were presented with 20 messenger bags, in sequence, on separate screens and asked to rate each bag on a five-point scale (definitely not buy, probably not buy, may or may not buy, probably buy, definitely buy). After completing the online task, they were taken to a separate room to complete the offline task. They were presented with the *same* set of 20 bags, physically laid out on a conference table, with a card next to each bag displaying a corresponding identifier and price. The experimenter walked them through all the features, showing each feature on a sample bag, and asked them to evaluate the bags and rate them on the same five-point scale. Online Appendix C presents additional details of the task.

We recruited 122 participants from a university subject pool for the study. To incentivize honest responses (Ding 2007), participants were told that they will be entered in a raffle and if they win, they will receive, for free, a bag configured according to their preferences (inferred from the responses they provided in the study) plus cash, for a total value of $180.[8]

5.2. Data Analysis: Parameter Estimation

From the study, we collected two data sets (online and offline), each set consisting of 20 ratings from the 122

Figure 1. List of Attributes and Image of Exterior Designs, Shown to Participants in the Online Task

- Exterior design: Black, Blue, Reflective, Colorful (illustrated on the right)
- Size: Small (10 × 19 × 14 in), Large (12 × 22 × 15 in)
- Price: \$120, \$140, \$160, \$180
- Strap pad: Yes, No
- Water bottle pocket: Yes, No
- Interior compartments: Empty bucket (no dividers), Divider for files, Crater laptop sleeve

participants. We fitted the following linear model separately to the two data sets to obtain the online and offline partworths:

$$y_{piz} = \gamma_z + \sum_j \beta_{zj} x_{ij} + \varepsilon_{pi},$$

where $z \in \{\text{online, offline}\}$, and y_{piz} is the rating provided by participant p for bag i online or offline.

Price is a continuous variable with one coefficient, and the remaining attributes are categorical variables, represented using dummy coding.[9] Table 3 presents the estimated online and offline partworths.

Let us first focus on the online partworths. All the estimated partworths were statistically significantly different from 0 at $p < 0.001$. As expected, participants had a negative price coefficient (−0.22). Participants also preferred black to the other exterior designs. For example, participants rated colorful exterior design 1.06 points lower on average than black. In the online study, participants also preferred large bags to small bags, having a water bottle pocket to not having one, and having a strap pad to not having one. The offline partworths have similar interpretations.

To test whether the partworths online differ from those offline, we fitted following model to the data pooled from the two studies:

$$y_{piz} = \gamma + \sum_j \beta_j x_{ij} + \delta z + \sum_j \delta_j z x_{ij} + \varepsilon_{piz}, \qquad (11)$$

where we abuse notation and let z denote a Boolean variable taking the value 0 for the data from the online study and 1 for the data from the offline study. The coefficients δ_j capture the difference between the offline and online partworths for feature j. We compared this model with a restricted (and nested) one obtained by restricting the coefficients $\delta = \delta_j = 0$, for all j:

$$y_{piz} = \gamma + \sum_j \beta_j x_{ij} + \varepsilon_{piz}. \qquad (12)$$

The F-test (analysis of variance test) rejected the null hypothesis that all the differences δ and δ_j, for all j, are zero at $p < 0.01$, indicating that the online and offline partworths differ statistically significantly.

The last column of Table 3 reports the coefficients δ_j, which capture the difference in partworths for feature j. We note that of the nine partworth estimates, six changed statistically significantly at $p < 0.01$ (all except Blue, Price, and Divider for files), and some of the coefficients changed by a large amount. In particular, the population preference for Colorful went up, Reflective went down, and size *reversed* (from Large to Small) after physical evaluation.

Finally, we carried out individual-level tests. We fitted the models in (11) and (12) to the data for each individual and compared them using a F-test. We observed that the models differed statistically significantly for 29.5% (64 out of 122) of them at $p < 0.01$ and 51.6% (63 out of 122) at $p < 0.05$. For completeness, we also fitted a mixed model with a random intercept for each participant and found the partworths to be essentially the same; the likelihood-ratio tests comparing the models in (11) and (12) but with random intercepts corresponding to participants also resulted in the same conclusion, but with a slightly different p-value.

We also tested whether the offline partworths persist when the consumers go back online, using a smaller study in which we asked a group of 20 other participants to do the tasks in reverse order: first the offline task, followed by the online task. For this group, we found that an F-test comparing models in (11) and (12) could not reject the null hypothesis that the online and offline partworths differ at $p < 0.01$; see Table EC.2 in Online Appendix C for the estimated coefficients. Furthermore, the individual-level tests revealed that the models in (11) and (12) were *not* statistically significantly different for *all* of the 20 individuals at $p < 0.01$ and $p < 0.05$. The results from this second group of participants provide evidence that the attribute partworth used for the purchase decision depends only on *whether*

Table 3. Differences Between the Online and Offline Partworths

Attribute	Level	Online (w^{on})	Offline (w^{off})	Difference
Exterior design	Reflective	−0.31**	−0.60**	−0.28*
	Colorful	−1.06**	−0.71**	+0.36**
	Blue	−0.22**	−0.11	+0.11
	Black			
Size	Large	0.27**	−0.31**	−0.58**
	Small			
Price	\$120, \$140, \$160, \$180	−0.011**	−0.008**	+0.004
Strap pad	Yes	0.51**	0.25**	−0.26**
	No			
Water bottle pocket	Yes	0.45**	0.17**	−0.28**
	No			
Interior compartments	Divider for files	0.41**	0.52**	+0.11
	Crater laptop sleeve	0.62**	0.88**	+0.26*
	Empty bucket/no dividers			
Intercept		3.72**	3.39**	−0.33

Notes. Results are based on the 122 participants who completed the online task first, followed by the offline task. The levels with no coefficients were set to zero in dummy encoding. The differences are statistically significant for several attribute-level combinations.
*$p < 0.01$; **$p < 0.001$.

the customer has been exposed to the attribute in a physical product, rather than on the channel in which the purchase decision is made. Once the customer has been exposed to the attribute level, he will apply the new partworths to both his online and offline purchasing decisions. Fitting a mixed model as described above did not change our conclusions.

We draw the following conclusions from the conjoint study. Consumers use different partworths when evaluating products online and offline and once a product is examined offline, there is evidence to suggest that consumers apply the offline partworths to both online and offline product evaluation. The relevant parameters can be estimated using well-established market research tools, making our method readily applicable in practice.

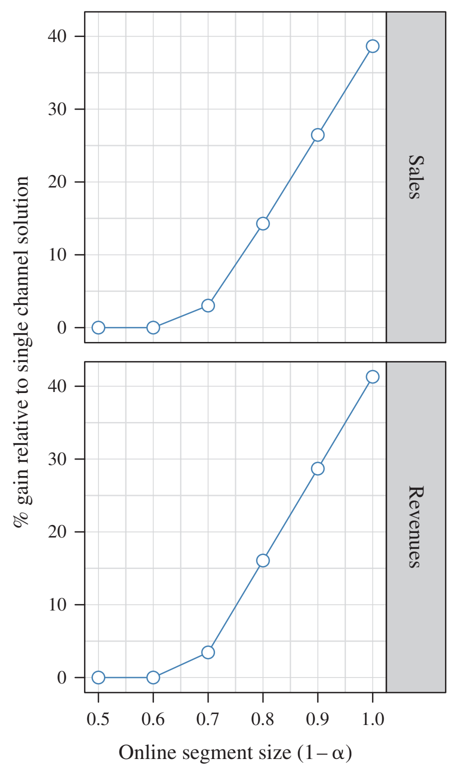
5.3. Assortment Optimization: Impact on Sales and Revenues

Using the parameters obtained from the conjoint study, we demonstrate how the firm's sales and revenues are affected if the offline assortment is selected without taking into account the online channel. We used the IP formulation obtained in Section 3.3 to optimize the assortment with and without the online channel for various sizes of the offline segment of consumers. Our results demonstrate that the gain in revenues from accounting for the online channel can be significant if a large portion of the population visits both channels.

Because we are assuming that customer choices are described by the MNL model, we verified its fit by carrying out a fivefold cross-validation on both the online and offline conjoint data. We measured the out-of-sample error in terms of the standard mean absolute percentage error metric, which measures the average relative error in predicting market shares. We found that the MNL model has about a 3.7% error rate, indicating that it is a good fit. The details of the verification are in Online Appendix C, Section C.2.

5.3.1. Setup. To compute the optimal assortments, we used the following parameter values. The utility parameters $\tilde{w}_{kl}^{on}$ and $\tilde{w}_{kl}^{off}$, for all nonprice attribute levels k, l (presented in Table 3), were obtained from the conjoint study. We obtained the partworth revenues from the prices that Timbuk2 posted on their website for their customizable messenger bags. The price for the base configuration is \$140, which corresponds to the product configuration Black, Small, No strap pad, No water bottle pocket, and Empty bucket. Because of the dummy coding, the utility of the base configuration is equal to the intercept + $\beta_{\text{price}} \cdot \$140$. Here we use the offline intercept (because the intercepts do not differ significantly), and $\beta_{\text{price}} = -0.008$ is the offline price coefficient. The partworth revenues $\tilde{\pi}_{kl}$ for the non-dummy attribute levels were obtained from the additional prices over the base that Timbuk2 charges: \$10 for Reflective or Colorful, \$10 for Large, \$15 for Strap pad, \$5 for Water bottle pocket, \$10 for Laptop compartment, and \$0 for Divider for files. For every non-dummy attribute level k, l, we set $w_{kl}^{c} = \tilde{w}_{kl}^{c} + \beta_{\text{price}} \cdot \tilde{\pi}_{k,l}$, for $c \in \{\text{on}, \text{off}\}$. Because only differences in mean utilities matter, we absorbed the utility of the base configuration into the utility of the no-purchase option, which was set to achieve reasonable market shares of about 40% in the offline channel (range between 41% and 56% in both channels depending on the size of the offline customer segment). With these parameter values, we

Figure 2. (Color Online) Gains in Revenues and Sales from Accounting for the Online Channel



	Online segment size $(1-\alpha)$				
	Sales max.			Revenue max.	
Attribute level	0.5 – 0.6	0.7	≥ 0.8	0.5 – 0.6	≥ 0.7
Exterior design					
Reflective*	×	×		×	
Colorful	×	×	×	×	×
Blue	×	×	×	×	×
Black	×	×	×	×	×
Size					
Large*	×			×	
Small	×	×	×	×	×
Strap pad					
Yes*	×	×		×	
No	×	×	×	×	×
Water bottle pocket					
Yes*	×	×		×	
No	×	×	×	×	×
Interior compartments					
Divider for files	×	×	×	×	×
Crater laptop sleeve	×	×	×	×	×
Empty bucket/no dividers	×	×	×	×	×

Notes. The table marks by × the attribute levels that are present in the sales and revenue maximizing subsets. The features marked with asterisks are overvalued.

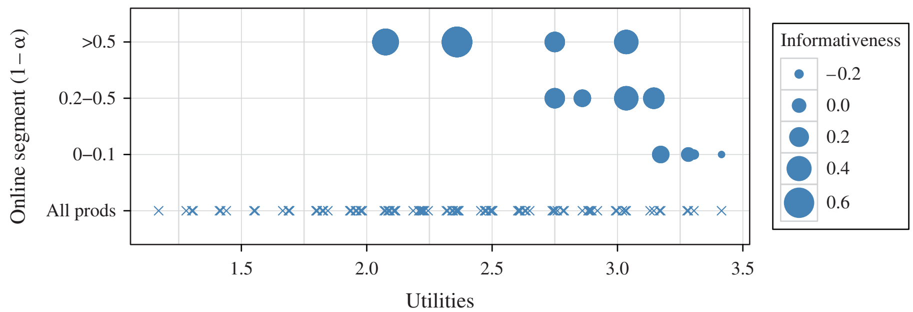
carried out sales and revenue maximization and compared the optimal solutions with the single-channel benchmark.

5.3.2. Results. Using the IP-based heuristic (with the number of grid points $J_{on} = J_{off} = 10$) described in Section 3.3, we computed both sales and revenue maximizing assortments by varying the online segment sizes from 0.5 to 1. We compared the optimal sales and revenues against the benchmark sales and revenues, respectively, obtained from maximizing the offline channel (ignoring channel interactions). The benchmark solution was obtained by applying our IP-based heuristic with $\alpha = 1$.

Figure 2 reports the gains from accounting for the presence of the online channel in the firm sales and revenues. The upward trend indicates that as the number of customers visiting both channels increases, it becomes more important to account for the online channel in determining the offline assortment. When all customers visit both channels, the gain in both sales and revenue is about 40%. As the number of these customers decreases and more customers purchase only from the offline store, accounting for the online channel becomes less relevant.

Figure 2 also shows the attribute levels contained in the sales and revenue maximizing assortments for different values of α. Focusing on the sale maximizing subsets, we note that when the offline segment is large, all features are included. This is because this regime is dominated by the offline segment, for which it is indeed optimal to include all feasible products. As the online segment grows, the assortment changes to exclude overvalued features: Reflective, Large, Strap pad, and Water bottle pocket. This result is consistent with the result of Theorem 3.1. Excluding overvalued attributes benefits the sales from the online segment but hurts sales from the offline segment. This tension is strongest for over-valued attributes that have positive offline partworths (in our case, Strap pad and Water bottle pocket) because excluding them decreases the offline sales most significantly. As a result, the decision whether to include these attributes is particularly sensitive to the value of α. The interpretation of the results for the revenue maximizing subsets is similar. When the offline segment is large, all features are included, which is optimal for the offline channel because the firm's market share is "small enough." When the online segment becomes larger, the firm will exclude overvalued features because they bring in the

Figure 3. (Color Online) Utilities of Products Included in the Optimal Assortments of Size at Most Six at Different Sizes of the Online Segment



Notes. The last row plots the utilities of all the 96 products in the universe. Products represented by larger-sized points are more "informative," measured as the difference between the offline and online partworths. The optimal assortments contain a mix of "popular" (high utility) and "informative" products.

most revenues and hiding them increases the chance of their sale.

To gain insights into the products included in the sales maximizing offline assortment, we constrained its cardinality to be 4, which allowed us to examine the individual products included. Figure 3 shows the sales maximizing subset of products at different values of the online segment size. Each product in the optimal subset is represented by a point whose horizontal axis value denotes the product's "popularity" and the radius denotes the product's "informativeness." We use the offline utility of the product as a measure of its popularity and the difference between its offline and online utilities as a measure of its informativeness. The broad insight we obtain from our results is that the optimal assortment is a mix of popular and informative products. The popular products have high utilities and generate sales in the offline channel, whereas the informative products expose customers to undervalued attributes and generate sales in the online channel. When the offline segment dominates, i.e., the offline channel generates most of the firm's sales, it is optimal to fill the capacity with popular products; we see this on the graph in the second to last row, where the assortment consists of products with high utilities. On the other hand, when the online segment dominates, it is optimal to offer informative products at the expense of popular ones; we see this on the graph where the points become larger (more informative products) but move to the left (less popular) in the upper rows.

6. Conclusions and Future Work

This work focused on a firm's showcase decision: selecting a subset of products to offer in an offline channel from a larger product line offered through the online channel to maximize expected profits across both channels. A key component of our consumer demand model is that utility partworths change when customers learn about products by inspecting them physically in a brick-and-mortar store. In the context of this demand model, we formalized the decision problem, established computational hardness, and proposed approximation algorithms with theoretical guarantees. In addition, we used a demonstrative case study with messenger bags to estimate consumers' utility parameters in a conjoint study. Through this case study, we demonstrated that accounting for channel interactions can result in substantial gains in expected revenue (up to 40% in our case); the composition of the optimal assortment can also be significantly different. By laying out a framework for product showcasing, this work provides a platform for other interesting aspects of omnichannel retailing. Next, we discuss two specific directions in which this work can be extended.

This paper assumes that consumers exogenously decide whether to visit one or both of the channels. However, a consumer's decision to visit the offline (online) channel may depend on the products she examines online (offline) and offers of "in-store exclusives" by the firm to encourage store visits. Accounting for these effects by endogenizing the store visit decision is a promising future direction that is particularly relevant when the firm sells through multiple channels.

Furthermore, the utility model proposed in this work provides a framework for modeling product returns. Given that many online retailers, such as Warby Parker, Zappos, or Bonobos, offer generous return policies, the offline channel can be viewed as a way to mitigate costs of product returns. When consumers purchase from these retailers, they decide what to order based on their online evaluation of the available items. However, once they receive their order, they determine what they want to keep based on physical evaluation.

Finally, our utility model ignores interactions between attributes. While our proposed modeling framework readily extends (the partworths of the interaction terms change upon exposure to one of the attributes), the algorithmic methods may face computational challenges because the full-factorial assumption may be violated. Extending our algorithms to handle cases with constraints on the feasible products is a promising future direction.

Acknowledgments

The authors thank department editor Vishal Gaur, the associate editor, and the three anonymous referees for their detailed and thoughtful comments, which greatly improved this paper. The authors are also grateful to the seminar participants of London Business School, University of California at San Diego, and New Economic School for their constructive comments that have helped improve the manuscript. They also thank Prof. John Hauser for his help in designing the conjoint study.

Endnotes

[1] In fact, there is a growing trend of firms such as Bonobos, Warby Parker, Birchbox, etc., which started online, but then opened brick-and-mortar stores to showcase their product lines.

[2] This problem maximizes the firm's market share. However, for brevity, we use the term "sales" interchangeably with "market shares."

[3] For instance, it can be shown that the minimum cardinality assortment M that achieves $\mathbf{S}$ is of size $\max_{k=1}^{K}|S_k|$.

[4] Unlike an FPTAS, the computational complexity of the algorithm may scale exponentially in $1/\varepsilon$ in a PTAS.

[5] For instance, the assortment $\{(0,1),(1,0)\}$ cannot be expressed as the Cartesian product $S_1 \times S_2$ for *any* two subsets $S_1 \subseteq \mathcal{L}_1 = \{0,1\}$ and $S_2 \subseteq \mathcal{L}_2 = \{0,1\}$.

[6] The MILPs were solved using Gurobi Optimizer version 6.0.2 on a computer with a 3.5 GHz Intel Core i5 processor, 16 GB of RAM, and the Mac OSX Yosemite operating system.

[7] We could not compute profit maximizing subsets because we do not have cost data.

[8] The cash component was intended to eliminate any incentive for the participants to provide higher ratings for more expensive items to win a more expensive prize.

[9] We set the levels Black, Small, No strap pad, No water bottle pocket, and Empty bucket with no dividers to zero for attributes *Exterior design*, *Size*, *Strap pad*, *Water bottle pocket*, and *Interior compartment*, respectively. For the categorical variables, the coefficients of the "default" levels (set to zero) are not identified. Their combined effect is included in the intercept term. In total, nine coefficients and one intercept term were estimated for each data set.

References

Alptekinoğlu A, Corbett CJ (2010) Leadtime-variety tradeoff in product differentiation. *Manufacturing Service Oper. Management* 12(4):569–582.

Alptekinoğlu A, Grasas A (2014) When to carry eccentric products? Optimal retail assortment under consumer returns. *Production Oper. Management* 23(5):877–892.

Balakrishnan PV, Jacob VS (1996) Genetic algorithms for product design. *Management Sci.* 42(8):1105–1117.

Belloni A, Freund R, Selove M, Simester D (2008) Optimizing product line designs: Efficient methods and comparisons. *Management Sci.* 54(9):1544–1552.

Bront JJM, Méndez-Díaz I, Vulcano G (2009) A column generation algorithm for choice-based network revenue management. *Oper. Res.* 57(3):769–784.

Brynjolfsson E, Hu Y, Rahman MS (2009) Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Sci.* 55(11):1755–1765.

Cohen MC, Lobel I, Paes Leme R (2016) Feature-based dynamic pricing. Working paper, New York University Stern School of Business, New York.

Davis J, Gallego G, Topaloglu H (2013) Assortment planning under the multinomial logit model with totally unimodular constraint structures. Department of Industrial Engineering and Operations Research, Columbia University, New York.

Davis JM, Gallego G, Topaloglu H (2014) Assortment optimization under variants of the nested logit model. *Oper. Res.* 62(2): 250–273.

Désir A, Goyal V (2014) Near-optimal algorithms for capacity constrained assortment optimization. Working paper, Columbia University, New York.

Ding M (2007) An incentive-aligned mechanism for conjoint analysis. *J. Marketing Res.* 44(2):214–223.

Dobson G, Kalish S (1988) Positioning and pricing a product line. *Marketing Sci.* 7(2):107–125.

Dobson G, Kalish S (1993) Heuristics for pricing and positioning a product-line using conjoint and cost data. *Management Sci.* 39(2):160–175.

Feldman JB, Topaloglu H (2015) Capacity constraints across nests in assortment optimization under the nested logit model. *Oper. Res.* 63(4):812–822.

Forman C, Ghose A, Goldfarb A (2009) Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Sci.* 55(1):47–57.

Fruchter GE, Fligler A, Winer RS (2006) Optimal product line design: Genetic algorithm approach to mitigate cannibalization. *J. Optim. Theory Appl.* 131(2):227–244.

Gallego G, Topaloglu H (2014) Constrained assortment optimization for the nested logit model. *Management Sci.* 60(10): 2583–2601.

Gallego G, Li A, Lius Beltran J (2016) Product line design and pricing under logit model. Working paper, Columbia University, New York.

Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness* (WH Freeman and Company, New York).

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

Ghoniem A, Maddah B (2015) Integrated retail decisions with multiple selling periods and customer segments: Optimization and insights. *Omega* 55:38–52.

Ghoniem A, Maddah B, Ibrahim A (2016) Optimizing assortment and pricing of multiple retail categories with cross-selling. *J. Global Optim.* 66(2):291–309.

Green PE, Krieger AM (1985) Models and heuristics for product line selection. *Marketing Sci.* 4(1):1–19.

Green PE, Rao VR (1971) Conjoint measurement for quantifying judgmental data. *J. Marketing Res.* 8(3):355–363.

Green PE, Srinivasan V (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* 54(4):3–19.

Jagabathula S (2014) Assortment optimization under general choice. Working paper, New York University Stern School of Business, New York.

Jagabathula S, Rusmevichientong P (2017) A nonparametric joint assortment and price choice model. *Management Sci.* 63(9): 3128–3145.

Kohli R, Krishnamurti R (1987) A heuristic approach to product design. *Management Sci.* 33(12):1523–1533.

Kohli R, Krishnamurti R (1989) Optimal product design using conjoint analysis: Computational complexity and algorithms. *Eur. J. Oper. Res.* 40(2):186–195.

Kohli R, Sukumar R (1990) Heuristics for product-line design using conjoint analysis. *Management Sci.* 36(12):1464–1478.

Kök A, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Oper. Res.* 55(6):1001–1021.

Kuhfeld WF, Tobias RD, Garratt M (1994) Efficient experimental design with marketing research applications. *J. Marketing Res.* 31(4):545–557.

Lawler EL (1979) Fast approximation algorithms for knapsack problems. *Math. Oper. Res.* 4(4):339–356.

Li G, Rusmevichientong P, Topaloglu H (2015) The *d*-level nested logit model: Assortment and price optimization problems. *Oper. Res.* 62(2):325–342.

Mehra A, Kumar S, Raju JS (2013) "Showrooming" and the competition between store and online retailers. Working paper, Indian School of Business, Hyderabad, India.

Randall T, Ulrich K, Reibstein D (1998) Brand equity and vertical product line extent. *Marketing Sci.* 17(4):356–379.

Rodríguez B, Aydın G (2011) Assortment selection and pricing for configurable products under demand uncertainty. *Eur. J. Oper. Res.* 210(3):635–646.

Rusmevichientong P, Shen ZJM, Shmoys DB (2009) A PTAS for capacitated sum-of-ratios optimization. *Oper. Res. Lett.* 37(4):230–238.

Rusmevichientong P, Shen ZJM, Shmoys DB (2010) Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Oper. Res.* 58(6):1666–1680.

Rusmevichientong P, Shmoys D, Tong C, Topaloglu H (2014) Assortment optimization under the multinomial logit model with random choice parameters. *Production Oper. Management* 23(11): 2023–2039.

Subramanian S, Sherali HD (2010) A fractional programming approach for retail category price optimization. *J. Global Optim.* 48(2):263–277.

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

Ulu C, Honhon D, Alptekinoğlu A (2012) Learning consumer tastes through dynamic assortments. *Oper. Res.* 60(4):833–849.

Yen JY (1971) Finding the *k* shortest loopless paths in a network. *Management Sci.* 17(11):712–716.