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

MANAGEMENT SCIENCE
Vol. 69, No. 10, October 2023, pp. 5830–5847
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Model-Free Assortment Pricing with Transaction Data

Ningyuan Chen,[a] Andre A. Cire,[a,b] Ming Hu,[a,*] Saman Lagzi[c]

[a] Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada; [b] Department of Management, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada; [c] Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599

*Corresponding author

Contact: ningyuan.chen@utoronto.ca, https://orcid.org/0000-0002-3948-1011 (NC); andre.cire@rotman.utoronto.ca, https://orcid.org/0000-0001-5993-4295 (AAC); ming.hu@rotman.utoronto.ca, https://orcid.org/0000-0003-0900-7631 (MH); saman.lagzi16@rotman.utoronto.ca, https://orcid.org/0000-0002-0001-8572 (SL)

Received: January 5, 2021
Revised: September 16, 2021; March 16, 2022
Accepted: May 13, 2022
Published Online in Articles in Advance: January 16, 2023

https://doi.org/10.1287/mnsc.2022.4651

Abstract. We study the problem when a firm sets prices for products based on the transaction data, that is, which product past customers chose from an assortment and what were the historical prices that they observed. Our approach does not impose a model on the distribution of the customers' valuations and only assumes, instead, that purchase choices satisfy incentive-compatible constraints. The uncertainty set of potential valuations of each past customer can then be encoded as a polyhedral set, and our approach maximizes the worst case revenue, assuming that new customers' valuations are drawn from the empirical distribution implied by the collection of such polyhedra. We study the single-product case analytically and relate it to the traditional model-based approach. Then, we show that the optimal prices in the general case can be approximated at any arbitrary precision by solving a compact mixed-integer linear program. We further design three approximation strategies that are of low computational complexity and interpretable. In particular, the cutoff pricing heuristic has a competent provable performance guarantee. Comprehensive numerical studies based on synthetic and real data suggest that our pricing approach is uniquely beneficial when the historical data has a limited size or is susceptible to model misspecification.

History: Accepted by Omar Besbes, revenue management and market analytics.
Funding: This work was supported by the Rotman School of Management (TD-MDAL Research Grant) and the Natural Sciences and Engineering Research Council of Canada [Grants RGPIN-2015-06757, RGPIN-2020-04038, RGPIN-2020-06054, RGPIN-2021-04295].
Supplemental Material: The e-companion and data files are available at https://doi.org/10.1287/mnsc.2022.4651.

Keywords: data-driven • incentive-compatible • robust optimization • pricing • assortment • approximation algorithm

1. Introduction

Online retailing has seen steady growth in the last decade. According to the survey by Gaubys (2020), the global market share of e-commerce is expected to have surpassed 20% in 2022, and the trend is only accelerating. Retailers, however, face various challenges when transitioning to an online business model. First, the granularity of the data gathered from past customers, from which firms observe what customers bought, the products they viewed, and their prices at the time of the purchase, far exceeds that of the off-line business counterpart. Second, the firm is able to roll out products rapidly, and many new products (or new configurations of old products) can be displayed each day (Caro et al. 2020). Seemingly contradictory to the first point, there may be little purchase information for a firm to take timely actions on these new products, such as price correction and adjustments.

In an attempt to address this challenge, we study the pricing decision of an assortment of products for a firm based on its historical transaction data. The data record the assortment of products viewed by past customers and their prices, which may vary across customers because of promotions. The data also record the decision made by the customers: the purchase of at most one of the products. However, those customers who do not make any purchase may not need to be observed and recorded in the transaction data.

The common approach to handle this situation is what we refer to as "model-estimate-optimize." The firm first builds a discrete choice model to characterize how customers form their utilities and make choices. For example, in the multinomial logit (MNL) model, the probability of a customer choosing product j from a set of n products priced at $(p_1, \dots, p_n)$ can be expressed as

$$\frac{\exp(\alpha_j - \beta p_j)}{1 + \sum_{k=1}^{n} \exp(\alpha_k - \beta p_k)}$$

for some parameters $\{\alpha_k\}_{k=1}^n$ and β representing the average attractiveness of the products and the price sensitivity, respectively. It is equivalent to a random utility model in which a random customer has utility $\alpha_k - \beta p_k + \epsilon_k$ for product k, where ϵ_k follows an independent and identically distributed Gumbel distribution and customers choose the product, or no purchase, with the highest utility. The second step is to estimate. Based on the historical data, the firm can estimate the parameters (see Train 2009 for the detailed steps). Finally, the firm optimizes its pricing decisions by setting the optimal prices for future customers using the estimated model. This approach enjoys wide popularity because of its simplicity and computational efficiency.

However, in the setting we consider, the model-estimate-optimize approach falls short in three aspects. First, when the data size is small relative to the number of products, the estimation of the parameters (such as α_j in the MNL model) may be noisy and unstable. It is not clear whether the firm can rely on the estimation to make adequate pricing decisions. Second, the discrete choice model may not be able to capture the behavioral pattern in the data, which is referred to as model misspecification. In this case, the firm can choose a more complex model, potentially with more parameters, to minimize the misspecification error. This inevitably exacerbates the first issue as pointed out in Abdallah and Vulcano (2020). Third, the model-based approach is sensitive to whether the no purchases are being observed and recorded in the data in contrast to our robust approach.

In this paper, we propose a data-driven approach to the assortment-pricing problem. As opposed to the model-estimate-optimize approach, we consider no prior models on how customers form their utilities for the products. Instead, when a customer is observed to choose a product from a set of products with given prices, we assume that a (model-free) incentive-compatibility condition specifies the customer's potential valuations for the products. For example, when a product is purchased by a customer, the utility of that product must be necessarily at least as high as that of any of the products that are not purchased. Leveraging this condition, the data for a transaction define a polyhedral set containing the vector of valuations $(v_1, \ldots, v_n)$ of the particular customer, which is referred to as the incentive-compatible (IC) polyhedron. When a new customer arrives, without knowing anything about the customer's preferences, we uniformly sample from the IC polyhedra from customers in the historical data and set prices such that the expected revenue derived from the arriving customer is maximized when the customer's valuations, drawn from any sampled IC polyhedron, lead to the least possible revenue.

The contribution of this work is fourfold. First, to the best of our knowledge, this is the first study that leverages incentive-compatibility constraints to define a valuation polyhedron for pricing. With this novel approach, we circumvent the need for prespecified customer models as is customary in the random utility literature, investigating the optimal pricing directly based on the data. We believe this approach sheds new insights into the formulation of other related data-driven problems.

Second, by exploiting the structure of the polyhedra, we present a disjunctive model of the pricing problem and several structural results associated with the optimal prices. We show that the disjunctive model can be approximated at any desired precision by a compact bilinear program, itself solvable by a compact mixed-integer programming model after an appropriate reformulation. For scalability purposes, we also present low-complexity, interpretable approximation algorithms that are shown to achieve strong theoretical and numerical performance. We also study special cases of practical interest in which the optimal prices can be obtained efficiently.

Third, we build the connection between our model-free approach and the traditional model-based approach when the firm sells only one product. In particular, the revenue garnered by our approach within a given underlying model can be expressed in a closed form. As a result, when the number of historical customers approaches infinity and prices shown to past customers are randomly drawn from a uniform distribution, under mild technical conditions, we are able to establish a tight constant bound of $1/2$ for the performance of our model-free pricing relative to the optimal model-based revenue. This result showcases the robustness of our approach under a model-based framework. Moreover, we show that if the no-purchases are not recorded in the data, our approach can lead to strictly higher revenue than the best the model-based approach could do.

Fourth, we conduct a comprehensive numerical study based on synthetic and real data. In particular, we use the well-studied IRI data set (Bronnenberg et al. 2008) and fit different choice models, including the linear, MNL, and mixed logit models. We then generate a small number of customer purchases based on the models. After applying our approach to the data, the generated revenues significantly outperform the incumbent prices in the data for all the models, demonstrating the benefit of the data-driven, model-free approach: the insensitivity to model misspecification and stable performance when the data size is limited.

Our findings in the numerical study suggest that (1) our best approximation algorithm consistently recovers at least 96% of the optimal robust revenue in various settings, and its computational complexity scales linearly with the number of samples. (2) When the historical data are generated from commonly used discrete choice models, such as the MNL model, our data-driven approach performs well compared with the optimal prices under the correctly specified estimated model, especially for a limited data size. On the other hand, when the discrete choice model is misspecified, our approach is more robust and outperforms the misspecified optimal prices. (3) Applying our data-driven approach to real data sets leads to an increase in revenues over the incumbent prices in the data set.

2. Related Work

Our study is broadly related to three streams of literature. The first steam is the papers studying the estimation and optimization of the discrete choice models with prices, which provide the basis for the model-estimate-optimize approach. For the estimation of popular discrete choice models including the MNL model, Train (2009) provides an excellent review. Recently, two new choice models have drawn attention of scholars in the operations research community: the rank-based model (Farias et al. 2013) and the Markov chain choice model (Blanchet et al. 2016). Because of their flexibility, the estimation is not as straightforward as others, such as the MNL model. Several studies propose various algorithms to address the issue (van Ryzin and Vulcano 2015, 2017; Şimşek and Topaloglu 2018). As for optimal pricing, the MNL model is studied in, for example, Hopp and Xu (2005) and Dong et al. (2009). For the nested logit model, which is a generalization of the MNL model, Li and Huh (2011), Gallego and Wang (2014), and Li et al. (2015) investigate its optimal pricing problem. Zhang et al. (2018) study the optimal pricing problem of the generalized extreme value random utility models with the same price sensitivity. In contrast, Mai and Jaillet (2019) use a robust framework to tackle the same problem for extreme value utilities.

Although we also adopt a robust approach, we do not rely on a particular distribution, and the decision is fully guided by the data. Similarly, Rusmevichientong and Topaloglu (2012) and Jin et al. (2020) study the robust optimization problem of pricing or assortment planning in the MNL model. Several papers incorporate pricing into the rank-based model (Rusmevichientong et al. 2006, Jagabathula and Rusmevichientong 2017) and the Markov chain model (Dong et al. 2019) and study the optimal pricing problem. More recently, Yan et al. (2022) use the transaction data to fit a general discrete choice model of a representative consumer and solve the optimal pricing based on the fitted model by mixed-integer linear programming. Whereas the problem they study is essentially similar to ours, their approach relies on the model-estimate-optimize paradigm and, hence, is conceptually distinct from the proposed approach here.

In the context of revenue management, the definition of the term "data-driven" typically depends on the problem context. When the agent makes decisions in the process of data collection, the data-driven approach is usually associated with a framework that integrates such a process with decision making so that the agent is learning the unknown parameters or environment and maximizing revenues.[1] The previous works by Bertsimas and Vayanos (2017), Zhang et al. (2022), Cohen et al. (2018, 2020), Ettl et al. (2020), and Ban and Keskin (2021) fall into this category. It is connected to a large body of literature on demand learning and dynamic pricing. We refer to den Boer (2015) for a comprehensive review of earlier papers in this setting.

In contrast, our study essentially handles an off-line setting, in which the data is collected and given. Two recent papers also provide alternatives to the model-estimate-optimization approach. Bertsimas and Kallus (2020) leverage statistical methods, such as k nearest neighbors and kernel smoothing, to integrate past observations into the current decision-making problem given a covariate. It circumvents the step of estimating a statistical model. Elmachtoub and Grigas (2022) achieve a similar goal by skipping the minimization of the estimation error and directly focusing on the decision error. They design a new loss function that combines the errors in both stages: estimation and optimization. Both papers study a setting in which the optimization is conditioned on a covariate. Our problem is less general than their formulation and does not have a covariate, which allows us to utilize the special structure of the multiproduct pricing problem (i.e., the incentive compatibility of customers) that does not apply to their general approach. Ban and Rudin (2019) propose an algorithm integrating historical demand data and newsvendor-optimal order quantity without estimating the demand model separately.

There are a few papers using model-free approaches in revenue management. Allouah et al. (2022) study the problem that the seller observes samples from the buyer's valuation but is agnostic to the underlying distribution. The optimal price is solved for a family of possible distributions in the maximin robust sense. Chen et al. (2019) adopt the estimate-then-model approach, which is model-free, to estimate choice models using random forests. They do not consider the optimization stage. Our problem is similar to the work by Ferreira et al. (2016), who study the demand forecasting and price optimization of a new product by a retailer. However, it focuses on a single product, and it is unclear how to adapt it to multiple products. A few studies apply analytics to promotion planning (see, e.g., Cohen et al. 2017, Cohen and Perakis 2020). They consider more practice-based assumptions than ours and use various approximations for the demand model.

This paper uses the incentive compatibility of past customers as a building block and, hence, is related to the literature on auction designs, especially those papers using a data-driven or robust formulation. Bandi and Bertsimas (2014) study the multi-item auction design with budget-constrained buyers and use a robust formulation for the set of valuations. The uncertainty set is constructed using the historical information, such as means and covariance matrices. Because of the polyhedral structure of the uncertainty sets, the robust optimization problem is tractable. Our formulation investigates the worst case revenue in the IC polyhedron (see Section 3), which is similar to the idea of robust optimization for valuations drawn from uncertainty sets. However, there are three distinctions. First, in Bandi and Bertsimas (2014), the uncertainty set is constructed on the valuations of all bidders for a single product, whereas in our problem, the valuations of a single customer for all the products fall into a polyhedron. This is because of the different applications. Second, our data-driven approach averages over the empirical distribution of historical customers, whereas the historical information is used to construct the uncertainty set in Bandi and Bertsimas (2014). Third, the optimization problem in our problem cannot be solved efficiently, and we resort to approximation algorithms. Derakhshan et al. (2021) consider a data-driven optimization framework to find the optimal personalized reservation price of buyers when past bids are input to the algorithm. They do not impose assumptions on the valuation distributions and maximize the revenue when the valuation of the future customer is drawn from the empirical distribution of the historical data. This is similar to the motivation of this paper. However, because of the different contexts, the reduction and approximations have little in common with ours. Similarly, Koçyiğit et al. (2020) study the problem of designing an auction to sell an indivisible good to a group of bidders when the bidders' valuations come from an ambiguous distribution and the bidders' attitude toward this ambiguity is unknown. They design optimal mechanisms that are robust to the worst case realization of bidders' valuation and their attitudes. Koçyiğit et al. (2022) apply the robust approach to multiproduct pricing. The authors assume the valuation of a single customer is drawn from a rectangular uncertainty set and design mechanisms to maximize the worst case revenue ensuring incentive compatibility. They show that the optimal robust selling mechanism is to sell products separately with randomized posted prices. Finally, Allouah and Besbes (2020) study the single-item auction for a group of buyers when the seller does not have access to the valuation distribution of the buyers and buyers do not have any information about their competitors. Instead, the auction is designed for a general class of distributions with a competitive-ratio objective, which is conceptually similar to our model-free approach.

3. Problem Description

We consider a firm that observes the transaction data of m customers $\mathscr{C} = \{1, \dots, m\}$ with respect to an assortment of n products $\mathscr{P} = \{1, \dots, n\}$. Specifically, the firm's historical data includes the product prices $\mathbf{P}_i = (P_{i1}, \dots, P_{in}) > \mathbf{0}$ viewed by each customer $i \in \mathscr{C}$ as well as the product $c_i \in \mathscr{P}$ that the customer chose from that assortment. We consider that customers viewed all products $\mathscr{P}$ and purchased one product from that assortment (customers without purchases are censored in the data); both assumptions are made without loss of generality (Remarks 1 and 2). That is, we can incorporate the situation in which historical customers may have not observed some of the products in the assortment during their transactions, and removing those customers who made no purchase during their shopping session does not change our results. The goal of the firm is to set the product prices $\mathbf{p} = (p_1, \dots, p_n)$ for newly arriving customers that leverages this off-line historical data.

In this study, we investigate a pricing approach that operates on the full set of possible customer utilities under incentive-compatibility constraints. More precisely, let $v_{ij} \geq 0$ be the unknown valuation that customer $i \in \mathscr{C}$ assigns to each product $j \in \mathscr{P}$. We assume that the utility of purchasing j is given by $v_{ij} - P_{ij}$ (and, hence, quasilinear in v_{ij}) and such utilities must be compatible with observations from the data; that is, the set of all possible valuations of customer i is

$$\mathscr{V}_i \equiv \{(v_{i1}, \dots, v_{in}) \in \mathbb{R}^n_+ : v_{ic_i} - P_{ic_i} \geq 0,\\ v_{ic_i} - P_{ic_i} \geq v_{ij'} - P_{ij'}, \quad \forall j' \in \mathscr{P} \setminus \{c_i\}\}. \tag{1}$$

The first inequality in (1) indicates that the utility of purchasing product c_i is nonnegative. The second inequality specifies that valuations are incentive-compatible, that is, that the utility of purchasing product c_i must be either the same or larger than the utility for the remaining products $\mathscr{P} \setminus \{c_i\}$. We note that $\mathscr{V}_i$ is defined by a finite set of closed half-spaces and, hence, is a polyhedral set. We refer to $\mathscr{V}_i$ as the IC polyhedron of customer i.

For a newly arriving customer, however, the customer's valuation is not known to the firm. Based on historical data, it is reasonable to use the empirical distribution to form such an estimate. That is, we assume that the customer's valuation is equally likely to fall into one of the IC polyhedra $\mathscr{V}_1, \mathscr{V}_2, \dots, \mathscr{V}_m$. Given the new prices $\mathbf{p}$ and considering that the new customer's valuation falls into $\mathscr{V}_i$ for some i, we apply a robust approach in which the arriving customer picks product $j \in \mathscr{P}$ that yields the lowest possible revenue, and such a choice is consistent with $\mathscr{V}_i$ under prices $\mathbf{p}$. Such a revenue is described by the program

$$f_i(\mathbf{p}) \equiv \min_{\mathbf{v}_i \in \mathscr{V}_i, r \geq 0} \quad r \tag{DP}$$

$$\text{s.t.} \quad \bigvee_{j \in \mathscr{P}} \begin{pmatrix} r = p_j \\ v_{ij} - p_j \geq 0 \\ v_{ij} - p_j \geq v_{ij'} - p_{j'}, \quad \forall j' \in \mathscr{P} \end{pmatrix} \bigvee \begin{pmatrix} r = 0 \\ v_{ij} - p_j \leq 0, \quad \forall j \in \mathscr{P} \end{pmatrix}. \tag{2}$$

Model (DP) (short for disjunctive program) is a classic disjunctive program (Balas 1998), that is, the set of feasible solutions is defined by a disjunction of polyhedra representing the feasible valuations for each product. Specifically, with the objective min r, we select the product with the lowest revenue for which there exists a feasible valuation $\mathbf{v}_i \in \mathscr{V}_i$ that is also incentive-compatible under the new prices $\mathbf{p}$. This is modeled by the first disjunction term of (2), in which some product j (and, hence, price p_j) is selected only if its net utility (i.e., $v_{ij} - p_j$) is nonnegative and at least as large as that of choosing other products. The second disjunction term of (2) formulates the case in which the customer chooses no product—only possible if the utility of choosing any product under $\mathbf{p}$ is nonpositive.

Remark 1. As noted earlier, we assume that the data are censored, and each historical customer $i \in \mathscr{C}$ has purchased one product from the assortment. In particular, if some customer i has not purchased any products, we can assume that i belongs to the polyhedron with zero revenue in Model (DP), that is, the second disjunction term in (2). Thus, revenue driven by customer i is always zero for any prices, and therefore, customer i can be removed from the model.

Based on (DP), we assign a weight $1/m$ and sum over $i \in \mathscr{C}$ to take the expectation with respect to the empirical distribution. To maximize over $\mathbf{p}$, the objective function is expressed as

$$\tau^* \equiv \sup_{\mathbf{p} \geq 0} \frac{1}{m} \sum_{i \in \mathscr{C}} f_i(\mathbf{p}). \qquad \text{(OP)}$$

Problem (OP) (short for optimal pricing) highlights important distinctions from existing pricing approaches. First, it does not rely on a parametric discrete-choice model that explicitly specifies the distribution of customer utilities (see, e.g., Train 2009). Instead, we adopt a model-free approach and consider the worst case valuation for a given $\mathbf{p}$ under quasilinearity and incentive-compatible customer preferences, which are arguably weaker and more justifiable than existing parametric models to the best of our knowledge. Second, we do not attempt to estimate a nonparametric model (e.g., Jagabathula and Rusmevichientong 2017). Instead, the goal is to investigate the structure of (OP), trade-offs, and scenarios in which prices derived from (OP) are beneficial in comparison with existing models given its data-driven nature and emphasis on the historical data.

Remark 2. The choice model (DP) assumes that each customer $i \in \mathscr{C}$ sees the same assortment $\mathscr{P}$ and their historical prices $\mathbf{P}_i$. Whereas, in practice, customers may see different assortments, which can be subsets of $\mathscr{P}$, this assumption can be relaxed by setting a sufficiently large P_{ij} if customer i has not been offered product j (e.g., the sum of all historical prices). In that case, the structural results can be rewritten accordingly without loss of generality.

We next discuss two aspects of the pricing model.

Attaining the Maximum of (OP). We note that the maximum of (OP) may not be attainable because of the discontinuity of the customer choices in (2), hence, the use of supremum in the objective. For example, consider an instance with one customer ($m = 1$) and one product ($n = 1$), in which $P_{11} = P^*$ for some $P^* > 0$ and $c_1 = 1$. Thus, from (1), the customer valuation satisfies $v_{11} \geq P^*$. Suppose now we assign a price $p_1 \geq P^*$. It follows from (2) that the optimal revenue is zero because any valuation $P^* \leq v_{11} \leq p_1$ is incentive-compatible with the no-purchase option. However, for any $p_1 = P^* - \epsilon$ with $\epsilon > 0$, we have $\tau^* \to P^*$ as $\epsilon \to 0$. Thus, $f_i(\mathbf{p})$ is discontinuous in $\mathbf{p}$.

Interpretation as a Choice Model. One may take the worst case choices embedded in (DP) of all historical customers and assign an equal probability to them. This construction could potentially be interpreted as the choice behavior of the new customer. However, the essence of our approach is to take a conservative mapping from the price vector directly to the revenue, bypassing the choice model. We do not claim that (OP) can be used as a proper predictive model for consumer choices.

4. Model-Free Pricing for a Single Product

In this section, we investigate the special case of a single product, that is, $n = 1$. The traditional model-based approach for optimal pricing relies on a demand function that characterizes the distribution of customers' valuations. Given that such a notion is absent in our framework, our goal in this section is to explore the connection between the proposed pricing scheme and the model-based approach when a known demand function specifies consumer purchase behavior. That is, we attempt to rigorously characterize the performance of model-free pricing in a model-based setting. To that avail, in this section we provide a sufficient condition under which the model-free pricing achieves the optimal revenue in the model-based setting. Moreover, we provide tight performance guarantees for the model-free price in the model-based setting when historical prices are drawn independently from a uniform distribution. Finally, we characterize the sample complexity of our model-free approach. Namely, we show the rate, in terms of the number of samples, at which the model-free optimal price converges to its asymptotic optimal price.

Let $F : [0, +\infty) \to [0, 1]$ denote the cumulative distribution function (CDF) of the random customer valuation for the product. Thus, given a price $p \geq 0$, the probability that a customer purchases the product is

$1-F(p)$. The expected revenue at p, which is the target of model-based approaches, is $R(p) \triangleq p(1-F(p))$. Our objective is to compare the price from (OP) with $\arg\max_{p\geq 0} R(p)$. Note that the latter requires an infinite number of data points to estimate F exactly and is susceptible to data censoring, which we discuss in Lemma EC.1 in Section EC.4 of the e-companion.

To generate the historical data that can be used by our framework, we have to specify how the historical prices are drawn. We assume that the seller picks a price independently from a distribution with CDF $G(\cdot)$.[2] Observing a price $p \sim G$, the customer makes a purchase with probability $1-F(p)$. Because of the censoring, the historical data in our framework $\{(P_{i1}, c_i)\}_{i=1}^m$ can be thought of being generated independently with $\mathsf{P}(P_{i1} \geq p) = \int_p^{+\infty} (1-F(x))\mathrm{d}G(x)$ and $c_i = 1$ (we assume $P_{i1} = 0$ if the no-purchase option is chosen by customer i). As is shown in Section 5, the optimal price for our model-free approach satisfies $p_m^* \in \arg\max_{p\geq 0} p\sum_{i=1}^m \mathbb{I}(P_{i1} \geq p)$.

To compare the fundamentals of the model-based approach and our framework without the statistical noise, we let $m \to \infty$ and define p^* as $p^* \in \arg\max_{p\geq 0} p\left(\lim_{m\to\infty} \frac{\sum_{i=1}^m \mathbb{I}(P_{i1}\geq p)}{m}\right)$. Moreover, for the simplicity of analysis and exposition, we assume the following.

Assumption 1. *Both CDFs $F(\cdot)$ and $G(\cdot)$ have a probability density function denoted by $f(\cdot)$ and $g(\cdot)$, respectively.*

Our next result characterizes the optimal data-driven price p^* in our framework asymptotically and compares it to the optimal price out of the model-based approach, assuming the model is known.

Proposition 1. *Under Assumption 1, we have*

$$p^* \in \arg\max_{p\geq 0} p \int_p^{+\infty} g(x)(1-F(x))\mathrm{d}x. \tag{3}$$

Moreover, when $g(p) \propto f(p)/(1-F(p))$, the optimal price of our framework is also optimal for the model-based approach:

$$p^* \in \arg\max_{p\geq 0} R(p) = \arg\max_{p\geq 0} p(1-F(p)).$$

Proposition 1 reveals connections and differences between the two approaches. First, the optimal price of our framework depends on the distribution of the historical prices, which itself does not affect the optimal price of the model-based approach. The reason the historical prices impact the expected revenue in our data-driven setting is that our approach needs to recover the true valuation distribution from censored demand. To estimate $F(\cdot)$ well, one needs to know the fraction of no purchases for the historical prices because the observed demand is censored by the offered price.

Second, even when the data size is sufficiently large, the model-free pricing is not expected to converge to the optimal model-based prices in general. When the historical prices in our framework have a probability density function specified by the hazard rate function of the valuation distribution, the optimal prices of the model-based and model-free approaches coincide. However, one likely does not know the hazard rate function exactly; otherwise, it is possible to infer $F(\cdot)$ from such a hazard rate and use the optimal price of the model-based approach, assuming that the model is correctly specified.

4.1. Model-Free Pricing Under Uniformly Distributed Historical Prices

Next, we investigate a simple price distribution that does not require the information of F—the uniform distribution in the historical data—and study the performance of our framework. We consider the following assumption.

Assumption 2. *The distribution $f(\cdot)$ is supported on $[0,a]$ and $g(p) = 1/b$ for $p \in [0,b]$, where $0 \leq a \leq b$.*

Under Assumption 2, the first order condition for p^* in (3) implies that

$$R(p^*) = p^*(1-F(p^*)) = \int_{p^*}^a (1-F(x))\mathrm{d}x \triangleq S(p^*). \tag{4}$$

In other words, the optimal model-free price from our framework is an equal divider of the welfare: the revenue earned by the firm, $R(p^*)$, is equal to the surplus gained by the consumers, $S(p^*)$. Our question, therefore, can be framed as follows: if we plug the optimal price from our framework p^* (or, equivalently, an equal divider of the welfare) into $R(p^*)$, how large would the gap be between the optimal model-free profit $R(p^*)$ and the optimal model-based profit $\max_p R(p)$? Once equipped with Assumption 3, which we show later is not restrictive, Theorem 1 answers this question definitively.

Assumption 3. *The maximizer of (3), p^*, is the only solution to (4) in $[0,a]$ such that $p^*\int_{p^*}^a (1-F(x))\mathrm{d}x > 0$; $R(p)$ is unimodal in $[0,a]$ and has a unique maximizer $\hat{p}$.*

Theorem 1. *Suppose Assumptions 1–3 hold. We have $R(p^*)/R(\hat{p}) \geq \min\{\hat{p}/p^*, 1/2\}$. Moreover, the bound is asymptotically tight: for any $\epsilon > 0$, we can construct $F(\cdot)$ such that $R(p^*)/R(\hat{p}) \leq 1/2 + \epsilon$; for any $0 < x_1 < x_2 < 1/e$, we can construct $F(\cdot)$ such that $\hat{p} = x_1$ and $|p^* - x_2| < \epsilon$ such that $R(p^*)/R(\hat{p}) \leq \hat{p}^{1-\epsilon}/p^{*1-\epsilon}$.*

Proof Sketch. Based on $F(\cdot)$, we may have $p^* \leq \hat{p}$ or $p^* > \hat{p}$. Hence, we consider both cases separately. When $p^* \leq \hat{p}$, we show that $R(p^*)/R(\hat{p}) \geq 1/2$. Then, for any $\epsilon > 0$, we construct a distribution $F(\cdot)$ that leads to $R(p^*)/R(\hat{p}) \leq 1/2 + \epsilon$. Intuitively, the worst case performance of p^* is achieved when customers' valuation distribution is heavily concentrated at a.

When $p^* > \hat{p}$, we leverage the fact that p^* is the divider of welfare into equal portions of revenue and consumer surplus. At $\hat{p}$, consumer surplus is strictly larger than the firm's revenue. These observations lead to $R(p^*)/R(\hat{p}) \geq \hat{p}/p^*$. Moreover, for any $\epsilon > 0$, we construct a distribution $F(\cdot)$ that achieves $R(p^*)/R(\hat{p}) \leq \hat{p}^{1-\epsilon}/p^{*1-\epsilon}$. To elaborate on the intuition behind this worst case performance, beyond $\hat{p}$, both $S(p)$ and $R(p)$ deteriorate at a nearly identical rate, leading to p^* to be as far away from $\hat{p}$ as possible. □

We note that our bound is asymptotically tight as ϵ in Theorem 1 can be chosen to be arbitrarily close to zero. Theorem 1 establishes a performance bound when we use model-free pricing in the model-based setting with a demand function. Because the revenue maximizer $\hat{p}$ is typically larger than the welfare divider p^*, using our optimal price guarantees at least a half of the optimal revenue in this case. Note that our framework is completely model-free and does not assume a behavioral pattern of consumers. Theorem 1 suggests the robust performance of our approach even when there is an embedded model. Moreover, it also establishes the revenue bound for the welfare divider p^*, which may be of independent interest. The proof for the bound and the construction of tight instances are highly nontrivial. We defer it to the e-companion.

We point out that the bound established in Theorem 1 is applicable to a wide range of customer valuation distributions and can be improved significantly when customers' valuation distribution follows a specific distribution, such as the uniform distribution (see Proposition EC.2 in the e-companion for more details). The next lemma illustrates the applicability of the bound in Theorem 1 by establishing that Assumption 3 is not restrictive and is satisfied for a large array of distributions.

Lemma 1. *Suppose the distribution $f(\cdot)$ is supported on $[0, a]$ and has a strictly increasing generalized failure rate, that is, $xf(x)/(1-F(x))$ is strictly increasing on $[0, a]$. Then, Assumption 3 holds.*

Lemma 1 shows that Assumption 3 holds for distributions with a finite support that have a strictly increasing generalized failure rate. As Ziya et al. (2004) point out, assuming a strictly increasing generalized failure rate for customer valuation distributions is widely used in the literature of revenue management.

Another implication is that, when $\hat{p} \geq p^*$, Theorem 1 guarantees that $R(p^*)/R(\hat{p}) \geq 1/2$. The following lemma puts forward a sufficient condition for $\hat{p} \geq p^*$.

Lemma 2. *Suppose Assumptions 1–3 hold. Moreover, let $H(p) = \int_0^p (1-F(x))\mathrm{d}x / \int_0^a (1-F(x))\mathrm{d}x$ and assume that $H(\cdot)$ has a failure rate that is no less than that of $F(\cdot)$, that is, $h(p)/(1-H(p)) \geq f(p)/(1-F(p))$ for any $p \in [0, a]$, where $h(p) = (1-F(p))/\int_0^a (1-F(x))\mathrm{d}x$. Then, we have $\hat{p} \geq p^*$, and thus, $R(p^*)/R(\hat{p}) \geq 1/2$ by Theorem 1.*

As pointed out by Chen et al. (2020), the conditions of Lemma 2 are arguably mild and are satisfied for many distributions that have increasing failure rates and are commonly used in the marketing and operations management literature, for example, uniform and right-truncated exponential distributions.

Finally, if the no purchases are not recorded in the data, our approach can lead to strictly higher revenue than what the best model-based approach could do (see Section EC.4). This result demonstrates the advantage of our model-free robust approach in requiring no information on observing the no-purchases in contrast to the traditional model-based approach, which is sensitive to the information on the no-purchases (see Figure EC.3).

4.2. Sample Complexity of Model-Free Pricing

We investigate the sample complexity of model-free pricing, that is, how fast the optimal price of our framework approaches its asymptotic optimal price p^* given the increasing number of historical customers, m. The following proposition establishes the rate of convergence under some mild conditions.

Proposition 2. *Suppose Assumption 1 holds and $g(\cdot)$ is continuous and supported on $[0, b]$. Assume that p^* is the unique maximizer of $\max_{p\geq 0} p\int_p^b g(x)(1-F(x))\mathrm{d}x$ and there exists $\alpha > 0$ such that $p^*\int_{p^*}^b g(x)(1-F(x))\mathrm{d}x - p\int_p^b g(x)(1-F(x))\mathrm{d}x \geq \alpha(p-p^*)^2$ for all $p \in [0, b]$. Then, with m historical customers and for any $\epsilon > 0$, we have $\mathsf{P}(|p_m^* - p^*| \geq \epsilon) \leq 4e^{\frac{-\alpha^2\epsilon^4 m}{2b^2}}$. This suggests a sample complexity of $O\left(\left(\frac{b^2}{\alpha^2\epsilon^4}\right)\log\left(\frac{1}{\delta}\right)\right)$ to ensure that $|p_m^* - p^*| \geq \epsilon$ with probability at most δ.*

5. Reformulations of the Pricing Model (OP)

In this section, we investigate linear reformulations of (OP) that serve as the basis of our exact and tractable approximation strategies. We start in Section 5.1 with an analysis of the optimal policy structure for Problem (DP), presenting a small linear programming (LP) model to compute $f_i(\mathbf{p})$ that is compact with respect to the number of products and customers. We leverage this model in Section 5.2 to derive an alternative, compact mixed-integer linear program that approximates τ^* at any desired absolute error with respect to the optimal solution of (OP). Finally, we use this reformulation in Section 5.3 to show the optimal price structure of two commonly seen special cases.

5.1. Optimal Revenue from an IC Polyhedron

We first characterize the IC polyhedron $\mathscr{V}_i$ and $f_i(\mathbf{p})$ for given $i \in \mathscr{C}$ and $\mathbf{p}$. It can be interpreted as how customer

i behaves under the new price $\mathbf{p}$ based on the customer's historical choice. Proposition 3 uses classical polyhedral results to transform (DP) into an equivalent linear program.

Proposition 3. *Formulation* (DP) *is equivalent to the linear program*

$$f_i(\mathbf{p}) = \min_{\mathbf{v}_i^1,\ldots,\mathbf{v}_i^n,\mathbf{v}_i^{\emptyset},\mathbf{x}\geq 0} \sum_{j\in\mathscr{P}} p_j x_j \tag{DP-LP}$$

$$\text{s.t.} \quad v_{ij}^j - p_j x_j \geq 0, \quad \forall j \in \mathscr{P}, \tag{5}$$

$$v_{ij}^j - p_j x_j \geq v_{ij'}^j - p_{j'} x_j, \quad \forall j, j' \in \mathscr{P}, \tag{6}$$

$$v_{ij}^{\emptyset} \leq p_j x_{\emptyset}, \quad \forall j \in \mathscr{P}, \tag{7}$$

$$\sum_{j\in\mathscr{P}} x_j + x_{\emptyset} = 1, \tag{8}$$

$$v_{ic_i}^j - P_{ic_i} x_j \geq 0, \quad \forall j \in \mathscr{P} \cup \{\emptyset\}, \tag{9}$$

$$v_{ic_i}^j - P_{ic_i} x_j \geq v_{ij'}^j - P_{ij'} x_j, \quad \forall j, j' \in \mathscr{P} \cup \{\emptyset\}, \tag{10}$$

where $\mathbf{v}_i^j = (v_{i1}^j, \ldots, v_{in}^j)$ $\forall j \in \mathscr{P} \cup \{\emptyset\}$ *and* $\mathbf{x} = (x_1, \ldots, x_n, x_{\emptyset})$.

Model (DP-LP) (short for disjunctive program LP reformulation) is a linear program with $\mathcal{O}(n^2)$ variables and constraints. In particular, the variables $x_1, \ldots, x_n, x_{\emptyset}$ represent which product is picked by an arriving customer who draws the IC polyhedron $\mathscr{V}_i$, where $x_{\emptyset}$ encodes the no-purchase option. Constraint (8) ensures that the customer selects either one product or the no-purchase option. Constraints (5) and (6) imply that $\mathbf{v}_i^j$ is incentive-compatible with the choice x_j. They are equivalent to the jth disjunctive term of (2), whereas the other inequalities $j' \neq j$ of the same family become redundant whenever $x_j = 0$. This same reason applies analogously to the no-purchase option and Inequalities (7). Constraints (9) and (10) ensure that $\mathbf{v}_i^j \in \mathscr{V}_i$, that is, the valuation must be incentive-compatible with the historically chosen product c_i. The generated revenue $\sum_{j\in\mathscr{P}} p_j x_j$ is minimized over such $\mathbf{v}_i^j$ for all j and $\mathbf{x}$.

We now analyze the structure of (DP-LP) to draw insights into the optimal customer choices and reduce the size of the formulation. To this end, consider the set of valuations from (DP-LP) that are incentive-compatible with product $j \in \mathscr{P}$, that is, $\mathscr{W}_i^j(\mathbf{p}) \equiv \{\mathbf{v}_i^j \in \mathscr{V}_i : v_{ij}^j - p_j \geq 0, v_{ij}^j - p_j \geq v_{ij'}^j - p_{j'}, \forall j' \in \mathscr{P}\}$ and those incentive-compatible with the no-purchase option: $\mathscr{W}_i^{\emptyset}(\mathbf{p}) \equiv \{\mathbf{v}_i^{\emptyset} \in \mathscr{V}_i : v_{ij}^{\emptyset} \leq p_j, \ \forall j \in \mathscr{P}\}$. We now characterize in Lemma 3 when such valuation sets have at least one feasible point; that is, there exists a valuation vector for the products that is incentive-compatible with both the historical choice c_i under the historical price $\mathbf{P}_i$ and product j being chosen under the new price $\mathbf{p}$.

Lemma 3. *For any price* $\mathbf{p} \geq 0$, *the following statements a–c hold:*

a. *The no-purchase option is feasible to the ith customer type (i.e.,* $\mathscr{W}_i^{\emptyset}(\mathbf{p}) \neq \emptyset$*) if and only if* $p_{c_i} \geq P_{ic_i}$*, that is, the new price of the historically chosen product* c_i *remains the same or increases.*

b. *The purchase of the historically chosen product* c_i *by the ith customer type is always feasible (i.e.,* $\mathscr{W}_i^{c_i}(\mathbf{p}) \neq \emptyset$*) for all* $\mathbf{p} \geq 0$.

c. *The purchase of* $j \in \mathscr{P} \setminus \{c_i\}$ *by the ith customer type is feasible (i.e.,* $\mathscr{W}_i^j(\mathbf{p}) \neq \emptyset$*) if and only if* $p_j - p_{c_i} \leq P_{ij} - P_{ic_i}$*, that is, the price difference of j with respect to* c_i *remains the same or decreases.*

Lemma 3 provides easy-to-check conditions for whether the choice of a specific product or none is feasible. It also leads to a more compact formulation of (DP-LP). Intuitively, one can simply screen all products $j \in \mathscr{P}$ and the no-purchase option according to Lemma 3 for feasible options and choose the one with the lowest revenue. Formally, let $\mathbb{I}(C)$ be the indicator function of the logical condition C; that is, it is equal to one if C is true and zero otherwise. Proposition 4 applies Lemma 3 to a reformulation of (DP-LP) via a projective argument.

Proposition 4. *Formulation* (DP-LP) *is equivalent to the linear program*

$$f_i(\mathbf{p}) = \min_{\mathbf{x}\geq 0} \sum_{j\in\mathscr{P}} p_j x_j \tag{DP-C}$$

$$\text{s.t.} \quad \sum_{j\in\mathscr{P}} x_j = \mathbb{I}(p_{c_i} < P_{ic_i}), \tag{11}$$

$$x_j \leq \mathbb{I}(p_j - p_{c_i} \leq P_{ij} - P_{ic_i}), \ \forall j \in \mathscr{P}, \tag{12}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$.

Formulation (DP-C) (short for disjunctive program-combinatorial) reveals the combinatorial structure of the problem for prices $\mathbf{p}$ and the IC polyhedron of customer i. Specifically, the optimal objective value is the minimum price p_j among the feasible products and 0 if the no-purchase option is feasible, according to Lemma 3. Notice that (DP-C) is always feasible for any $\mathbf{p}$ because either $\sum_{j\in\mathscr{P}} x_j = 0$ or $x_{c_i} = 1$ is always a viable purchase option according to Lemma 3, (a) and (b). Furthermore, of particular importance to our methodology is the dual of (DP-C):

$$f_i(\mathbf{p}) = \max_{\mu_i\geq 0, \tau_i} \ \mathbb{I}(p_{c_i} < P_{ic_i})\tau_i - \sum_{j\in\mathscr{P}} \mathbb{I}(p_j - p_{c_i} \leq P_{ij} - P_{ic_i})\mu_{ij} \tag{DP-C-Dual}$$

$$\text{s.t.} \quad \tau_i - \mu_{ij} \leq p_j, \quad \forall j \in \mathscr{P}. \tag{13}$$

To draw insights on this dual problem, suppose prices are ordered as $p_1 \leq p_2 \leq \cdots \leq p_n$. If $p_{c_i} < P_{ic_i}$, then the no-purchase option is not feasible by Lemma 3(a) and the customer necessarily purchases one product from $\mathscr{P}$. In this case, the worst case revenue is p_{j^*}, where $j^* \equiv \min_{j\in\mathscr{P}}\{j : p_j - p_{c_i} \le P_{ij} - P_{ic_i}\}$. In an optimal solution (τ_i^*, μ_i^*), variable $\tau_i^* = p_{j^*}$ yields the revenue obtained when prices are set to $\mathbf{p}$. The solution μ_{ij}^* for each $j \in \mathscr{P}$ captures the lost objective value if the product is feasible, that is,

$$\mu_{ij}^* = \begin{cases} p_{j^*} - p_j, & \text{if } j < j^*, \\ 0, & \text{otherwise.} \end{cases}$$

We note that these solutions are optimal because all are nonnegative (because of the ascending order of prices), feasible to (13), and equal to the same solution value of (DP-C) as only terms indexed by $j \ge j^*$ have nonzero objective coefficients in the dual problem. The structure of the optimal duals also implies two immediate properties that we leverage in our reformulations.

Proposition 5. *For any $\mathbf{p} \ge 0$, the following statements hold:*

a. *We have $f_i(\mathbf{p}) \le \min\{p_{c_i}, P_{ic_i}\}$, which implies that the optimal revenue is bounded by the new price of the historically chosen product c_i.*

b. *Let $P^{\max} \equiv \max_{i\in\mathscr{C}} P_{ic_i}$ be the maximum historical price paid by any customer. There exists some $\mathbf{p}' \ge 0$ such that $f_i(\mathbf{p}') \ge f_i(\mathbf{p})$ and $p_j' < P^{\max}$ for all $j \in \mathscr{P}$, which implies that pricing any product higher than $P^{\max}$ does not lead to any increase in revenue.*

Given this dual interpretation and Proposition 5, we obtain an equivalent version of the dual that is written in terms of the single variable τ_i:

$$f_i(\mathbf{p}) = \max_{\tau_i \ge 0} \ \mathbb{I}(p_{c_i} < P_{ic_i})\tau_i \qquad \text{(DP-D)}$$

$$\text{s.t. } \ \tau_i \le p_j + \mathbb{I}(p_j - p_{c_i} > P_{ij} - P_{ic_i})P_{ic_i}, \ \ \forall j \in \mathscr{P}. \qquad (14)$$

If $p_j - p_{c_i} \le P_{ij} - P_{ic_i}$ for some $j \ne c_i$, then the jth product is feasible *for* purchase and Inequality (14) reduces to $\tau_i \le p_j$; that is, the maximum revenue is bounded by p_j. Otherwise, the inequality becomes redundant given Proposition 5(a). We also note that $\tau_i \le p_{c_i}$ for $j = c_i$ in (14), that is *product c_i* is always feasible *for customer i*, and the model is consistent with Proposition 5(a).

5.2. Robust Pricing Reformulation

We now develop reformulations to address our original pricing problem (OP). We consider the inner optimization of (OP) obtained by dropping the constant term $1/m$ and replacing $f_i(\cdot)$ by (DP-D), which has the same objective sense as the outer problem (hence, moving from a *sup min* problem to a *sup max* problem, which is simply *sup*):

$$\sup_{\mathbf{p},\,\boldsymbol{\tau}\ge 0} \ \sum_{i\in\mathscr{C}} \mathbb{I}(p_{c_i} < P_{ic_i})\tau_i \qquad \text{(OP-C)}$$

$$\text{s.t. } \ \tau_i \le p_j + \mathbb{I}(p_j - p_{c_i} > P_{ij} - P_{ic_i})P_{ic_i}, \quad \forall i \in \mathscr{C}, j \in \mathscr{P}. \qquad (15)$$

Multiplying the optimal value of (OP-C) (short for optimal pricing-combinatorial) by $1/m$ yields the optimal revenue τ^*. The difficulty in this problem is its nonconcave and discontinuous objective function, which is defined in terms of indicator functions on both open and closed half-spaces. We, however, exploit the constraint structure of (OP-C) to price products to any absolute error of τ^* by a two-step process, the first of which involves a significantly more computationally tractable model.

In particular, the indicator terms of (OP-C) can be reformulated in several ways (see, e.g., Belotti et al. 2016). For our purposes, we study the following equivalent bilinear mixed-integer model:

$$\sup_{\mathbf{p},\,\boldsymbol{\tau}\ge 0,\,\mathbf{y}} \ \sum_{i\in\mathscr{C}} y_{ic_i}\tau_i \qquad \text{(OP-B)}$$

$$\text{s.t. } \ \tau_i \le p_j + (1 - y_{ij})P_{ic_i}, \qquad \forall i \in \mathscr{C}, j \in \mathscr{P}, \qquad (16)$$

$$p_{c_i} < P_{ic_i} + (P^{\max} - P_{ic_i})(1 - y_{ic_i}), \ \forall i \in \mathscr{C}, \qquad (17)$$

$$p_j - p_{c_i} > P_{ij} - P_{ic_i} - (P^{\max} + P_{ij} - P_{ic_i})y_{ij}, \quad \forall i \in \mathscr{C}, j \in \mathscr{P} \setminus \{c_i\}, \qquad (18)$$

$$\mathbf{y} \in \{0,1\}^{m\times n}. \qquad (19)$$

To see the correspondence between the variables of (OP-B) (short for *optimal pricing* bilinear reformulation) and (OP-C), consider a set of new prices $\{p_1, \ldots, p_n\}$. The condition $\mathbb{I}(p_j - p_{c_i} > P_{ij} - P_{ic_i}) = 0$ for $j \in \mathscr{P} \setminus \{c_i\}$ implies $y_{ij} = 1$ because of (18), which may be interpreted as product j being feasible for customer i under the new prices as per Lemma 3(c). Moreover, when $\mathbb{I}(p_j - p_{c_i} > P_{ij} - P_{ic_i}) = 1$, both assignments $y_{ij} = 0$ and $y_{ij} = 1$ are feasible for (OP-B). However, $y_{ij} = 0$ leads to a (weakly) higher objective function value. Similarly, $\mathbb{I}(p_{c_i} < P_{ic_i}) = 0$ implies $y_{ic_i} = 0$ because of (17), indicating that the no-purchase option is available for customer i. Further, $\mathbb{I}(p_{c_i} < P_{ic_i}) = 1$ implies that both assignments of $y_{ic_i} = 1$ and $y_{ic_i} = 0$ are feasible, whereas $y_{ic_i} = 1$ always leads to a (weakly) higher objective function value.

To elaborate more on the equivalency of (OP-C) and (OP-B), we note that, for both programs, the objective and constraints for each customer i reduce to Problem (DP-D). Further, we also remark that Inequalities (17) and (18) are big-M constraints that rely on $P^{\max}$ defined in Proposition 5. For completeness, we formalize the validity of Model (OP-B).

Proposition 6. *At optimality, the solution values of (OP-C) and (OP-B) match.*

The set of feasible solutions to (OP-B), however, is not polyhedral because the linear Inequalities (17) and (18) are defined by open half-spaces. We propose to solve a parameterized version of (OP-B) by setting a precision parameter ϵ on the constraint violation of (OP-B), which allows us to replace the supremum by a maximum:

$$g(\epsilon) \equiv \max_{\mathbf{p},\boldsymbol{\tau}\ge 0,\mathbf{y}} \sum_{i\in\mathscr{C}} y_{ic_i}\tau_i \qquad \text{(OP-}\epsilon\text{)}$$

$$\text{s.t.} \quad \tau_i \le p_j + (1-y_{ij})P_{ic_i}, \quad \forall i\in\mathscr{C}, j\in\mathscr{P}, \quad (20)$$

$$p_{c_i} \le P_{ic_i} + (P^{\max} - P_{ic_i})(1-y_{ic_i}) - \epsilon, \quad \forall i\in\mathscr{C}, \quad (21)$$

$$p_j - p_{c_i} \ge P_{ij} - P_{ic_i} - (P^{\max} + P_{ij} - P_{ic_i})y_{ij} + \epsilon, \ \forall i\in\mathscr{C}, j\in\mathscr{P}\setminus\{c_i\}, \quad (22)$$

$$\mathbf{y}\in\{0,1\}^{m\times n}. \quad (23)$$

Because (OP-ϵ) restricts the feasible space of (OP-B) for $\epsilon > 0$, the optimal value must be no greater than the one from the original problem, that is, $g(\epsilon) \le m\tau^*$ for $\epsilon > 0$. Conversely, when $\epsilon = 0$, (OP-ϵ) serves as a relaxation, and therefore, $g(0) \ge m\tau^*$.

The challenge in solving $g(\epsilon)$ is to choose an appropriate $\epsilon > 0$ that leads to a sufficiently close approximation to the real supremum $m\tau^*$. Large values of ϵ may lead to poor approximations because of the discontinuity of (OP), whereas small values are not computationally tractable because of the numerical limitations of solvers. Theorem 2, however, shows that it suffices to solve $g(0)$ to obtain a sufficiently close value to $m\tau^*$ at any desired (absolute) error. It also prescribes a set of final prices with a formal guarantee that can be used by the firm.

Theorem 2. *Consider Formulation* (OP-ϵ) *with* $\epsilon = 0$. *The following statements hold:*

a. *The model has a finite optimal value attainable by some* $\mathbf{p}^0 \ge 0$.

b. *If a product* $j\in\mathscr{P}$ *is priced at zero by the optimal vector* $\mathbf{p}^0$, *we can reprice it at* $p_j^0 = \min_{i\in\mathscr{C},j\in\mathscr{P}} P_{ij} > 0$ *without losing optimality.*

c. *Let* $\mathbf{p}^0 > 0$ *be an optimal vector of positive prices after ordering, that is,* $0 < p_1^0 \le p_2^0 \le \cdots \le p_n^0$. *For any desired error* $\delta > 0$, *let* $\mathbf{p}'$ *be a vector of prices such that*

$$\mathbf{p}' = \left(p_1^0 - \frac{\delta'}{mn}, p_2^0 - 2\frac{\delta'}{mn}, p_3^0 - 3\frac{\delta'}{mn}, \ldots, p_n^0 - n\frac{\delta'}{mn}\right)$$

for any sufficiently small $0 < \delta' \le \delta$ *so that* $\mathbf{p}' > 0$. *Then,*

$$0 \le g(0) - m\tau^* \le g(0) - \sum_{i\in\mathscr{C}} f_i(\mathbf{p}') \le \delta,$$

where τ^* *is the optimal solution of the original pricing problem (OP).*

By Theorem 2, we are able to solve (DP) for any desired error δ using Reformulation (OP-ϵ) with $\epsilon = 0$. More precisely, the set of prices $\mathbf{p}^*$ can be obtained by

1. Solving (OP-ϵ) for $\mathbf{p}$ with $\epsilon = 0$, which is guaranteed to exist because of Theorem 2(a).
2. If any price is zero, increasing it to a positive value according to Theorem 2(b).
3. Applying the transformation from Theorem 2(c) to any desired error δ.

Whereas modern commercial solvers can address nonlinear models of the form (OP-ϵ), we also present an equivalent mixed-integer linear program by a standard big-M reformulation of the quadratic constraints. The program is the key to the analysis of our approximation algorithms.

$$g(0) = \max_{\mathbf{p},\boldsymbol{\tau},\bar{\boldsymbol{\tau}}\ge 0,\mathbf{y}} \sum_{i\in\mathscr{C}} \bar{\tau}_i \qquad \text{(OP-MIP)}$$

$$\text{s.t.} \quad (20), (21), (22), (23) \text{ with } \epsilon = 0, \quad (24)$$

$$\bar{\tau}_i \le y_{ic_i}P_{ic_i}, \quad \forall i\in\mathscr{C}, \quad (25)$$

$$\bar{\tau}_i \le \tau_i, \quad \forall i\in\mathscr{C}, \quad (26)$$

$$\bar{\tau}_i \ge \tau_i - (1-y_{ic_i})P_{ic_i}, \quad \forall i\in\mathscr{C}. \quad (27)$$

Prioritizing Market Share. We note that, in some market environments, the firm may want to ensure a certain purchase probability for the new customer through its pricing in addition to revenue maximization. The optimization program (OP-MIP) can be modified accordingly to tackle such a setting with market penetration considerations. For example, to guarantee a purchase probability of at least ρ for the new customer, it suffices to add the constraint $\sum_{i\in\mathscr{C}} y_{ic_i} \ge m\rho$ to (OP-MIP). Thus, our approach can accommodate the case when the firm prioritizes the market share consideration in its revenue maximization objective.

5.3. Special Cases

We now discuss two cases of practical interest in which (OP-MIP) can be solved analytically. First, we consider the scenario in which each individual customer is offered the same price for all products in the assortment. The prices, however, can be different per customer. This occurs when products are similar in nature and customers are offered personalized promotions over time. Proposition 7 states the structure of the optimal solution for this case.

Proposition 7. *Suppose that, for each customer* $i\in\mathscr{C}$, *all products* $j\in\mathscr{P}$ *have the same historical price* $P_{ij} = P_i$. *Furthermore, without loss of generality, assume prices are ordered, that is,* $P_1 \le P_2 \le \cdots \le P_m$. *The price vector* $\mathbf{p}^*$ *defined by* $p_j^* = P_{i^*}$ *for all* $j\in\mathscr{P}$, *where* $i^* = \arg\max_{i\in\mathscr{C}}\{(m-i+1)P_i\}$, *is optimal to (OP-MIP).*

Proposition 7 reveals a connection with the classic pricing literature. Specifically, if we perceive the set $\{P_1, \dots, P_m\}$ as the empirical distribution of the customer valuations, then our problem, when all products are offered at the same price, reduces to the well-studied revenue maximization problem $\max_p\{p \cdot d(p)\}$, where $d(p)$ is the demand function of price p.

As our second practical case, we consider a setting in which the price for any product is fixed over time. More precisely, prices may differ per product but not per customer. Proposition 8 also indicates that the optimal prices have a simpler structure; that is, it suffices to set them to their historical prices.

Proposition 8. *Suppose that, for each product $j \in \mathscr{P}$, all customers observe the same historical price P_j. The price vector $\mathbf{p}^*$ such that $p_j^* = P_j$ for all $j \in \mathscr{P}$ is optimal to (OP-MIP).*

6. Approximate Pricing Strategies and Analysis

Formulation (OP-MIP), albeit amenable to state-of-the-art commercial solvers, may still be challenging to solve because of its difficult constraint structure (e.g., the presence of big-M constraints). Moreover, Abdallah and Vulcano (2020) point out that, in retail, the number of SKUs in a family of products could be on the order of several hundred. Thus, even with a few data points per product, the size of (OP-MIP) can be significantly large. Whereas recent works focus on choice model estimation in such high-dimensional settings (Jiang et al. 2020), to the best of our knowledge, model-free pricing approaches are yet to be developed for such settings.

In this section, we analyze three interpretable and intuitive approximation algorithms that are of low-polynomial time complexity in the input size of the problem and, hence, scalable to large problem sizes. We discuss their benefits and worst case revenue performance in comparison with the optimal solution of (OP-MIP). Specifically, we evaluate in Sections 6.1 and 6.2 two standard heuristics based on historical prices and LP relaxations, respectively. In Section 6.3, we propose an approximation based on a simplified version of (OP-MIP), which provides the strongest approximation factor among the three policies, is efficient to compute, and is also interpretable.

We note that, in practice, the firm might want to use simpler pricing algorithms, such as offering the products at their average historical prices or at the prices observed by a random historical customer. However, it can be shown that such pricing algorithms can lead to an arbitrarily poor revenue performance in comparison with the optimal solution of (OP-MIP). See Propositions EC.3 and EC.4. Thus, we do not discuss them in detail in this section.

6.1. Conservative Pricing

A simple approximation a conservative firm may consider is to price all the products at their historically lowest purchase price. That is,

$$p_j = \min_{i \in \mathscr{C}: c_i = j} P_{ij}, \qquad \forall j \in \mathscr{P}. \tag{28}$$

By Inequality (20) for $j = c_i$, these prices guarantee that each customer i purchases at least one product (i.e., the no-purchase option is not chosen). Furthermore, it also follows that the final total revenue is at least $m\underline{P}$, where $\underline{P} \equiv \min_{i \in \mathscr{C}} P_{ic_i}$.

Such a pricing policy maximizes the demand at the cost of a lower profit margin and is, thus, referred to as "conservative pricing." We show that it has a straightforward worst case performance bound, which is also tight.

Proposition 9. *Let $\underline{P} \equiv \min_{i \in \mathscr{C}} P_{ic_i}$ and $\overline{P} \equiv \max_{i \in \mathscr{C}} P_{ic_i}$ be the minimum and maximum historical purchase prices, respectively. The revenue from the conservative pricing (28) is at least $\underline{P}/\overline{P}$ of the optimal value of (OP-MIP). Furthermore, this ratio is asymptotically tight as the worst case is achieved when the number of historical customers grows to infinity.*

As one may expect, when products are not similar in nature and their prices vary in a wide range, this approximation is too conservative and does not perform well. However, the ratio $\underline{P}/\overline{P}$ provides a useful benchmark that can be used to gauge the performance of other approximations.

6.2. LP Relaxation Pricing

It is a natural practice to consider the LP relaxation of Program (OP-MIP), that is, to replace the integrality constraint (23) by the continuous domain $\mathbf{y} \in [0,1]^{m \times n}$. The resulting LP model can be solved in (weakly) polynomial time, from which we can extract a candidate price vector $\mathbf{p}^{\text{LP}}$. We wish to evaluate the quality of $\mathbf{p}^{\text{LP}}$ with respect to the optimal prices of (OP-MIP). Notice that the remaining variables of (OP-MIP) can be easily determined when $\mathbf{p}$ is fixed to $\mathbf{p}^{\text{LP}}$.

In Example EC.1, we show that the resulting LP prices can have a worse worst case performance than the conservative pricing approach. However, even though the worst case revenue is not necessarily superior relative to conservative pricing, we show in Section 7 that the LP relaxation pricing usually outperforms conservative pricing numerically in most cases. Intuitively, conservative pricing is more concerned with the worst case, whereas LP relaxation can be close to maximizing the expected performance, specifically because we have tuned the big-M constrained to be very tight with respect to the input parameters.

6.3. Cutoff Pricing

In this section, we provide a heuristic by manipulating and reshaping the IC polyhedra. In particular, consider the historical customer i in (DP-C). Suppose we omit Constraint (12) when deciding which product to buy under $\mathbf{p}$ in the worst case. In other words, the customer does not fully follow the IC constraints; rather, as long as the price of the historically chosen product c_i is lower than its historical price, that is, $p_{c_i} \le P_{ic_i}$, all products are eligible for purchase. Therefore, if the customer plans to purchase a product (i.e., $p_{c_i} \le P_{ic_i}$), then the customer chooses the product with the lowest price, that is, product arg $\min_{j \in \mathscr{P}} p_j$, in the worst case.

We formulate this setting in the following model:

$$\max_{\mathbf{p},\, \boldsymbol{\tau} \ge 0} \left\{ \sum_{i \in \mathscr{C}} \mathbb{I}(p_{c_i} \le P_{ic_i})\tau_i \ : \ \tau_i \le p_j, \quad \forall j \in \mathscr{P} \right\}. \tag{OP-CP}$$

Under the assumed purchase rule, all historical customers choose the same product with the minimum price as long as they decide to purchase under $\mathbf{p}$. Thus, the key is to determine the minimum price p^*. Then, setting all products at this price ($p_i \equiv p^*$) maximizes the objective in (OP-CP). As we vary the value of p^*, the indicator functions only change values when $p^* = P_{ic_i}$. Letting $p^* = P_{ic_i}$ for some $i \in \mathscr{C}$, Formulation (OP-CP) simplifies to

$$\max_{i \in \mathscr{C}} \sum_{i' \in \mathscr{C}} \mathbb{I}(P_{i'c_{i'}} \ge P_{ic_i})\, P_{ic_i}, \tag{29}$$

which can be solved in $\mathcal{O}(m)$ time complexity by inspecting one P_{ic_i} at a time. We denote the optimal solution to (OP-CP) by $p^* = P_{ic_i}$ for some customer i. In particular, p^* can be perceived as a cutoff price: the historical customer i' does not purchase any product if and only if the historical price paid by customer i', $P_{i'c_{i'}}$, is below p^*. We leverage this to propose the cutoff pricing approximation $\mathbf{p}^{\mathrm{CP}}$ as follows. For every $j \in \mathscr{P}$,

$$p_j^{\mathrm{CP}} = \begin{cases} \min\limits_{i \in \mathscr{C}} \{P_{ic_i} : c_i = j,\, P_{ic_i} \ge p^*\}, & \\ \qquad\qquad \text{if } \{i \in \mathscr{C} : c_i = j, P_{ic_i} \ge p^*\} \ne \varnothing, & \\ \min\left\{ \max\left\{ \max\limits_{i \in \mathscr{C}} P_{ij}, p^* \right\}, \overline{P} \right\}, & \text{otherwise,} \end{cases} \tag{30}$$

where each product j is priced at its lowest historical purchase price that was greater than or equal to the cutoff price p^*. We note that not all products are priced at the cutoff price. Rather, they are often priced slightly higher than p^*. Such modifications do not impact the indicator functions and, thus, the optimal value of (OP-CP) but lead to better and less conservative empirical performances. We next show its performance in Proposition 10, recalling that $\underline{P} \equiv \min_{i \in \mathscr{C}} P_{ic_i}$ and $\overline{P} \equiv \max_{i \in \mathscr{C}} P_{ic_i}$ are the minimum and maximum historical purchase prices, respectively, and introducing $med(P)$ and $avg(P)$ as the median and mean of $\{P_{ic_i}\}_{i=1}^m$.

Proposition 10. *The cutoff pricing (30) generates a revenue that is at least* $\max\left\{\frac{1}{1+\log(\overline{P}/\underline{P})}, \frac{med(P)}{2avg(P)}\right\}$ *of the optimal value of (OP-MIP). Furthermore, this bound is asymptotically tight as it is achieved when both the number of products and number of historical customers grow to infinity.*

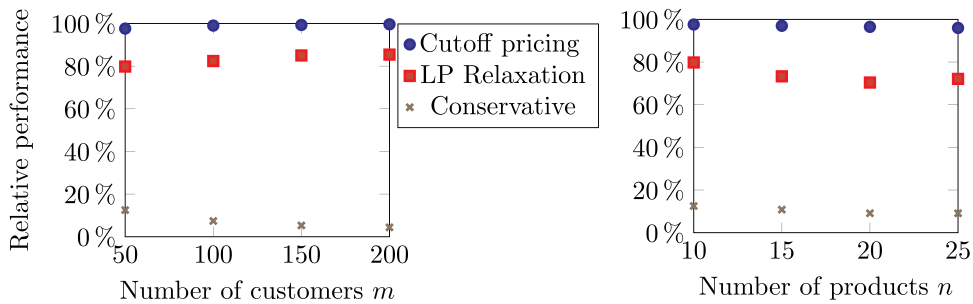
Proof Sketch. We create a random variable $\mathcal{X}$ whose support is on the set $\{P_{1c_1}, \ldots, P_{mc_m}\}$. We show that the performance bound of cutoff pricing can be translated to $\max_x x(1 - F_{\mathcal{X}}(x))/\mathbb{E}[\mathcal{X}]$, where $F_{\mathcal{X}}(x)$ is the CDF of $\mathcal{X}$. This is a classic problem and leads to the bound $\max\left\{\frac{1}{1+\log(\overline{P}/\underline{P})}, \frac{med(P)}{2avg(P)}\right\}$.

To show that the performance bound is asymptotically tight for any $\overline{P}$ and $\underline{P}$, we construct an example with n products and m customers that achieves a ratio arbitrarily close to $\frac{1}{1+\log(\overline{P}/\underline{P})}$, when both n and m grow arbitrarily large. Intuitively, this worst case performance happens when a small decrease in the price of historically lower priced products leads to all the historical customers with high historical purchase prices to choose them, hence showcasing the limitation of cutoff pricing in assuming a customer does not follow the IC constraints. □

Compared with Proposition 9, the cutoff pricing dramatically improves upon the worst case scenario of the conservative pricing, especially when $\overline{P} \gg \underline{P}$. Whereas Proposition 10 implies a strong performance bound for this pricing policy, we show that, for any $\underline{P}$ and $\overline{P}$, we can construct an asymptotic example such that the revenue from cutoff pricing is arbitrarily close to $\frac{1}{1+\log(\overline{P}/\underline{P})}$. Thus, we note that, when $\underline{P}/\overline{P}$ is near zero, in theory, the performance of cutoff pricing can be poor. However, this is unlikely to happen in practical settings, primarily because Proposition 10 suggests that cutoff pricing always generates a revenue with a factor that is at least a half of the ratio of the median to the mean of the historical purchase prices. Unless the purchase prices are highly skewed, this ratio is likely to be close to $1/2$.

Empirical observations of the price dispersion $\underline{P}/\overline{P}$ (see, e.g., Hosken and Reiffen 2004, Anania and Nisticò 2014, Dubois and Perrone 2015) often lead to a reasonable bound by Proposition 10. For example, in a study of grocery prices across the United States, Hosken and Reiffen (2004) investigate the frequency distribution of scaled prices for 20 categories of goods (the price of each good is scaled by its annual modal price) and show that the entirety of this distribution lies within $[0.6, 1.4]$. Moreover, in e-companion EC.1, we demonstrate bounds computed from 31 different categories in the IRI academic data set. We show that, for almost all

Figure 1. (Color online) Performance of the Three Approximation Algorithms



the categories, the bound implied by Proposition 10 falls into the range of 30%–50%. Finally, we also observe in Section 7 that cutoff pricing is superior numerically to the two earlier proposed heuristics.

Remark 3. In the case of model-free pricing with a single product, cutoff pricing recovers the optimal price. This follows from Proposition 7 and the fact that the cutoff price achieves the optimum of (29).

7. Numerical Analysis

We now present a numerical study of the proposed approaches on both synthetic and real data sets, evaluating our methodologies with respect to classical model-based methodologies in scenarios of practical interest. We start in Section 7.1 with an analysis of the empirical performance of the approximation pricing strategies from Section 6. In Section 7.2, we evaluate our data-driven pricing approach on "small data" regimes, which are typically challenging for classic model-based methods. Next, we consider a scenario in Section 7.3 in which the firm misspecifies the pricing model. Finally, in Section 7.4, we compare all approaches on a large-scale data set from the U.S. retail industry.

7.1. Approximation Performance

In this section, we use synthetic data to investigate the performance of the approximation algorithms developed in Section 6. We generate instances with $m \in \{50, 100, 150, 200\}$, $n \in \{10, 15, 20, 25\}$, and historical prices P_{ij} drawn uniformly at random from the interval (0, 10). Each customer chooses a product in the assortment with equal probability $1/n$. Based on the synthetic data, we compute the optimal value $g(0)$ from (OP-MIP) and the objective of the three approximations in Section 6. We report the objective values of the approximations relative to the optimal value $g(0)$ for 200 independent instances.

Figure 1 depicts the relative performance ratio in percentage (i.e., $100 \times$ optimal revenue/heuristic revenue) of the conservative, LP relaxation, and cutoff pricing when the number of customers m and the number of products n vary in the historical data. The figures suggest that conservative pricing performs poorly (achieving less than 15% of the optimal revenues on average), and cutoff pricing does the best among the three, obtaining at least 96% of the optimal value in all the cases. Increasing the number of customers or decreasing the number of products improves the performance of LP relaxation and cutoff pricing. Moreover, the numerical results suggest that cutoff pricing significantly outperforms conservative pricing as expected from its theoretical performance guarantee (Proposition 10) because cutoff pricing generally has an average performance that is an order of magnitude higher than that of conservative pricing. Finally, we observe that the numerical performance by cutoff pricing is far above that of its theoretical performance; we include additional tables in the e-companion EC.1 with the expected theoretical performance.

Table 1 shows the solution time of the conservative, LP relaxation, cutoff pricing, and optimal problem when

Table 1. The Average Solution Time of the Approximation Strategies and Optimal Solution

	Solution time in seconds			
(m, n)	Conservative	LP relaxation	Cutoff	Optimal
(50, 10)	0.041 (0.001)	0.453 (0.001)	0.040 (0.001)	1.477 (0.078)
(50, 15)	0.065 (0.001)	1.049 (0.005)	0.062 (0.001)	4.250 (0.195)
(50, 20)	0.080 (0.001)	1.762 (0.006)	0.083 (0.001)	5.691 (0.254)
(50, 25)	0.103 (0.001)	2.767 (0.007)	0.109 (0.001)	8.003 (0.313)
(100, 10)	0.081 (0.001)	0.929 (0.002)	0.078 (0.001)	17.260 (0.800)
(150, 10)	0.117 (0.001)	1.419 (0.003)	0.132 (0.002)	89.050 (5.400)
(200, 10)	0.165 (0.002)	2.018 (0.006)	0.181 (0.003)	525.800 (42.930)

Note. Standard errors are reported in parentheses.

the number of customers m and the number of products n vary in the historical data. The results suggest that, whereas the optimal solution time for midsize problems is not too egregious, it does not scale well when the problem size grows as is expected. However, cutoff pricing solves all problem sizes in under one second.

7.2. Small Sample Size

When the data size is small, model-based methods risk unstable estimations even when the model is correctly specified. In this section, we evaluate revenues obtained from the proposed pricing approaches in such cases, comparing with a classic MNL model (Train 2009).

We generate synthetic instances with $n = 10$ products and varying number of customers $m \in \{20, 30, \ldots, 500\}$. The probability of customer i choosing product j from the assortment is given by

$$\frac{\exp(\alpha_j - \beta P_{ij})}{1 + \sum_{k=1}^{n} \exp(\alpha_k - \beta P_{ik})},$$

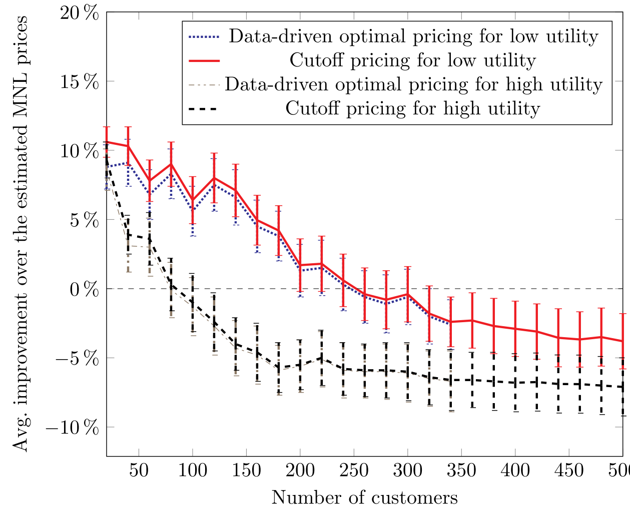
where $\beta = 0.5$. So the number of data points m is merely enough to estimate 10 parameters. We consider two sets of experiments for the remaining parameters α and the historical prices, representing low and high customer utility. Specifically, for the high-utility experiment, historical prices P_{ij} are drawn uniformly at random from the interval $[5.5, 8.5]$, and $\{\alpha_j\}_{j=1}^{10}$ are independent and drawn uniformly at random from the interval $[1, 3]$. In the low-utility experiment, historical prices are drawn uniformly at random from the interval $[2.5, 4.5]$ and $\{\alpha_j\}_{j=1}^{10}$ are independent and drawn uniformly at random from the interval $[-2, 0]$. Table 2 summarizes the setup of the two experiments. Note that, in both cases, this choice of parameters guarantees that the optimal price of the MNL model is within the range of the historical prices. In Section EC.1, we also study the effect of the dispersion of historical prices in the data on the comparative performance of our data-driven framework.

We simulate 200 independent instances for each m. In each instance, we calculate the optimal solution to (OP-MIP) and cutoff pricing. Moreover, we use the BIOGEME package developed by Bierlaire (2003) to estimate the parameters of the MNL model with the historical data and then calculate the optimal price based on the fitted model. Note that they are the prices obtained from the model-estimate-optimize approach under the correct model specification. We then evaluate

Table 2. Instance Parameters for the Study in Section 7.2

Experiment	α	P_{ij}
Low utility	$[-2, 0]$	$[2.5, 4.5]$
High utility	$[1, 3]$	$[5.5, 8.5]$

Figure 2. (Color online) Performance of the Data-Driven Optimal Pricing (OP-MIP) and Cutoff Pricing Relative to the Optimal MNL Prices Estimated from the Data



Note. The difference in the revenues is converted to percentage by dividing it by the optimal revenue of the model before being averaged.

the three sets of prices with respect to the ground-truth model and compare their expected revenues.

In Figure 2, we illustrate the average difference of the revenues of cutoff (data-driven optimal) pricing and the estimated MNL prices relative to the optimal MNL revenues when the parameters are known. For the data-driven optimal pricing policy, we do not compute it for more than 350 customers because of the prohibitive computational cost. Note that in the figure, if the average improvement quantity is positive, then it implies that our approach outperforms the estimated MNL prices. From the figure, when the number of customers is less than 70, both data-driven pricing schemes outperform the estimated MNL prices in both experiments. Note that the MNL prices are estimated based on the correct specification of the model. This experiment further suggests that data-driven approaches may be beneficial with respect to model-based approaches when the data size is small. It is also expected that, as the data size grows, the model-estimate-optimize approach would eventually converge to the optimal prices of the true model when the model is correctly specified. In this regime, data-driven approaches may not be beneficial. We note the MNL prices are estimated from uncensored data, favoring the model-based approaches and not affecting the data-driven prices (Remark 1). In Figure EC.3 and Section EC.4, we investigate the case when the MNL prices are estimated from censored data and demonstrate the benefit of our data-driven approaches in terms of the robustness to censored data.

7.3. Model Misspecification

Another potential benefit of the proposed approach is that it is agnostic to the underlying model and, thus,

Table 3. Instance Parameters for the Study in Section 7.3

Experiment	α	P_{ij}
Low utility	$[-2, 0]$	$[2.5, 4.5]$
High utility	$[1, 3]$	$[5.5, 8.5]$

less sensitive to model misspecification. We assess this scenario in the next experiment with synthetic instances. We consider $m = 50$ customers and $n = 10$ products in two experiment sets, also representing low and high utilities. The historical prices P_{ij} are drawn uniformly at random from $[2.5, 4.5]$ and $[5.5, 8.5]$ in the two experiments. For each customer, we generate choices using a mixed logit model with two classes (Train 2009). More precisely, given $(P_{i1}, \ldots, P_{in})$, the probability of customer i choosing product j is

$$\frac{1}{2} \cdot \frac{\exp(\alpha_{1j} - \beta_1 P_{ij})}{1 + \sum_{k=1}^{n} \exp(\alpha_{1k} - \beta_1 P_{ik})} + \frac{1}{2} \cdot \frac{\exp(\alpha_{2j} - \beta_2 P_{ij})}{1 + \sum_{k=1}^{n} \exp(\alpha_{2k} - \beta_2 P_{ik})}.$$

Here, we set $\beta_1 = 0.5$ and $\beta_2 = 2$. In each of the 200 instances, we randomly draw $(\alpha_{1j}, \alpha_{2j})$ independently from $[-2, 0]$ (for $P_{ij} \in [2.5, 4.5]$) and $[1, 3]$ (for $P_{ij} \in [5.5, 8.5]$). The range choices of P and α guarantee that the historical price ranges cover the optimal MNL prices with $\beta = 0.5$ (see Table 3).

We investigate the case in which the model is misspecified. In particular, we fit the MNL model, instead of the mixed logit model, using BIOGEME to the historical data. We calculate the optimal prices for the fitted MNL model and compare its expected revenue to our data-driven approaches (optimal solution and cutoff pricing) under the mixed logit model.

Table 4 suggests a better performance of our model-free approach as the optimal prices of the MNL model are designed for a misspecified model. Note that, in this case, the misspecification error between the MNL model and the mixed logit model with two classes is arguably mild. We may expect the benefit of model-free assortment pricing to be more substantial when there are strong irregular patterns in the data that cannot be captured by the assumed model. Combined with the results in Section 7.2, this further suggests that our proposed pricing models can be beneficial when the data size is small or the firm has little confidence in the modeling of the demand.

7.4. Real Data Sets

In this section, we apply model-free assortment pricing to the IRI Academic data set (Bronnenberg et al. 2008). The IRI data collects weekly transaction data from 47 U.S. markets from 2001 to 2012, covering more than 30 product categories. Each transaction includes the week and store of the purchase, the universal product code of the purchased item, the number of units purchased, and the total paid dollars. We investigate the product category of razors and the transactions from the first two weeks in 2001. We focus on this category primarily because it forms a proper assortment, that is, customers are unlikely to purchase more than one unit if they purchase any. To construct the assortments, we focus on the top 10 (out of 45) purchased products from all stores during the two weeks, that is, $n = 10$. The purchases of all other products are treated as "no purchase." An assortment is, thus, defined as the products of the same store in the same week when the customer visits. Moreover, we follow the procedures in van Ryzin and Vulcano (2015) and Şimşek and Topaloglu (2018): for each purchase record, four no-purchase records of the same assortment are added to the data set. Although model-free pricing does not depend on the censored demand, the benchmark model-based approaches do. It is ideal to create a reasonable fraction of customers who do not buy any products. After data preprocessing, there are in total 18,217 transactions with 2,460 unique sets of assortment price vectors (store/week combinations).

For the performance evaluation, we resort to a model (estimated from the data) that describes how consumers choose products and calculate the expected revenue under this model. We fit three models to the data to highlight the model-free or model-insensitive nature of our approach. More precisely, in the first MNL model we estimate $\{\alpha_j\}_{j=1}^{10}$ and β in the choice probability

$$\frac{\exp(\alpha_j - \beta P_{ij})}{1 + \sum_{k=1}^{10} \exp(\alpha_k - \beta P_{ik})}.$$

Table 4. Expected Revenues of Data-Driven Assortment Pricing and Misspecified MNL Model

	Expected revenue in the mixed logit model	
Pricing method	Low-utility experiment	High-utility experiment
Data-driven optimal	0.725 (0.005)	2.393 (0.007)
Cutoff pricing	0.734 (0.005)	2.415 (0.007)
MNL optimal pricing	0.635 (0.014)	2.113 (0.035)

Note. Standard errors are reported in parentheses.

In the second mixed logit model, we estimate $\{w_l, \beta_l, \alpha_{l1}, \ldots, \alpha_{l,10}\}_{l=1}^{2}$ in the choice probability

$$w_1 \frac{\exp(\alpha_{1j} - \beta_1 P_{ij})}{1 + \sum_{k=1}^{10} \exp(\alpha_{1k} - \beta_1 P_{ik})} + w_2 \frac{\exp(\alpha_{2j} - \beta_2 P_{ij})}{1 + \sum_{k=1}^{10} \exp(\alpha_{2k} - \beta_2 P_{ik})}.$$

Both models are estimated using BIOGEME. We also fit a linear demand model with parameters $\{\alpha_j, \beta_j\}_{j=1}^{10}$ and $\{\beta_{jk}, \gamma_{jk}\}_{j \neq k}$. The choice probability of product j in the linear model is

$$\alpha_j - \beta_j P_{ij} + \sum_{k \neq j} (\beta_{jk} I_{ik} P_{ik} + \gamma_{jk}(1 - I_{ik})),$$

fitted using the ordinary least squares. Here, $I_{ik} \in \{0, 1\}$ is the indicator for whether product k is included in the assortment seen by customer i. Note that we fit the choice probability separately for product $j = 1, \ldots, 10$ and the no-purchase probability is one minus their sum.

We generate three data sets for the estimated models considering $m = 50$ and $n = 10$. The historical prices P_{ij} are drawn from $\{0.9\overline{P_j}, 0.95\overline{P_j}, \overline{P_j}, 1.05\overline{P_j}, 1.1\overline{P_j}\}$, where $\overline{P_j}$ is the average price of product j in the IRI data. The customer choices are then generated using one of the three models. We calculate the prices using the optimal solution to (OP-MIP) and cutoff pricing and plug them into the three models to evaluate their expected revenues. We also compute the expected revenue of the incumbent prices, which is the average product prices in the IRI data under the three models.

Note that we apply our proposed approach to the data sets generated from the estimated models and then evaluate the expected revenues of the model-based prices, the optimal solution to (OP-MIP), or cutoff pricing under the corresponding models. A seemingly more straightforward approach is to directly apply model-free assortment pricing to the original IRI data set and then obtain the optimal prices. However, the major concern is the conflation of the model misspecification error and the efficacy of the pricing schemes. Suppose the model-free prices calculated using the IRI data perform poorly under, say, the MNL model. There might be two reasons: (1) the estimated MNL model accurately captures the pattern in the IRI data set but the data-driven approach fails to approximate the optimal prices of the MNL model, or (2) the MNL model does not fit the data and the data-driven approach, which is solely based on the data, cannot possibly approximate the optimal prices of the MNL model. We lack a reliable way to disentangle the two factors, the goodness of fit versus the performance of our approach. By the simulated data sets using the three models, we can control the goodness of fit and isolate the performance of the data-driven pricing schemes. Nevertheless, because the three models are fitted using the IRI data, they are expected to capture the choice patterns in reality to a large degree.

We also do not compare model-free assortment pricing with the optimal prices under the three models, but only the incumbent prices. This is because the optimal prices of the three models are not realistic. For example, the optimal prices under the MNL model are \$59.45 for all products, and the average optimal price under the linear demand model is \$99.70. However, the price range of the top 10 purchased products in the IRI razors data is \$3.29–\$7.51, and the model-free assortment pricing recommends prices (both optimal and cutoff) between \$5.68 and \$6.64. That is, the estimated demand models cannot extrapolate the demand outside the price range although they may approximate the demand patterns inside the price range well. Thus, the resulting optimal prices are not implementable, further suggesting the stability of the model-free approach.

Table 5 shows our results for this experiment with 200 instances for each of the three estimated models. The linear model generates significantly lower expected revenues than the other two, possibly because the estimated demand is lower in the region around the incumbent prices in the linear model. Compared with the incumbent prices, data-driven assortment pricing significantly improves the expected revenue under all three demand models. It suggests that our approach offers a robust improvement in this setting.

8. Practical Considerations and Concluding Remarks

We point out a few practical considerations when our approach is applied to real-world problems. The first issue is censoring, that is, when a customer walks away

Table 5. Expected Revenues of Data-Driven Assortment Pricing and the Incumbent Prices Under the Three Models Fitted Using the IRI Data

	Fitted demand model		
Pricing method	MNL	Mixed logit	Linear demand
Data-driven optimal	1.507 (0.007)	1.513 (0.008)	1.197 (0.018)
Data-driven cutoff pricing	1.730 (0.007)	1.731 (0.007)	1.477 (0.019)
Incumbent prices	1.464 (—)	1.468 (—)	0.944 (—)

Note. Standard errors are reported in parentheses.

without purchasing and, thus, cannot be observed in the data. Many pricing approaches struggle to handle censoring and completely ignoring data censoring may result in price distortion. Fortunately, our approach can handle data censoring well. Indeed, as mentioned in Remark 1, customers who do not buy anything can be removed from the data set without affecting the resulting prices. As a result, the data-driven prices do not depend on the censored customers.

One implicit assumption we make is that consumers have price-sensitivity identical to one, reflected in the quasilinear utilities. In fact, this assumption can be easily relaxed. The IC polyhedron can be constructed in the same fashion as long as an individual customer has the same known price sensitivity for all products. In this case, dividing (1) by the same factor results in the same polyhedron, and the theoretical results still hold.

Another assumption we make is that the firm does not observe the new incoming customer's information and, hence, assumes the customer behaves similarly to one of the previously seen customers with equal probability. In practice, the firm may potentially know that the customer is a returning customer. Then, one can analytically solve Model (OP-MIP) for that customer. If the firm has seen the customer beforehand, it is optimal under the worst case valuations of the customer to set the prices equal to the ones the customer observed in the customer's historical purchase. Moreover, using consumer features to predict shopping behavior is a popular practice in modern retailing (see Elmachtoub et al. 2021 for the value of personalized pricing). For example, an arriving consumer may have a similar background to a segment of past customers. To incorporate consumer features, we may put different weights (as opposed to equal weights) on the IC polyhedra $\mathscr{V}_1, \mathscr{V}_2, \ldots, \mathscr{V}_m$ in the formulation based on how similar the arriving consumer is to a past one. Computationally, our approach can still accommodate this case. It remains an exciting research direction to capture the consumer features and properly reflect them in the weights.

Incorporating product features is another exciting direction. Consumers form valuations for the base model of a product and certain add-on features. The valuation for a product configuration can be obtained as the sum of valuations of the base model and the added features. In this setting, the IC polyhedron of valuations can still be formulated for customers in the transaction data. In contrast to our problem, the pricing should concern the configurations instead of a separate base model and features. Yet it remains an open problem if nonlinear pricing can be analyzed and efficiently solved.

Whereas we are mainly interested in a situation in which the number of customers is low relative to number of products, scalability is still important from a practical point of view. We demonstrate in Section 7 that our mixed-integer programming reformulations can handle midscale problems with hundreds of samples. For large-scale data, our best approximation (cutoff pricing) is of low complexity and scales linearly with the number of samples. Whereas it provides interpretable and intuitive prices without the need for a commercial solver, such an approximation also has a robust theoretical guarantee and good empirical performance as suggested by our numerical study.

Acknowledgments

The authors gratefully acknowledge Omar Besbes (department editor), the associate editor, and the anonymous referees for their valuable comments, which helped improve this paper significantly.

Endnotes

[1] This is referred to as the online problem, which should not be confused with the notion of online retailing.

[2] We argue that this is a reasonable assumption. To estimate $F(\cdot)$ accurately in a model-based approach, the historical prices need to span the whole price range, presumably by conducting an experiment with random prices.

References

Abdallah T, Vulcano G (2020) Demand estimation under the multinomial logit model from sales transaction data. *Manufacturing Service Oper. Management* 23(5):1196–1216.

Allouah A, Besbes O (2020) Prior-independent optimal auctions. *Management Sci.* 66(10):4417–4432.

Allouah A, Bahamou A, Besbes O (2022) Pricing with samples. *Oper. Res.* 70(2):1088–1104.

Anania G, Nisticò R (2014) Price dispersion and seller heterogeneity in retail food markets. *Food Policy* 44:190–201.

Balas E (1998) Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Appl. Math.* 89(1–3):3–44.

Ban GY, Keskin NB (2021) Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. *Management Sci.* 67(9):5549–5568.

Ban GY, Rudin C (2019) The big data newsvendor: Practical insights from machine learning. *Oper. Res.* 67(1):90–108.

Bandi C, Bertsimas D (2014) Optimal design for multi-item auctions: A robust optimization approach. *Math. Oper. Res.* 39(4):1012–1038.

Belotti P, Bonami P, Fischetti M, Lodi A, Monaci M, Nogales-Gómez A, Salvagnin D (2016) On handling indicator constraints in mixed integer programming. *Comput. Optim. Appl.* 65(3):545–566.

Bertsimas D, Kallus N (2020) From predictive to prescriptive analytics. *Management Sci.* 66(3):1025–1044.

Bertsimas D, Vayanos P (2017) Data-driven learning in dynamic pricing using adaptive optimization. Working paper, MIT, Cambridge, MA.

Bierlaire M (2003) Biogeme: A free package for the estimation of discrete choice models. *Proc. Third Swiss Transportation Res. Conf.*

Blanchet J, Gallego G, Goyal V (2016) A Markov chain approximation to choice modeling. *Oper. Res.* 64(4):886–905.

Bronnenberg BJ, Kruger MW, Mela CF (2008) Database paper—The IRI marketing data set. *Marketing Sci.* 27(4):745–748.

Caro F, Kök AG, Martínez-de Albéniz V (2020) The future of retail operations. *Manufacturing Service Oper. Management* 22(1):47–58.

Chen N, Gallego G, Tang Z (2019) The use of binary choice forests to model and estimate discrete choice models. Preprint, submitted August 2, https://dx.doi.org/10.2139/ssrn.3430886.

Chen H, Chen YJ, Li Y, Zhang XM, Zhou S (2020) Digital goods reselling: Implications on cannibalization and price discrimination. Preprint, submitted December 9, https://dx.doi.org/10.2139/ssrn.3712041.

Cohen MC, Perakis G (2020) Optimizing promotions for multiple items in supermarkets. *Channel Strategies and Marketing Mix in a Connected World* (Springer), 71–97.

Cohen MC, Lobel I, Paes Leme R (2020) Feature-based dynamic pricing. *Management Sci.* 66(11):4921–4943.

Cohen MC, Lobel R, Perakis G (2018) Dynamic pricing through data sampling. *Production Oper. Management* 27(6):1074–1088.

Cohen MC, Leung NHZ, Panchamgam K, Perakis G, Smith A (2017) The impact of linear optimization on promotion planning. *Oper. Res.* 65(2):446–468.

den Boer AV (2015) Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys Oper. Res. Management Sci.* 20(1):1–18.

Derakhshan M, Golrezaei N, Paes Leme R (2021) Linear program-based approximation for personalized reserve prices. *Management Sci.* 68(3):1849–1864.

Dong L, Kouvelis P, Tian Z (2009) Dynamic pricing and inventory control of substitute products. *Manufacturing Service Oper. Management* 11(2):317–339.

Dong J, Simsek AS, Topaloglu H (2019) Pricing problems under the Markov chain choice model. *Production Oper. Management* 28(1):157–175.

Dubois P, Perrone H (2015) Price dispersion and informational frictions: Evidence from supermarket purchases. Preprint, submitted October 28, https://dx.doi.org/10.2139/ssrn.2682593.

Elmachtoub AN, Grigas P (2022) Smart "predict, then optimize". *Management Sci.* 68(1):9–26.

Elmachtoub AN, Gupta V, Hamilton M (2021) The value of personalized pricing. *Management Sci.* 67(10):6055–6070.

Ettl M, Harsha P, Papush A, Perakis G (2020) A data-driven approach to personalized bundle pricing and recommendation. *Manufacturing Service Oper. Management* 22(3):461–480.

Farias VF, Jagabathula S, Shah D (2013) A nonparametric approach to modeling choice with limited data. *Management Sci.* 59(2):305–322.

Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing Service Oper. Management* 18(1):69–88.

Gallego G, Wang R (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Oper. Res.* 62(2):450–461.

Gaubys J (2020) Ecommerce market share from 2017 to 2023. Accessed December 27, 2022, https://www.oberlo.ca/statistics/ecommerce-share-of-retail-sales.

Hopp WJ, Xu X (2005) Product line selection and pricing with modularity in design. *Manufacturing Service Oper. Management* 7(3):172–187.

Hosken D, Reiffen D (2004) Patterns of retail price variation. *RAND J. Econom.* 35(1):128–146.

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

Jiang ZZ, Li J, Zhang D (2020) A high-dimensional choice model for online retailing. Preprint, submitted September 17, https://dx.doi.org/10.2139/ssrn.3687727.

Jin Q, Lin JY, Zhou SX (2020) Price discounts and personalized product assortments under multinomial logit choice model: A robust approach. *IISE Trans.* 53(4):453–471.

Koçyiğit Ç, Rujeerapaiboon N, Kuhn D (2022) Robust multidimensional pricing: Separation without regret. *Math. Programming* 196:841–874.

Koçyiğit Ç, Iyengar G, Kuhn D, Wiesemann W (2020) Distributionally robust mechanism design. *Management Sci.* 66(1):159–189.

Li H, Huh WT (2011) Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing Service Oper. Management* 13(4):549–563.

Li G, Rusmevichientong P, Topaloglu H (2015) The *d*-level nested logit model: Assortment and price optimization problems. *Oper. Res.* 63(2):325–342.

Mai T, Jaillet P (2019) Robust multi-product pricing under general extreme value models. Preprint, submitted December 19, https://arxiv.org/abs/1912.09552.

Rusmevichientong P, Topaloglu H (2012) Robust assortment optimization in revenue management under the multinomial logit choice model. *Oper. Res.* 60(4):865–882.

Rusmevichientong P, Van Roy B, Glynn PW (2006) A nonparametric approach to multiproduct pricing. *Oper. Res.* 54(1):82–98.

Şimşek AS, Topaloglu H (2018) An expectation-maximization algorithm to estimate the parameters of the Markov chain choice model. *Oper. Res.* 66(3):748–760.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).

van Ryzin G, Vulcano G (2015) A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Sci.* 61(2):281–300.

van Ryzin G, Vulcano G (2017) An expectation-maximization method to estimate a rank-based choice model of demand. *Oper. Res.* 65(2):396–407.

Yan Z, Natarajan K, Teo CP, Cheng C (2022) A representative consumer model in data-driven multiproduct pricing optimization. *Management Sci.* 68(8):5798–5827.

Zhang M, Ahn HS, Uichanco J (2022) Data-driven pricing for a new product. *Oper. Res.* 70(2):847–866.

Zhang H, Rusmevichientong P, Topaloglu H (2018) Multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. *Oper. Res.* 66(6):1559–1570.

Ziya S, Ayhan H, Foley RD (2004) Relationships among three assumptions in revenue management. *Oper. Res.* 52(5):804–809.