Submitted to *Management Science*

manuscript

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

A Representative Consumer Model in Data-Driven Multi-Product Pricing Optimization

Zhenzhen Yan

Nanyang Technological University, Singapore, yanzz@ntu.edu.sg, https://sites.google.com/view/zhenzhenyan/home

Karthik Natarajan

Singapore University of Technology and Design, Singapore, karthik_natarajan@sutd.edu.sg,

Chung Piaw Teo

National University of Singapore, Singapore, bizteocp@nus.edu.sg,

Cong Cheng

Hebei University of Technology, China, chc_5588@163.com,

We develop a data-driven approach for the multi-product pricing problem, using the theory of a representative consumer in discrete choice. We establish a set of mathematical relationships between product prices and demand for each product in the system, including that of the outside option. We provide identification conditions to recover the underlying representative consumer model and show that with sufficient pricing experiments, the approach can identify the underlying demand model (more precisely, the associated pertubation function in the representative consumer model) accurately, up to a constant shift and a given tolerance level. This holds even when the demand data obtained are noisy realization of the theoretical demand. We use this approach to solve the multi-product pricing problem using a (mixed integer) linear optimization method. Extensive tests using both synthetic and industry data clearly demonstrates the benefits of this approach, which addresses the issue of model misspecification in traditional pricing methods using discrete choice models, and circumvents the computational issues associated with pricing methods that assume a known consumer valuation of each product.

Key words: Representative Consumer Model, Data-Driven Pricing, Identification, Sample Complexity Analysis

1. Introduction

Pricing as a strategy is commonly used to shape demand in a market, and has been extensively studied in the literature (see Talluri and van Ryzin (2004) for the use of pricing techniques in revenue management). The primary goal is to understand how demand across n products is influenced by the product prices, and to choose the right prices to maximize the seller's total profit. More precisely, given a limited number of observations on product prices $\boldsymbol{p}_t := (p_{t1}, \ldots, p_{tn})$ and normalized demand vectors[1] $\hat{\boldsymbol{x}}_t(\boldsymbol{p}_t) := (\hat{x}_{t1}(\boldsymbol{p}_t), \ldots, \hat{x}_{tn}(\boldsymbol{p}_t))$, with $t = 1, \ldots, T$, can we determine the optimal prices to maximize profit, assuming that product features and market conditions, except for prices, remain unchanged during this period?

These questions are important because it is now common for e-retailers to test market response to product pricing through A/B tests by adjusting their product prices periodically. Einav et al. (2011) analyze targeted pricing and auction design variations using eBay sales data and find that of the hundreds of millions of listings on a given day, more than half will reappear as a separate listing, often with modified sales parameters such as price. The availability of data from such pricing experiments presents companies with the opportunity to learn market responses to prices and increase profits through better pricing solutions.

One of the key difficulties in this problem is rationalizing the demand substitution patterns across different products. To this end, a large stream of research focuses on finding tractable instances of the product pricing problem using specific discrete choice models (cf. Section 2.2), derived based on a random utility maximization (RUM) framework. Many of these choice models result in a closed-form representation of choice probabilities, which renders model calibration and price optimization problems tractable. The maximum log-likelihood estimation method is often used to estimate parameters for such models.

More sophisticated choice models, such as the random coefficient logit model, have been proposed in the marketing and economics literature to capture complex substitution behaviour and provide more flexible modeling capability. However, estimation of the random coefficient logit model is much more complicated. Berry et al. (1995) developed a two-step regression method to estimate the random coefficient logit model from aggregate sales data, and this is a popular method in industry for developing demand models and estimating price elasticities. However, solving the price optimization problem with the random coefficient logit model is, to the best of our knowledge, a

[1] The normalized demand is calculated based on the demand for each product and the outside options. It corresponds to market share of the product in the market. Specifically, the sum of the normalized demand of all products and the outside options is equal to one. In this paper, for ease of exposition, we often use the term demand and normalized demand interchangeably, whenever the context is clear.

substantial open challenge (cf. van de Geer and Bhulai (2016); Li et al. (2019); and Marandi and Lurkin (2020) for a partial analysis).

In this literature, model misspecification is an important concern. For instance, in the estimation of a mixed logit choice model, consumer types must be predetermined, and firms often need a fair amount of understanding of the consumer market to arrive at an appropriate model. What if the demand model used in the price optimization model is misspecified? This is a salient issue, especially if the data available are noisy observations of demand under different product prices. In this paper, we address this model misspecification issue as follows. First, instead of assuming a particular discrete choice model, we use a more flexible approach to model demand. In particular, we restrict our attention to a class of representative consumer models with separable but general convex perturbation functions to achieve both flexibility in the model identification problem and tractability in the price optimization problem. These models go beyond the classic expected utility maximization framework (cf. Fudenberg et al. (2015)) and lead to flexible candidates to fit the data. Second, we use a nonparametric method to identify the perturbation functions, which provides additional flexibility in learning the demand model. The proposed nonparametric approach to deal with aggregate demand data contributes to the stream of literature on nonparametric models for individual choice data (Farias et al. (2013)) and analysis of aggregate demand data with semiparametric estimation methods (Berry et al. (1995); Shi et al. (2018)). We provide a small example in the Appendix (see Example 1 in Appendix B) to illustrate the effect of model misspecification on pricing problems and demonstrate the effectiveness of the proposed approach compared with presuming a particular discrete choice model.

Although some important instances of tractable pricing models have been developed in the discrete choice model literature, in general the pricing problem is hard to solve. In fact, Chen et al. (2018) have shown that the multi-product pricing problem is NP-hard even if the marginal distributions of the utilities are of small support size and the utilities of the products are mutually independent. This literature assumes that the valuation of each product is known, possibly for a subset of customers in the market, and the pricing problem now reduces to a bilevel optimization problem, in which the second-level problem assumes that each customer chooses a product to maximize her surplus, based on the prices set at the first level (cf. Hanson and Martin (1990)). This often resulted in a large-scale mixed-integer programming model (MIP) that cannot be easily solved. By focusing on aggregate demand and not how each consumer responds to prices, our approach offers a new way to circumvent this technical difficulty. We refer interested readers to the Appendix for an example (Example 2 in Appendix B) that compares the performance of the proposed approach with that of the widely used MIP approach in the literature.

The complexity of the problem is further exacerbated by the fact that pricing decisions in practice are often constrained by business strategies and operational concerns. For instance, when adjusting prices, the new pricing decisions must be confined to a certain range of the existing prices, since drastic changes in pricing decisions may turn customers away. This practical constraint leads to upper and lower bounds on pricing decisions. In this paper, we propose an MIP-based pricing model, which can easily incorporate different side constraints on prices due to business strategies and other concerns. In the absence of side constraints, we further show that the pricing problem in our approach reduces to a linear program.

In summary, the main objective of this paper is to develop a data-driven pricing approach that can mitigate the model misspecification issue in calibrating the demand model and ensure computational tractability in pricing the products, with potential constraints on prices.

1.1. A Representative Consumer Model in Discrete Choice

We propose to model the demand function using a flexible discrete choice model of a representative consumer. Consider a multi-product pricing problem in which the set of products offered by the seller is denoted by $\mathcal{N} = \{1, \ldots, n\}$ and $\{0\}$ is the outside option. Let the price of product j be p_j and $x_j(\boldsymbol{p})$ be the probability of choosing product j given the price vector $\boldsymbol{p}$. We use v_j to denote the deterministic utility of product j (excluding the price feature), and α is the price-sensitivity parameter. The term $v_j - \alpha p_j$ is known as the base utility of product j in the econometrics literature and denoted by u_j. Without loss of generality, we assume $v_0 = p_0 = 0$ ($u_0 = 0$) for the outside option. Consider the following deterministic optimization to obtain the choice probabilities $\boldsymbol{x}(p)$[2].

$$\max_{\boldsymbol{x} \in \Delta_n} \sum_{j=0}^{n} (v_j - \alpha p_j) x_j - C(\boldsymbol{x}), \tag{1}$$

where the optimization is over the n-dimensional unit simplex defined as

$$\Delta_n = \left\{ \boldsymbol{x} \in \mathbb{R}_+^{n+1} \;\middle|\; \sum_{j=0}^{n} x_j = 1 \right\}. \tag{2}$$

Model (1) provides an alternative model to calculate choice probabilities under a random utility model using a deterministic optimization model. In model (1), the consumer optimally chooses a probability distribution over the n products and the outside option. Her overall utility is modeled as the sum of the expected base utility and some nonlinear, deterministic perturbation function $C(\cdot)$ of the probability vector she selects. The perturbation function $C(\boldsymbol{x})$ is assumed to be convex. It plays the role of a regularization function and creates an incentive for choice diversification.

[2] Technically, $\boldsymbol{x}(p)$ depends on v_j. We suppress the dependence on v_j since it is a fixed constant in our problem.

For any random utility choice model, whose utility function is the sum of a base utility and a random noise term (with strictly positive density function), Hofbauer and Sandholm (2002) show that the problem can be mapped to a representative consumer model with a strictly convex perturbation function $C(\boldsymbol{x})$, such that the optimal solution to (1) is exactly the choice probabilities. In particular, Anderson et al. (1992) use this approach to recreate the Multinomial Logit (MNL) choice probabilities, with $C(\boldsymbol{x}) = \sum_j x_j \log x_j$. The modeling flexibility of the representative consumer model enables us to learn the underlying choice probabilities by fitting deterministic perturbation functions, instead of recovering the underlying random utility models.

When the convex perturbation function is separable (i.e., $C(\boldsymbol{x}) = \sum_j C_j(x_j)$), we have a separable representative consumer model (SRCM). Fudenberg et al. (2015) have shown that SRCM can be interpreted as a form of ambiguity-averse preferences for an agent who is uncertain about her true utility, which goes beyond the classic expected utility maximization framework. SRCM renders model calibration accurate and pricing problems tractable.

Remark 1. In Model (1), we have delineated the role $(p_1, \ldots, p_n)$ played in the representative consumer model to facilitate subsequent discussion of the impact of pricing on the choice probabilities $\boldsymbol{x}$. Furthermore, we have assumed implicitly that the pricing decisions do not affect the perturbation function $C(\cdot)$. This holds, for instance, in the MNL choice model. In cases in which this does not hold, we can nevertheless apply the above by performing an iterative search for the optimal pricing solution, solving the above problem in each iteration using a set of pricing experiments obtained from local perturbation of the current prices. We refer readers to Section 6 for more details.

Remark 2. In practice, the normalized demand data denoted by $\hat{\boldsymbol{x}}$ often deviates from the theoretical demand function denoted by $\boldsymbol{x}$ from a SRCM family. We name the normalized demand data in such a case noisy demand, and attribute the error between them to a measure of data quality in our problem.

Our problem is related to the quest to identify the underlying representative consumer model to rationalize a set of data. This problem is non-trivial even when the normalized demand data are generated exactly from a SRCM model. Allen and Rehbeck (2019) provide nonparametric identification results for a similar model—maximizing the sum of expected base utility and a perturbation term, assuming the perturbation term is independent of the product attributes. More specifically, they show that by varying the product attributes and through the corresponding normalized demand, the base utility u_j, as a linear function of the features, can be identified up to a location/scale normalization assuming no observation errors in the demand data.

Our paper is related to Allen and Rehbeck (2019), but we focus only on varying the price of the products $(\boldsymbol{p}_t, t=1,\ldots,T)$ and collect the corresponding normalized demand data $(\hat{x}_t, t=1,\ldots,T)$ to identify the function $\sum_j v_j x_j - C(\boldsymbol{x})$[3] up to a constant shift in location[4] and a threshold of ϵ in shape[5]. We refer readers to Section 4 for more details. Note that we could not identify v_j and $C(\cdot)$ separately, since we vary only the price features. But we show that a constant shift in location does not affect the optimal prices in our pricing problem. We further provide the number of pricing experiments required to ensure that the identified perturbation function is within ϵ threshold in shape from the true function.

In summary, our main contributions in this paper are as follows:

- We use the first-order optimality conditions in the SRCM model to obtain a mathematical relationship between prices and theoretical demand for each product. We also determine the number of pricing experiments needed to identify the perturbation functions up to a constant shift in location and a threshold of ϵ in shape. When the normalized demand data do not coincide with the theoretical demand function (i.e., the choice probabilities of a representative consumer in the market), the identification conditions may not hold exactly. In this case, we provide a nonparametric method to estimate the perturbation functions that minimize the worst-case deviation from the identification conditions.
- Building on the identification results, we provide a piecewise linear approximation of the objective function and develop a mixed-integer linear program for the pricing problem. We further show that a constant shift of the identified perturbation function does not affect the pricing solution. The proposed data-driven pricing framework has an additional advantage, in that it can easily handle other more complicated pricing constraints that can be formulated by mixed-integer linear constraints.
- We test the performance of the proposed methods in a comprehensive numerical study, using synthetic and industry data from the automobile and fast food sectors. While our approach is feature agnostic, the pricing solutions obtained from the approach we proposed can be mapped to a pricing strategy that relates to specific features in the problem. For instance, in the automobile example, our solution essentially translates to a pricing strategy that depends on the brand of the vehicle, even though the brand feature was not explicitly modeled in the problem.

The rest of the paper is organized as follows. In Section 2, we provide a literature review of the estimation and pricing problem with discrete choice models. In Section 3, we formulate the

[3] By a slight abuse of notation, we also call this the perturbation function, when the context is clear.

[4] The function f is a constant shift in location from g if $f(x) = g(x) + \alpha$, for some α, and for all x in the domain.

[5] We say a function f is within a threshold of ϵ in shape from g if $|f(x) - g(x)| \leq \epsilon$ for all x in the domain.

pricing problem using the representative consumer model and characterize the conditions when the pricing problem becomes convex and polynomial time solvable. In Section 4, we develop a data-driven approach for pricing with/without noisy demand data, and provide sample complexity analysis of the proposed approach. In Section 5, we provide computational results to demonstrate the performance of the proposed approach.

2. Literature Review

The paper uses a representative consumer's choice decision to model demand. The key in the model calibration problem is to identify or estimate the underlying representative consumer's choice model. This section provides a brief review of the most related identification and estimation methods, as well as several pricing methods often used in the literature.

2.1. Identification and Estimation of The Choice Model

There are two main streams of modeling consumer stochastic choices in the literature. One is based on a random utility framework and the other arises as deliberate randomization. We will review the identification and estimation literature based on these two perspective of utility modeling.

In a random utility framework, a consumer's utility of a product includes a base utility and a random error term. The base utility is commonly modeled as a linear function of product attributes. The consumer makes a choice decision to maximize the accrued utility from the product selection. Extensive literature has focused on estimating the consumer's utility function by identifying the coefficients of product attributes in the base utility and the distribution of the random error term, to gain a better understanding of the demand model and predict the choices made by the market (population of consumers). Lewbel (1998) provides identification conditions for a general utility model with a base utility linear in product attributes and an unknown distribution of latent error variables that may be related to the product attributes. Based on this, he further derives a root n consistent, asymptotically normal semiparametric estimator for the coefficients of product attributes in the base utility. A large family of estimation literature in discrete choice models presumes a distribution of random error terms and reduces the estimation problems to a parametric estimation of the coefficients in the base utility (cf. Bunch (1987); Ben-Akiva and Lerman (2018); and Gallego et al. (2019)). The maximum log-likelihood method is widely used in this literature for parameter estimation.

Another stream models the consumer's stochastic choice as deliberate randomization by subjects. Here, the utility function is modeled as a deterministic function that includes an expected base utility linear in choice probabilities and a nonlinear perturbation function of the probabilities. A representative consumer model (RCM) is a typical model in this stream, which is often represented

in a perturbed-utility form, i.e., the sum of expected utility and a perturbation function. Fudenberg et al. (2015) argue that the perturbed-utility representation provides a tractable alternative to model the consumer's stochastic choice behavior. The nonlinear perturbation function serve as a reward for consumers to choose a randomization policy across products. They provide two characterization conditions under which stochastic choice corresponds to a perturbed-utility representation: revealed preference theory and a generalized Luce's IIA condition. In particular, under a separable additive perturbation function, they further interpret the perturbed-utility model as a form of ambiguity-averse preferences for a consumer who is uncertain about her true utility. In fact, Hofbauer and Sandholm (2002) show that a deterministic representation in the form of a RCM can be obtained for choice probabilities from any additive random utility model, regardless of the distribution of the random error term as long as it admits a strictly positive density. Our paper adapts this flexible modeling perspective and focuses on learning the deterministic perturbation function instead of the distribution function of the random error terms in a random utility model.

Allen and Rehbeck (2019) provided nonparametric identification results for an additive perturbed-utility model, in which the expected utility and the perturbation term are in an additive form. They assume that the perturbation term is independent of product attributes. They show that the identification can only be up to a location/scale normalization, assuming there are no observation errors in the demand data. Our paper is closely related to this work. We provide identification conditions based on the optimality conditions of a SRCM, in which the perturbation function is also assumed to be independent of product attributes. Moreover, we further consider the case when the normalized demand data may not coincide exactly with the theoretical demand function, which adds another layer of complexity to the identification problem. We borrow the idea from Keshavarz et al. (2011) to build an estimation model to handle this case, which is to minimize the deviation from the optimality condition of the representative consumer model. Keshavarz et al. (2011) consider a general convex optimization problem (e.g., a representative consumer model) that depends on some parameters (e.g., prices). Given observations of several values of the parameters (e.g., prices) and their corresponding optimal or nearly optimal decisions (e.g., exact or noisy observations of choice probabilities), they provide an estimation model to impute the objective function (e.g., the associated perturbation function in the representative consumer model). The main idea of the estimation model is to minimize deviation from its optimality conditions.

There has also been growing interest in semiparametric and nonparametric estimation methods. Building on the pioneering work of the maximum score estimator of Manski (1975), Fox (2007) showed that the semiparametric maximum-score estimator of choice models is consistent when using individual choice data on selected choice sets. Recently, Shi et al. (2018) developed semiparametric identification and estimation approaches to multinomial choice models in a panel data setting with

individual fixed effects, including both individual choice data and aggregate demand data. In Farias et al. (2013), a nonparametric approach is developed to estimate the choice model from a set of ranking data and individual transaction data. In their approach, the choice model is viewed as distributions over a product's rankings.

2.2. Pricing with Choice Models

Let w denote the cost of the product and $F(p)$ the probability that a buyer has a valuation less than or equal to a price p. The optimal solution (cf. Myerson (1981)) is to sell the item at a fixed take-it-or-leave-it price p^* where $p^* = \arg\max_{p\geq 0}(p-w)(1-F(p))$. Log-concavity of the complementary distribution function $1-F(\cdot)$ implies quasi-concavity of the profit function in the price variable and concavity of the profit function in the choice probability variable, which renders this single-dimension optimization problem tractable (see Bagnoli and Bergstrom (2005)). Generalizing this result to multiple products, however, is much more complicated.

Hanson and Martin (1996) show that the multi-product profit function, unfortunately, is not even a quasi-concave function in the price variables for the MNL model. Although a non-convex optimization problem, Akçay et al. (2010) show that the profit function is unimodal in prices and demonstrate the efficiency of using first-order conditions to determine optimal prices. An alternate approach to this problem was developed by Song and Xue (2007) and Dong et al. (2009), who show that the pricing problem is a convex optimization problem for MNL in terms of the choice probability variables, and reformulate the problem by optimizing over these variables rather than the prices.

Building on this technique, Li and Huh (2011) study the multi-product pricing problem under a nested logit (NL) choice model. They show that the multi-product pricing problem is a convex optimization problem in the choice probability variables in certain cases. For related results, see Gallego and Wang (2014); Li et al. (2015); and Huh and Li (2015). Li and Webster (2017) study the class of paired combinatorial logit (PCL) models that allows for more general correlation structures, compared with the NL model, and identify conditions under which unique optimal price solution is computable. Zhang et al. (2018) show that the pricing problem with the GEV model can also be efficiently solved based on an explicit formula for the optimal markup in terms of the Lambert-W function. Alptekinoğlu and Semple (2016) further investigate the pricing problem under the exponomial choice model and show that the pricing problem is a convex optimization model. Our paper contributes to the literature by establishing a general convexity condition for the pricing problem. In addition, we further propose a piecewise linear approximation to the pricing model and turn the pricing problem into a mixed-integer linear program, which enables us to incorporate various types of price constraints in practice.

The piecewise linear approximation technique has been widely used in approximating nonlinear optimization problems. Various papers have been devoted to investigating the performance of the approximation (cf. Huchette and Vielma (2017) and Misener and Floudas (2012)). Magnanti and Stratila (2012) show that a piecewise linear function well approximates a separable concave function. In particular, they consider a problem of minimizing a nonnegative separable concave function over a compact feasible set. By approximating the problem with a piecewise-linear minimization problem over the same feasible set, it achieves an optimal value within $O(1+\epsilon)$ optimality.

3. Pricing with a Representative Consumer Model

In this section, we formulate the pricing problem using the representative consumer model, which provides a generic optimization formula for multi-product pricing problems and serves as a basis for the data-driven pricing framework. We further characterize the conditions when the pricing problem becomes convex and polynomial time solvable.

The seller's expected profit maximization problem in the multiple product setting is formulated as

$$\max_{\boldsymbol{p}\geq 0, p_0=0} \sum_{j=1}^{n}(p_j - w_j)x_j(\boldsymbol{p}), \tag{3}$$

where $\boldsymbol{p}$ is a price vector and $\boldsymbol{x}(\boldsymbol{p})$ is the demand function based on the price $\boldsymbol{p}$, with demand across products normalized to one. We use a general discrete choice model of a representative consumer specified in (1) to model the demand function.

One of the challenges with using the arbitrary convex perturbation function in (1) is that to fit such a function, one would typically need a large number of data points to obtain reasonably accurate solutions for pricing purposes. We focus instead on a special case of the representative consumer model in which the perturbation term is separable and convex, with $C(\boldsymbol{x}) := \sum_{j=0}^{n} C_j(x_j)$, and each $C_j(\cdot)$ is a convex function (see Fudenberg et al. (2015); Mishra et al. (2014); and Feng et al. (2017)). We assume that $C_j(\cdot)$ does not depend on the price attributes $\boldsymbol{p}$, since our goal is to identify $C_j(\cdot)$ by varying $\boldsymbol{p}$. We also assume that the other attributes of the products are fixed throughout.

Remark 3. By a slight abuse of notation, we can use $C_j(x_j)$ to denote $C_j(x_j) - v_j x_j$ without loss of generality, thereby focusing solely on the impact of the pricing attributes on choice substitution.

In this case, under the separable perturbation function, the optimality condition of (1) provides a closed-form relationship between prices and demand (See (5) in Theorem 1). Substituting the equations into the seller's pricing problem, we formulate the pricing problem in terms of demand

following the widely used approach in the literature (cf. Song and Xue (2007) and Li and Huh (2011)). The pricing problem reduces to:

$$\begin{aligned} \max_{\boldsymbol{x}} \ & \sum_{j=1}^{n} -w_j x_j - \frac{1}{\alpha}\sum_{j=1}^{n} x_j C_j'(x_j) + \frac{1}{\alpha}(1-x_0)C_0'(x_0) \\ \text{s.t.} \ & \sum_{j=0}^{n} x_j = 1, \\ & x_j \geq 0, \qquad \forall j = 0,1,\ldots,n, \end{aligned} \tag{4}$$

where $\boldsymbol{w}$ denotes the cost vector of the products.

THEOREM 1. *Suppose the perturbation function satisfies the following:*

C1. $xC_j'(x)$ *is a convex function in* x *for each* $j = 1,\ldots,n$.

C2. $xC_0'(1-x)$ *is a concave function in* x.

Then, the pricing problem (4) is a convex optimization problem in demand $\boldsymbol{x}$ *and the optimal prices can be computed in polynomial time. Moreover, if the optimal solution is* $\boldsymbol{x}^*$*, then the optimal pricing strategy is*

$$p_j^* = \frac{-C_j'(x_j^*) + C_0'(x_0^*)}{\alpha}, \ \forall j = 1,2,\ldots,n, \tag{5}$$

which comes from the optimality condition of (1).

We refer readers to Appendix A for all detailed proofs in the paper.

To see that the conditions in the theorem are mild, it is useful to relate the representative consumer model to the following semiparametric choice model, proposed by Natarajan et al. (2009). Unlike the random utility model, in which the distribution of the random error term $\tilde{\boldsymbol{\epsilon}}$ must be specified, the semiparametric choice model considers a set of distributions Θ for $\tilde{\boldsymbol{\epsilon}}$. Specifically, the maximum expected utility function under the semiparametric model is

$$\sup_{\theta\in\Theta} \mathbb{E}_{\tilde{\boldsymbol{\epsilon}}\sim\theta}\left[u_j + \tilde{\epsilon}_j\right].$$

Feng et al. (2017) establish the equivalence between the representative consumer model and the semiparametric model. In particular, if we restrict Θ to a set of distributions specified by the marginal distributions $F_j(\cdot), j = 0,\ldots,n$, Natarajan et al. (2009) show that the corresponding convex reformulation is equivalent to a representative consumer model with separable and convex perturbation function. In particular, the perturbation function is the form of

$$C_j(x_j) = -\int_{1-x_j}^{1} F_j^{-1}(t)dt. \tag{6}$$

In fact, for each differentiable convex function $C_j(\cdot) : [0,1] \to \mathbb{R}$, there exists a valid cumulative probability distribution $F_j(\cdot)$ such that (6) holds. Note that $C_j'(x_j) = F_j^{-1}(1-x_j)$ holds in general.

We use $C_j(\cdot)$ and $F_j(\cdot)$ interchangeably in this paper, depending on the ease of exposition. Mishra et al. (2014) show that by appropriately defining the marginal distribution function $F_j(\cdot)$, which corresponds to a unique $C_j(x_j)$, the optimal solution to the representative consumer model recreates commonly used discrete choice models in the literature, including the MNL and NL choice models as special cases.

The observation that the choice probabilities that arise from a classical random utility model can be recreated using an alternate representative consumer model has important ramifications. In our case, since the pricing model essentially hinges on demand, which depends on the choice probabilities of a representative consumer and not the underlying choice process of each customer, we can use pricing experiments and the response function from the market to recover the representative consumer model, to facilitate the search for the optimal pricing solution. This allows us to circumvent the technical difficulty in recovering the underlying random utility model in the demand estimation problem.

Proposition 1. *C1 and C2 hold if the corresponding marginal distributions satisfy the following conditions: (i) the tail distribution $\bar{F}_j(\cdot)$ for $j=1,\ldots,n$ is log-concave and (ii) the distribution $F_0(\cdot)$ is log-concave, where $\bar{F}_j(x)=1-F_j(x)$ for $j=1,\ldots,n$.*

Since log-concavity is satisfied by many common probability distributions (see Bagnoli and Bergstrom (2005)), this result identifies a large class of problems for which the pricing problem is now tractable. We refer readers to Appendices C.3 and C.4 for the connection between this approach and the classical pricing problem under MNL and NL model.

4. A Data-driven Approach to Price Optimization

In this section, we assume that we are given a set of product prices, the corresponding demand (for the n products), and the market share of the outside options across T pricing experiments. The data obtained are denoted by $\{p_{tj}, \hat{x}_{tj}\}$ for $t=1,\ldots,T, j=1,\ldots,n$, where p_{tj} and $\hat{x}_{tj}$ denote the observed price and the normalized demand of product j in epoch t. The normalized demand is calculated based on the demand for each product and the outside options. Specifically, the sum of the normalized demand of all products and the outside options is equal to one. We use $\hat{x}_{t0}$ to denote the market share of the outside option in each epoch t and assume it is available in the data. The assumption that the data set includes outside market shares seems strong. But based on our discussion with industry partners, companies can usually infer the outside market share from various sources, and view this information as critical to their pricing decisions. For instance, external sources of market intelligence[6] are available to estimate the outside market share. Furthermore,

[6] https://www.b2binternational.com/publications/competitor-intelligence/

even if there are errors in the normalized demand estimation, our method is able to recover the underlying choice model rather accurately, provided the number of experiments T is large enough.

For ease of exposition, we set the price sensitivity parameter $\alpha = 1$ throughout the rest of this section. For a given set of prices $\boldsymbol{p}$, we let $\boldsymbol{x}^{\boldsymbol{C}}(\boldsymbol{p})$ denote the corresponding theoretical demand, from a SRCM (1) specified by perturbation functions $\boldsymbol{C}(\cdot)$. We use $\hat{\boldsymbol{C}}(\cdot)$ to represent the estimated perturbation functions. To make the problem tractable, we need the following assumptions regarding the class of perturbation function $\boldsymbol{C}$ admissible in our model:

Assumption 1.

The mapping from prices $\boldsymbol{p}$ in the nonnegative orthant to its demand model from a SRCM (denoted by $\boldsymbol{x}^{\boldsymbol{C}}$ in the interior of the simplex) is surjective. Furthermore, the demand function $\boldsymbol{x}^{\boldsymbol{C}}$ is continuous: if $\boldsymbol{p}^{\epsilon} \to \boldsymbol{p}$ as $\epsilon \to 0$, then $\boldsymbol{x}^{\boldsymbol{C}}(\boldsymbol{p}^{\epsilon}) \to \boldsymbol{x}^{\boldsymbol{C}}(\boldsymbol{p})$.

Assumption 2.

The function $C_j'(\cdot)$ is non-decreasing and Lipschitz continuous, i.e., for all $x_1, x_2 \in (0,1)$

$$|C_j'(x_1) - C_j'(x_2)| \leq D|x_1 - x_2| \text{ for some } D > 0.$$

All functions $C(\cdot)$ that satisfy the above assumptions are considered admissible (denoted by $\mathcal{M}_I$), where index $\mathcal{I}$ is to indicate that we are considering a family of non-decreasing functions, since $\boldsymbol{C}(\cdot)$ is assumed to be convex in model (1). Assumption 1 imposes natural conditions on "smooth" consumer choice functions. Assumption 2 imposes regularity conditions on the perturbation functions in the representative consumer model. Specifically, Assumption 2 restricts the first derivative of the convex perturbation function $C_j(\cdot)$ to be Lipschitz continuous. The assumption holds for commonly used choice models. To see this, from Equation (6), we can infer that for any convex perturbation function $C_j(\cdot)$, the derivative is given by $C_j'(x_j) = -F_j^{-1}(1 - x_j)$. Hence the assumption reduces to the Lipschitz continuity of the inverse function of a probability distribution function, which holds for a large family of continuous probability distribution functions.

The performance of any data-driven algorithm must depend on the quality of data. To analyze the performance of the proposed data-driven algorithm, we need the following assumption.

Assumption 3.

In each epoch t, the sample demands (observed demands) $\hat{\boldsymbol{x}}_t$ are within an ϵ_S distance of the theoretical demands $\boldsymbol{x}_t^{\boldsymbol{C}}(\boldsymbol{p}_t)$ from a SRCM specified by $\boldsymbol{C}(\cdot)$, i.e.,

$$\sqrt{\sum_{j=0}^{n} (\hat{x}_{tj} - x_{tj}^{\boldsymbol{C}}(\boldsymbol{p}_t))^2} \leq \epsilon_S, \forall t = 1, \ldots, T.$$

The parameter ϵ_S characterizes the quality of data. If the normalized demands are inferred from a sufficiently large number of purchase records at each epoch generated from a SRCM model, and the market share of outside options can be accurately estimated, we expect ϵ_S to be small.

4.1. Identification Problem

As demonstrated in Allen and Rehbeck (2019), the identification problem is challenging even when demand is generated exactly from a SRCM model. The goal in this section is to provide identification conditions based on the optimality conditions of SRCM for perturbation functions $\boldsymbol{C}(\cdot)$ leading to the observed data. In particular, we infer the true function $\boldsymbol{C}'(\cdot) := (C'_1(\cdot), \ldots, C'_n(\cdot), C'_0(\cdot))$ from the following conditions: Find $\bar{C}$ such that for all $\boldsymbol{p} = (p_1, \ldots, p_n)$ in the nonnegative orthant,

$$\sum_{j=1}^{n} \left(p_j - (-\bar{C}'_j(x_j^{\boldsymbol{C}}(\boldsymbol{p})) + \bar{C}'_0(x_0^{\boldsymbol{C}}(\boldsymbol{p})))\right)^2 = 0. \tag{7}$$

We say that the solution $\{\bar{C}'_j : j = 0, 1, \ldots, n\}$ is a *constant shift* from $\{C'_j : j = 0, \ldots, n\}$ if $\bar{C}'_j(x) = C'_j(x) + \Delta$ for all x and j, and for some constant Δ. Recall that $\mathcal{M}_I$ denote the set of admissible perturbation functions and include the ground truth C'. For ease of exposition, we use x_{tj} to denote $x_j^{\boldsymbol{C}}(\boldsymbol{p}_t)$ when the context is clear.

THEOREM 2. *If $\bar{\boldsymbol{C}}' \in \mathcal{M}_I$ solves (7) for all $\boldsymbol{p}$ in the nonnegative orthant, then $\bar{\boldsymbol{C}}'$ is a constant shift of $\boldsymbol{C}'$.*

Proof: The problem is feasible, since the data are generated from the optimality condition in (5) under the ground truth $C'_j(\cdot)$. Consider two different solutions $\boldsymbol{C}'^1(\cdot)$ and $\boldsymbol{C}'^2(\cdot)$ to (7). For notational simplicity, for a price $\boldsymbol{p}_t$, we use y_{tj} to denote $C'^1_j(x_{tj})$ and y^*_{tj} to denote $C'^2_j(x_{tj})$ for $j = 0, \ldots, n$. Suppose $y_{tj} = y^*_{tj} + \Delta_{tj}$.

Take any two price vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. For $t = 1, 2$, $p_{tj} = -y_{tj} + y_{t0} = -y^*_{tj} + y^*_{t0}$ according to (7), which implies that we have $\Delta_{tj} = \Delta_{t0}$ for all $j = 1, \ldots, n$. We denote it as Δ_t. We show next that $\Delta_1 = \Delta_2$. To see this, W.L.O.G suppose $x_{20} \le x_{10}$, then $x_{2j} \ge x_{1j}$ for some $j \in \{1, \ldots, n\}$ since the demand vector is in a simplex. From the surjectivity of the admissible solutions, as assumed in Assumption 1, we can always find a new price vector $\boldsymbol{p}_3$ such that $x_{30} = x_{20}$ but $x_{3j} = x_{1j}$, which implies $C'^1_j(x_{3j}) = C'^1_j(x_{1j})$ and $C'^1_0(x_{30}) = C'^1_0(x_{20})$, i.e., $y^*_{3j} + \Delta_3 = y^*_{1j} + \Delta_1$ and $y^*_{30} + \Delta_3 = y^*_{20} + \Delta_2$. Hence $\Delta_3 = \Delta_1$ and $\Delta_3 = \Delta_2$, since $y^*_{3j} = C'^2_j(x_{3j}) = C'^2_j(x_{1j}) = y^*_{1j}$ and $y^*_{30} = C'^2_0(x_{30}) = C'^2_0(x_{20}) = y^*_{20}$. Finally, we have $\Delta_2 = \Delta_3 = \Delta_1$. Therefore $\bar{C}'_j(x) = C'_j(x) + \Delta$ for any x and j by taking $\boldsymbol{C}'^2(\cdot)$ as the true function, which is one of the solutions to (7). Q.E.D.

Theorem 2 indicates that Equation (7) for all $\boldsymbol{p}$ in the nonnegative orthant is sufficient to identify a function up to a constant shift. Yet this involves an infinite number of equations. Interestingly, we can do this with fewer pricing experiments if we allow errors of small ϵ beyond a constant shift to the ground truth.

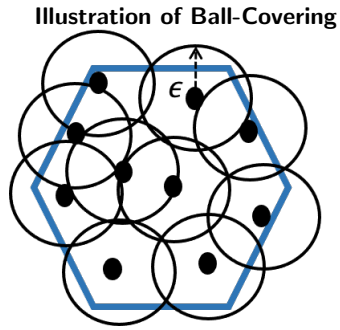
DEFINITION 1 (IDENTIFY UP TO A CONSTANT SHIFT WITHIN A THRESHOLD ϵ). A function $\bar{C}'_j$ is a constant shift of ground truth C'_j within a threshold of ϵ if $|\bar{C}'_j(x) - C'_j(x) - \Delta| \le \epsilon$ for all x in the simplex, for some constant Δ.

This identification criterion aligns with Allen and Rehbeck (2019), concentrating the identification on the location error (a constant shift) and shape error, which is characterized by the threshold ϵ. As we will show in the next section, a constant shift in location does not affect the pricing solution; hence the effect of the identification on the pricing solution depends on the threshold of shape error ϵ. The smaller the ϵ, the better the pricing solution we can obtain. To achieve a smaller ϵ in the identification problem, a larger number of pricing experiments will be required. We will establish the relation between the sample size and threshold ϵ in the subsequent analysis.

First, we show in Theorem 3 that the global structure of $\boldsymbol{C}'$ can be pegged down to within the threshold of ϵ, uniformly across all points in the simplex, using a polynomial number (in the number of products n) of pricing experiments. The analysis is based on an inspiring recent result in the approximate Carathéodory theorem (see Vershynin (2018)). We present the result in Lemma 1 for the sake of completeness.

Lemma 1. *(Covering polytopes by balls)[Vershynin (2018) Corollary 0.0.4]. A unit simplex in $n+1$ dimension ($\{\boldsymbol{x} : \sum_{i=0}^{n} x_i = 1, x_i \geq 0\}$) can be covered by at most $N := (n+1)^{\lceil 1/\epsilon^2 \rceil}$ Euclidean balls of radii $\epsilon > 0$.*

Figure 1 Illustration of Ball-Covering Problem



The ball covering is obtained using the set

$$\mathcal{N} := \left\{ \frac{\sum_{j=1}^{k} \boldsymbol{v}_j}{k} : \boldsymbol{v}_j \text{ a vertex in the simplex} \right\},$$

allowing for duplication in the choice of $\boldsymbol{v}_j$ and with $k = \lceil 1/\epsilon^2 \rceil$. The family of $\epsilon-$balls (at most N balls) centered at $\mathcal{N}$ covers the simplex in $n+1$ dimensions (see Figure 1, in which the black points represent the points in $\mathcal{N}$). This follows directly from a probabilistic argument, which shows that for every point $\boldsymbol{x}_c$ in the simplex and every integer k, one can find vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ such that $\left\| \boldsymbol{x}_c - \frac{1}{k} \sum_{j=1}^{k} \boldsymbol{v}_j \right\|_2 \leq \frac{1}{\sqrt{k}}$.

Let $\boldsymbol{x}_t \in \mathcal{N}$ for $t = 1, \ldots, N$ denote the demand vector located in the centers of the balls covering the simplex, and $\boldsymbol{p}_t, t = 1, \ldots, N$ denote their corresponding price vectors. Let Δ_t denote the corresponding gaps between the candidate choice function $\bar{\boldsymbol{C}}'$ and the true function $\boldsymbol{C}'$ at these points. We want to find $\bar{C}$ such that

$$\sum_{t=1}^{N} \sum_{j=1}^{n} \left(p_{tj} - (-\bar{C}'_j(x_{tj}) + \bar{C}'_0(x_{t0}))\right)^2 = 0. \tag{8}$$

We start the argument by showing that for any solution $\bar{\boldsymbol{C}}'$ to (8), we have $\Delta_t = \Delta$ for all $t = 1, \ldots, N$. Then we conclude the identification up to a constant shift within a threshold ϵ from the Lipschitz continuity assumption of the perturbation functions. To see the first point, note that for any two points $\boldsymbol{x}_a, \boldsymbol{x}_b$ in $\mathcal{N}$, we can write

$$\boldsymbol{x}_a = \frac{\sum_{j=1}^{k} \boldsymbol{v}_j^a}{k}, \boldsymbol{x}_b = \frac{\sum_{j=1}^{k} \boldsymbol{v}_j^b}{k}$$

for some collection of extreme points $\boldsymbol{v}_j^a$ and $\boldsymbol{v}_j^b$.

Claim 1. *When $n \geq 2$, we can use the structure of the ball covering to find $\boldsymbol{x}_c$ such that the demand vectors of the three points satisfy $x_{cl} = x_{al}$ for some product l and $x_{cl'} = x_{bl'}$ for another l', and $\boldsymbol{x}_c \in \mathcal{N}$.*

To see this, we define $\mathbb{I}_i^a = \{j : \boldsymbol{v}_{ji}^a = 1\}$ and $\mathbb{I}_i^b = \{j : \boldsymbol{v}_{ji}^b = 1\}$. In other words, we use $\mathbb{I}_i^a$ to denote the set of indices of extreme points generating $\boldsymbol{x}_a$ with the ith entry being 1. We can hence rewrite the formula of $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ as follows: $\boldsymbol{x}_a = \frac{\sum_{i=0}^{n} \sum_{j \in \mathbb{I}_i^a} \boldsymbol{v}_j^a}{k}$ and $\boldsymbol{x}_b = \frac{\sum_{i=0}^{n} \sum_{j \in \mathbb{I}_i^b} \boldsymbol{v}_j^b}{k}$. Note that by the pigeonhole principle, there must exist a $l \in \{0, \ldots, n\}$ such that the cardinality of $|\mathbb{I}_l^a| \leq \frac{k}{n+1}$. Furthermore, by omitting product l, there exists another product $l' \in \{0, \ldots, n\}$ and $l' \neq l$ such that the cardinality of $|\mathbb{I}_{l'}^b| \leq \frac{k}{n}$. Consider a partition $\mathbb{I}^c$ such that for $i \neq l, l'$, $|\mathbb{I}_l^c| = |\mathbb{I}_l^a|$, $|\mathbb{I}_{l'}^c| = |\mathbb{I}_{l'}^b|$, and $\sum_{i=0}^{n} |\mathbb{I}_i^c| = k$. $\boldsymbol{x}_c = \frac{\sum_{i \neq l, l'} \sum_{j \in \mathbb{I}_i^c} \boldsymbol{v}_j^c + \sum_{j \in \mathbb{I}_l^a} \boldsymbol{v}_j^a + \sum_{j \in \mathbb{I}_{l'}^b} \boldsymbol{v}_j^b}{k}$ satisfies the aforementioned properties.

Following the same argument as in Theorem 2 we can prove

$$\Delta_a = \Delta_{al} = \Delta_{cl} = \Delta_c = \Delta_{cl'} = \Delta_{bl'} = \Delta_b,$$

which implies that $\Delta_t = \Delta$ for all $t = 1, \ldots, N$. Note that all points x in the simplex are within ϵ neighborhood of a point in $\mathcal{N}$. By Lipschitz continuity, we obtain the following theorem:

Theorem 3. *If a Lipschitz continuous function $\bar{\boldsymbol{C}}' \in \mathcal{M}$ (with a Lipschitz constant D) solves (8), which needs $N := (n+1)^{\lceil 1/\epsilon^2 \rceil}$ carefully designed prices to ensure that the corresponding demand can induce balls of radii $\epsilon > 0$ to cover the simplex, then under Assumptions 1 and 2, $\bar{\boldsymbol{C}}'$ is a constant shift of $\boldsymbol{C}'$, within a threshold of $O(\epsilon)$, for all x in the unit simplex.*

Theorem 3 identifies the theoretical demand model characterized by $\boldsymbol{C}'$ via specially constructed pricing experiments. However, the experiments cannot be constructed without knowing $\boldsymbol{C}'$, since we will not be able to obtain the corresponding prices to induce those demand vectors in $\mathcal{N}$. Fortunately, when the price vector $\boldsymbol{p}$ is randomly generated, we can expect the corresponding demand to be close to every vector in $\mathcal{N}$ with sufficiently many pricing experiments. To see that, for each price, the corresponding demand $\boldsymbol{x}^{\boldsymbol{C}}(\boldsymbol{p})$ must be in an $\epsilon-$ball centered at a point $\boldsymbol{x}_c$ in $\mathcal{N}$ by the ball-covering theory in Lemma 1. We say the ball centered at a point $\boldsymbol{x}_c$ in $\mathcal{N}$ is "collected" if the point $\boldsymbol{x}_c$ is closest to the demand $\boldsymbol{x}^{\boldsymbol{C}}(\boldsymbol{p})$ by an experimented price $\boldsymbol{p}$. Suppose the ball centered at a point $\boldsymbol{x}_c$ is collected with probability q_c, then by performing the pricing experiments sufficiently many times (say T), we can expect to collect each of the balls centered at $\boldsymbol{x}_c$ in $\mathcal{N}$ at least once by the analysis of a coupon collector's problem (cf. Ferrante and Saltalamacchia (2014)). In particular, when $q_c = 1/N$—i.e., when each ball is collected with equal probability in an experiment—the number of pricing experiments needed is $T \sim O(N\log(N)) = O\left(\lceil 1/\epsilon^2\rceil (n+1)^{\lceil 1/\epsilon^2\rceil}\log(n+1)\right)$ (cf. Ferrante and Saltalamacchia (2014)), which is a polynomial number in the number of products n. If q_c is different for different balls, the number of experiments needed to cover every vector in $\mathcal{N}$ is $T(\boldsymbol{q}) := \sum_{i=1}^{N}\frac{1}{q_i} - \sum_{i<j}\frac{1}{q_i+q_j} + \cdots + (-1)^{N+1}\frac{1}{q_1+\cdots+q_N}$, which is at least the number needed in the uniform case (cf. Corollary 2 in Section 3.4 in Ferrante and Saltalamacchia (2014)).

Consider a set of $T \geq T(\boldsymbol{q})$ pricing experiments (with random prices). Our problem therefore reduces to the following: Find $\bar{\boldsymbol{C}}'$ such that

$$\sum_{t=1}^{T}\sum_{j=1}^{n}\left(p_{tj} - (-\bar{C}'_j(x_{tj}) + \bar{C}'_0(x_{t0})\right)^2 = 0. \tag{9}$$

THEOREM 4. *If a Lipschitz continuous function $\bar{\boldsymbol{C}}' \in \mathcal{M}$ (with a Lipschitz constant D) solves (9) for sufficiently many price experiments (large enough T), under Assumptions 1 and 2, $\bar{\boldsymbol{C}}'$ is a constant shift of $\boldsymbol{C}'$, within a threshold of $O(\epsilon)$, for all x in the unit simplex. In particular, when the demands are uniformly generated, the number of price experiments needed is $T \sim O\left(\lceil 1/\epsilon^2\rceil (n+1)^{\lceil 1/\epsilon^2\rceil}\log(n+1)\right)$.*

Theorem 4 indicates that with sufficiently many price experiments, we can identify $\boldsymbol{C}'$ up to a constant shift within ϵ. The sample size needed depends on how the price experiment is conducted. As shown earlier, if the price experiments lead to uniformly distributed demand, only a polynomial number (in the number of products) of experiments are needed.

Last, we examine the issue when there is no solution to the identification condition (9).

Estimation Problem When there are noise in the demand data obtained, there may not be solution to the identification condition (9) with observed demand data. Hence a natural idea is to consider an estimation problem that minimizes deviation from the identification condition.

Note that the objective in the pricing problem can be written as

$$\sum_{j=1}^{n} -w_j x_j - \frac{1}{\alpha}\sum_{j=1}^{n} x_j(C_j'(x_j) - C_0'(x_0)).$$

If we are able to identify $\bar{C}(\cdot)$ such that

$$\sum_{j=1}^{n}\left(\bar{C}_j'(x_j) - \bar{C}_0'(x_0) - \left(C_j'(x_j) - C_0'(x_0)\right)\right)^2 \le \eta, \tag{10}$$

for all $\boldsymbol{x}$ in the simplex, then by Cauchy-Schwarz inequality, the term $\left(\sum_{j=1}^{n} x_j(\bar{C}_j'(x_j) - \bar{C}_0'(x_0)) - \sum_{j=1}^{n} x_j(C_j'(x_j) - C_0'(x_0))\right)^2$ is bounded above by

$$\left(\sum_{j=1}^{n} x_j^2\right)\left(\sum_{j=1}^{n}\left(\bar{C}_j'(x_j) - \bar{C}_0'(x_0) - \left(C_j'(x_j) - C_0'(x_0)\right)\right)^2\right) \le \eta.$$

Hence the pricing problem with objective function

$$\sum_{j=1}^{n} -w_j x_j - \frac{1}{\alpha}\sum_{j=1}^{n} x_j(\bar{C}_j'(x_j) - \bar{C}_0'(x_0))$$

is within $O(\sqrt{\eta})$-neighborhood of the true objective function.

To see how we can achieve (10), we minimize the worst-case deviation measured by l_1 norm from the identification conditions in the experiments as follows:

$$\min_{\bar{C}'(\cdot)\in M_I} \max_{t=1,\ldots,T}\left\{\sum_{j=1}^{n}\left|p_{tj} - (-\bar{C}_j'(\hat{x}_{tj}) + \bar{C}_0'(\hat{x}_{t0}))\right|\right\}, \tag{11}$$

where p_{tj} and $\hat{x}_{tj}$ denote the observed price and noisy observation of the demand for product j in epoch t; i.e., $\hat{x}_{tj} = x_{tj} + \epsilon_{tj}$. The loss function is defined as the worst-case deviation from the identification conditions in (7) among the experimental prices. We use l_1 norm for the sake of tractability, which leads to a linear program for the estimation problem. It is worth mentioning that a similar idea has been used by Keshavarz et al. (2011) to compute the imputed convex objective from a set of samples with input $\boldsymbol{p}$ (price) and the corresponding decision $\hat{\boldsymbol{x}}$ (noisy demand) by minimizing a convex penalty function for deviation away from the optimality conditions. However, Keshavarz et al. (2011) assume a parametric form in basis functions, whereas we consider the candidates in a set of monotonic increasing functions, denoted by M_I, without any parametric assumptions on the candidate functions.

Note that from the optimality condition connecting $\boldsymbol{p}$ and its theoretical demand $\boldsymbol{x}$ in (5), we can write $p_{tj} - (-\bar{C}_j'(\hat{x}_{tj}) + \bar{C}_0'(\hat{x}_{t0}))$ as

$$\begin{aligned}
&-C_j'(x_{tj}) + C_0'(x_{t0}) - (-\bar{C}_j'(\hat{x}_{tj}) + \bar{C}_0'(\hat{x}_{t0})) \\
&\quad = -C_j'(x_{tj}) + C_0'(x_{t0}) - (-C_j'(\hat{x}_{tj}) + C_0'(\hat{x}_{t0})) + (-C_j'(\hat{x}_{tj}) + C_0'(\hat{x}_{t0})) - (-\bar{C}_j'(\hat{x}_{tj}) + \bar{C}_0'(\hat{x}_{t0}))
\end{aligned} \tag{12}$$

By a slight abuse of notation, assume $\bar{C}'(\cdot)$ is the optimal solution to (11). In this case, since the true solution $C'(\cdot)$ is a feasible solution, we have

$$\max_{t=1,\dots,T}\sum_{j=1}^{n}\left|p_{tj}-(-\bar{C}'_j(\hat{x}_{tj})+\bar{C}'_0(\hat{x}_{t0}))\right|\le\max_{t=1,\dots,T}\sum_{i=1}^{n}\left|p_{tj}-(-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0}))\right|.$$

Together with (12) and the triangle inequality $|a-b|\le|a|+|b|$, it follows that for any $t=1,\dots,T$,

$$\sum_{j=1}^{n}\left|\left(-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0})\right)-\left(-\bar{C}'_j(\hat{x}_{tj})+\bar{C}'_0(\hat{x}_{t0})\right)\right|\le 2\max_{t=1,\dots,T}\sum_{i=1}^{n}\left|-C'_j(x_{tj})+C'_0(x_{t0})-(-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0}))\right|.$$

We bound the term on the right using Assumption (3) and the Lipschitz continuity of $C'_j(\cdot)$ to obtain

$$\sum_{j=1}^{n}\left|-C'_j(x_{tj})+C'_0(x_{t0})-\left(-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0})\right)\right|\le 2nD\epsilon_S,$$

for all $t=1,\dots,T$, which implies

$$\sum_{j=1}^{n}\left|-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0})-\left(-\bar{C}'_j(\hat{x}_{tj})+\bar{C}'_0(\hat{x}_{t0})\right)\right|\le 4nD\epsilon_S,\text{for } t=1,\dots,T. \tag{13}$$

From the elementary inequality $\|x\|_2\le\|x\|_1$, we have

$$\sqrt{\sum_{j=1}^{n}\left(-C'_j(\hat{x}_{tj})+C'_0(\hat{x}_{t0})\right)-\left(-\bar{C}'_j(\hat{x}_{tj})+\bar{C}'_0(\hat{x}_{t0})\right)^2}\le 4nD\epsilon_S,\text{for } t=1,\dots,T. \tag{14}$$

In other words, (10) is satisfied for the points in the experiments, with $\eta=16n^2D^2\epsilon_S^2$.

Similar to the argument in Theorem 4, we can generalize the analysis to any $\boldsymbol{x}$ in the simplex by projecting $\hat{\boldsymbol{x}}_t$ in the experiment to the center of each of the balls covering the simplex with sufficiently large T. Hence, by solving (11) with sufficiently large T, we obtain a good estimate of the objective function in the pricing problem, as established in Theorem 5.

THEOREM 5. *If a Lipschitz continuous function $\bar{\boldsymbol{C}}'\in\mathcal{M}_I$ (with a Lipschitz constant D) solves (11) to optimality for sufficiently many price experiments (large enough T), under Assumptions 1 to 3, the objective function in the pricing problem from $\bar{\boldsymbol{C}}'$ is within $4D\sqrt{2n(n\epsilon_S^2+\epsilon^2)}-$neighborhood of the true objective function for all $\boldsymbol{x}$ in the unit simplex.*

Theorem 5 indicates that the optimal solution to the proposed estimation model (11) can provide a good approximation to the pricing problem despite the location/shape error in the identified perturbation functions. The error depends on the data quality measured by ϵ_S and the number of price experiments T, which affects ϵ. In particular, we show in Corollary 1 that the location error in the identified perturbation function does not affect the pricing problem.

COROLLARY 1. *If the identified function is a constant shift from the true function, the corresponding price optimization problem generates the true optimal price.*

The corollary directly follows from the analysis of Theorem 5. If the identified function is a constant shift from the true function, i.e., $\bar{C}'_j(x_j) = C'_j(x_j) + \Delta$ for $j = 0, \ldots, n$ and $x_j \in (0,1)$, (10) is satisfied for all $\boldsymbol{x}$ in the simplex, with $\eta = 0$. Hence the corresponding objective function in the pricing problem is the same as the true one. Corollary 1 indicates that the location error in the identification problem does not affect the pricing solution. In other words, although there are multiple solutions to the identification problem, they all lead to the same optimal demand and optimal prices as long as they are of a constant shift from the true function. This requires that the experiments be suitably designed and the price sample size be sufficiently large, as analyzed earlier.

Last, the model (11) provides a novel perspective on choice estimation. In most of the literature on estimating choice models from aggregate demand data (cf. Bell (1995)), a regression model is applied but the loss function is commonly defined as the deviation between the noisy observation $\hat{x}_{tj}$ and the theoretical demand $x_j(\boldsymbol{p}_t)$. In contrast, our estimation model is designed not only to predict what will happen but also to take into account the impact of prediction $x_j(\boldsymbol{p}_t)$ in the higher-level price optimization problem.

4.2. Price Optimization

Recall that our pricing problem is to find $\boldsymbol{x}$ to maximize the objective

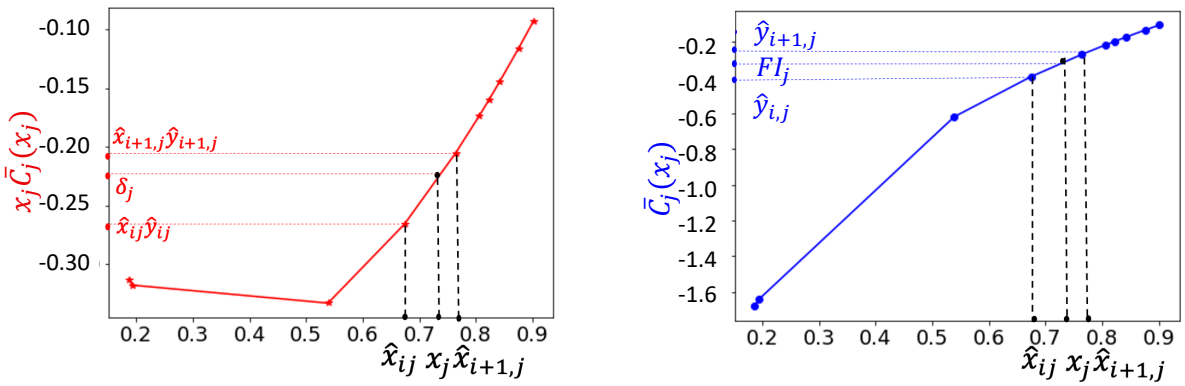
$$\sum_{j=1}^{n} -w_j x_j - \frac{1}{\alpha}\sum_{j=1}^{n} x_j C'_j(x_j) + \frac{1}{\alpha}(1-x_0)C'_0(x_0).$$

Solving the estimation model (11) provides the fitted value of $C'_j(\cdot)$ at each sampled point $\hat{x}_{tj}$. We use $\hat{y}_{tj}$ to represent the output from the estimation model, which is $\bar{C}'_j(\hat{x}_{tj})$. We use these values to construct a piecewise linear approximation to $C'_j(x)$ to optimize the prices of the products. As illustrated in Figure 2, for any x_j, it can be represented as a convex combination of two adjacent data points, e.g., $x_j = \lambda\hat{x}_{ij} + (1-\lambda)\hat{x}_{i+1,j}$. Then $C'_j(x_j)$ is approximated by $\lambda\hat{y}_{ij} + (1-\lambda)\hat{y}_{i+1,j}$. Similarly, $x_j C'_j(x_j)$ is approximated by $\lambda\hat{x}_{ij}\hat{y}_{ij} + (1-\lambda)\hat{x}_{i+1,j}\hat{y}_{i+1,j}$. We use the same approach to obtain a piecewise linear approximation of $(1-x_0)C'_0(x_0)$.

To formally formulate the pricing model based on the estimation results, we denote the indices of the sorted (in the ascending order) data set for a given j for $j = 0, \ldots, n$ as $\boldsymbol{r}^j = (r^j_1, \ldots, r^j_T)$. Let $\hat{\boldsymbol{x}}$ be a set of breakpoints, and $\hat{\boldsymbol{f}}$[7] the corresponding function values. For each point x, we are interested

[7] This can be $x\hat{C}'_j(x)$ or $\hat{C}'_j(x)$, depending on the data and setting.

Figure 2 Illustration of Pricing Model (15)



in recovering the corresponding function value at x, using a piecewise linear approximation of the points $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{f}})$. To this end, we define $PF\left(x; \hat{\boldsymbol{x}}; \hat{\boldsymbol{f}}\right)$ as follows:

$$PF\left(x; \hat{\boldsymbol{x}}; \hat{\boldsymbol{f}}\right) = \left\{ y = \sum_{t=1}^{T} \lambda_t \hat{f}_t \;\middle|\; \exists(\boldsymbol{\lambda}, \boldsymbol{z}) : \begin{array}{l} \lambda_1 \le z_1, \quad \lambda_t \le z_{t-1} + z_t, \forall t = 2, \ldots, T-1 \\ \lambda_M \le z_{T-1}, \; z_t \in \{0,1\}, \forall t = 1, \ldots, T-1, \\ \sum_{t=1}^{T} \lambda_t = 1, \quad \lambda_t \ge 0, \forall t = 1, \ldots, T, \\ \sum_{t=1}^{T-1} z_t = 1, \quad x = \sum_{t=1}^{T} \lambda_t \hat{x}_t, \end{array} \right\}.$$

The intuition behind the constraints is as follows: For a point (x, y) on the piecewise linear curve, the y value is the same convex combination of the $\hat{f}_t$ values as the x value is a convex combination of the $\hat{x}_t$ values. Specifically, if $x = \lambda_t \hat{x}_t + \lambda_{t+1}\hat{x}_{t+1}$, with $\lambda_t + \lambda_{t+1} = 1$, $\lambda_t, \lambda_{t+1} \ge 0$, then $y = \lambda_t \hat{f}_t + \lambda_{t+1}\hat{f}_{t+1}$. The binary variable $\boldsymbol{z}$ is introduced to indicate the interval in which the point is located. Thus, $PF\left(x; \hat{\boldsymbol{x}}; \hat{\boldsymbol{f}}\right)$ is a singleton, whose value denotes a piecewise approximation of the y value at x. Define $\circ$ as the element-wise product between two vectors. Let Ω denote the set of side constraints on the pricing solution. Then the price optimization problem can be formulated in

the following manner:

$$
\begin{aligned}
\Pi := \max_{\boldsymbol{x},\boldsymbol{\delta},\boldsymbol{FI}} \; & -\sum_{j=1}^{n} w_j x_j - \sum_{j=1}^{n} \delta_j + \delta_0 \\
\text{s.t. } & \delta_j \in PF\left(x_j; \hat{\boldsymbol{x}}_{r^j}; \hat{\boldsymbol{x}}_{r^j} \circ \hat{\boldsymbol{y}}_{r^j}\right), && \forall j = 1,\ldots,n, \\
& \delta_0 \in PF\left(x_0; \hat{\boldsymbol{x}}_{r^0}; (\mathbf{1} - \hat{x}_{r^0}) \circ \hat{\boldsymbol{y}}_{r^0}\right) \\
& FI_j \in PF\left(x_j; \hat{\boldsymbol{x}}_{r^j}; \hat{\boldsymbol{y}}_{r^j}\right), && \forall j = 0,\ldots,n, \\
& (-FI_1 + FI_0, \ldots, -FI_n + FI_0) \in \Omega, \\
& \sum_{j=0}^{n} x_j = 1 \\
& x_j \le \hat{x}_{r_T^j j}, && \forall j = 0,\ldots,n, \\
& x_j \ge \hat{x}_{r_1^j j}, && \forall j = 0,\ldots,n, \\
& \boldsymbol{x} \ge \mathbf{0}.
\end{aligned}
\tag{15}
$$

The first constraint provides a piecewise linear approximation of the function $x_j C_j'(x_j)$ at x_j (see the left figure in Figure 2), and the second constraint provides the approximation of $(1-x_0)C_0'(x_0)$ at x_0. Similarly, function inverse value $FI_j, j = 0,\ldots,n$ provides the approximation of $C_j'(x_j)$ at x_j (see the right figure in Figure 2). Hence the optimal price p_j can be represented as $-FI_j + FI_0$. We encapsulate all of the price constraints in Ω. For example, this may include bound constraints on the prices; e.g., $\underline{u}_j \le p_j \le \bar{u}_j$ or $p_i \le p_j$. Finally, we limit x to lie within the range of the data, since we have no additional information beyond the range unless we make some additional assumptions. When the set Ω is described through linear and integer constraints, this problem is solvable as a mixed integer linear program. Optimizing piecewise linear approximation to the concave function is known to approximate the separable concave optimization well (cf. Magnanti and Stratila (2012)).

It is worth observing that if Ω only includes nonnegativity constraints, the optimization model can be further simplified as a linear program by taking advantage of the convexity of the objective function. Notice that for a convex function, a linear approximation of the function value at point x_j can be calculated as

$$
\max_t \;\; \hat{x}_{r_t^j j}\hat{y}_{r_t^j j} + (\hat{x}_{r_{t+1}^j j}\hat{y}_{r_{t+1}^j,j} - \hat{x}_{r_t^j j}\hat{y}_{r_t^j j})/(\hat{x}_{r_{t+1}^j,j} - \hat{x}_{r_t^j j})(x_j - \hat{x}_{r_t^j j})
$$

and for a concave function, the corresponding approximated value at point x_0 is obtained from

$$
\min_t \;\; (1-\hat{x}_{r_t^0 0})\hat{y}_{r_t^0 0} + \left((1-\hat{x}_{r_{t+1}^0 0})\hat{y}_{r_{t+1}^0,0} - (1-\hat{x}_{r_t^0 0})\hat{y}_{r_t^0 0}\right)/(\hat{x}_{r_{t+1}^0,0} - \hat{x}_{r_t^0 0})(x_0 - \hat{x}_{r_t^0 0}).
$$

Therefore, the optimization problem can be modeled as the following linear programming problem:

$$
\begin{aligned}
\max_{\boldsymbol{x},\boldsymbol{\delta}} \; & -\sum_{j=1}^{n} w_j x_j - \sum_{j=1}^{n} \delta_j + \delta_0 \\
\text{s.t. } & \delta_j \geq \hat{x}_{r_t^j j}\hat{y}_{r_t^j j} + \frac{\hat{x}_{r_{t+1}^j j}\hat{y}_{r_{t+1}^j,j} - \hat{x}_{r_t^j j}\hat{y}_{r_t^j j}}{\hat{x}_{r_{t+1}^j,j} - \hat{x}_{r_t^j j}}(x_j - \hat{x}_{r_t^j j}), \forall t = 1,\ldots,T-1, j = 1,\ldots,n, \\
& \delta_0 \leq (1-\hat{x}_{r_t^0 0})\hat{y}_{r_t^0 0} + \frac{(1-\hat{x}_{r_{t+1}^0 0})\hat{y}_{r_{t+1}^0,0} - (1-\hat{x}_{r_t^0 0})\hat{y}_{r_t^0 0}}{\hat{x}_{r_{t+1}^0,0} - \hat{x}_{r_t^0 0}}(x_0 - \hat{x}_{r_t^0 0}), \forall t = 1,\ldots,T-1, \\
& \sum_{j=0}^{n} x_j = 1 \\
& x_j \leq \hat{x}_{r_T^j j}, \forall j = 0,\ldots,n, \\
& x_j \geq \hat{x}_{r_1^j j}, \forall j = 0,\ldots,n, \\
& \boldsymbol{x} \geq \mathbf{0}.
\end{aligned}
\tag{16}
$$

Finally, we show that if the identified function is a constant shift from $\boldsymbol{C}'(\cdot)$ at each sample point, i.e., $\bar{C}'_j(\hat{x}_{tj}) = C'_j(\hat{x}_{tj}) + \Delta$ for a constant Δ for $j = 0,\ldots,n$ and $t = 1,\ldots,T$, the pricing solution to the price optimization model is unique despite multiple admissible solutions from the identification problem.

Proposition 2. *If the identified function is a constant shift from the true function at each sample point, the corresponding MIP model (15) generates the same optimal demand $\boldsymbol{x}^*$, optimal price $\boldsymbol{p}^*$ and optimal profit Π^* as the piecewise linear approximation model (16) for the true pricing model.*

Proposition 2 indicates that the quality of the pricing solution from the proposed approach only depends on the performance of the piecewise linear approximation in the true pricing model if the identified function is a constant shift from the true function at each sample point. On the other hand, Magnanti and Stratila (2012) demonstrate that piecewise linear approximation approximates the separable concave optimization well with the number of pieces polynomial in the input size of the polyhedron formed by the constraints in Ω.

4.3. Validation Model

The proposed approach also allows us to perform validation by dividing the data into two sets: in-sample data and out-sample data. We first apply estimation model (11) to the in-sample data to get pointwise estimates of the perturbation functions in SRCM. After that, based on the estimates,

we solve the following validation model (17) to predict the demand under each price $\boldsymbol{p}^0$ in the out-sample data set:

$$
\begin{aligned}
Err(\boldsymbol{p}^0) := \min_{\boldsymbol{x},\boldsymbol{FI},\boldsymbol{\phi}} \; & \sum_{j=1}^{n} \phi_j \\
\text{s.t.} \; & -FI_j + FI_0 - p_j^0 \le \phi_j, && \forall j = 1,\dots,n, \\
& -(-FI_j + FI_0 - p_j^0) \le \phi_j, && \forall j = 1,\dots,n, \\
& FI_j \in PF\left(x_j; \hat{\boldsymbol{x}}_{rj}; \hat{\boldsymbol{y}}_{rj}\right), && \forall j = 0,\dots,n, \\
& \sum_{j=0}^{n} x_j = 1 \\
& x_j \ge 0, && \forall j = 0,\dots,n.
\end{aligned}
\tag{17}
$$

The main idea of prediction model (17) is to find the $\boldsymbol{x}$ whose corresponding price under the fitted piecewise linear function is closest to the observed price in l_1 norm, i.e., leading to the smallest deviation from identification condition (7) in l_1 norm. Here we use l_1 norm for the sake of tractability, under which the validation model becomes a mixed-integer linear program.

5. Computational Experiments

We compare our framework with three widely used approaches for pricing problems—pricing with consumer valuation samples, pricing with discrete choice models, and pricing with elasticity models—using a variety of data sets. The comparison is made on both synthetic data and industry data with the format of aggregate and normalized demand. We use MDM to denote the proposed data-driven pricing approach in the computational experiments. We use *CBC* solver for the linear programs and mixed-integer linear programs and *CVXOPT* for nonlinear convex programs in the numerical studies.

5.1. Pricing with Customer Valuation Samples

In this section, we test the MDM method on a data set with known customers' valuation of each product. Performance is evaluated by comparing with a well-established MIP approach proposed by Hanson and Martin (1990) for pricing with valuations. We also fit the data to an MNL choice model, and solve for the corresponding pricing problem.

Experiments:

- We sample the customers' valuation of each product from either a uniform distribution or a log-normal distribution independently. We use ν to denote the proportion of products with log-normal distributed valuations.
- In a population of 1,000 customers, we assume that we have access to the valuation of 100 customers (10% of the population) and solve a related MIP model (cf. Hanson and Martin (1990);

see Appendix C.1 for details) to get the optimal pricing solution. We use this as one benchmark in the study.

- For the MNL model, we first sample prices uniformly around the mean valuation of each product in the market, and obtain the corresponding market share (i.e., choice probability) of each product from the valuation of the 1,000 customers in the population. We solve the pricing problem for MNL using standard approach. In particular, we fit the price and demand data via an MNL estimation model (39) and then obtain another set of prices (denoted by MNL) by solving the pricing with the MNL model (37).
- We next apply the estimation model (11) to fit the perturbation functions and then solve the pricing model (15) to get a pricing solution, denoted by MDM.

We refer the readers to Appendix D for details in the numerical implementation. We perform the experiments 50 times, with new data generated in each round, to obtain a stable estimate of the expected performance of each method. We further test different combinations of products whose valuations are generated from the uniform and log-normal distributions and use ν to denote the proportion of products with uniform distributed valuations. We vary ν in $\{0, 0.5, 1\}$ to test the impact of a different mix of uniform and log-normal distributions on performance. We also vary the number of pricing experiments performed (i.e., Price Size) to understand its impact on performance.

We normalize the profit obtained from each method, using the profit accrued from the MIP approach as a base unit, and compute the average of the normalized profits over 50 simulations. Figure 3 reports the comparison of different approaches with $n = 4$ products. It is worth mentioning that the general findings remain valid even when we increase n.

Findings:

- MDM method outperforms the other two methods, and that performance tends to increase with more pricing experiments, especially for the variant when ν is high, although the improvement diminishes after 5-10 experiments.
- The MIP approach performs well when $\nu = 0$ (valuations are drawn from uniform distributions only), and MDM performs only marginally better than the MIP approach in this case. When the valuations of some products are from log-normal distributions, the MDM approach consistently dominates the MIP approach. The MNL approach is dominated by both MIP and MDM approach.
- Table 1 reports the computational times for all three approaches. The computation times for the MDM and MNL method are reported using 100 pricing experiments. The MDM approach is computationally more efficient than the MIP approach.

Figure 3 Profit Ratio Comparison under Different Proportions of Log-normal Distributed Valuations

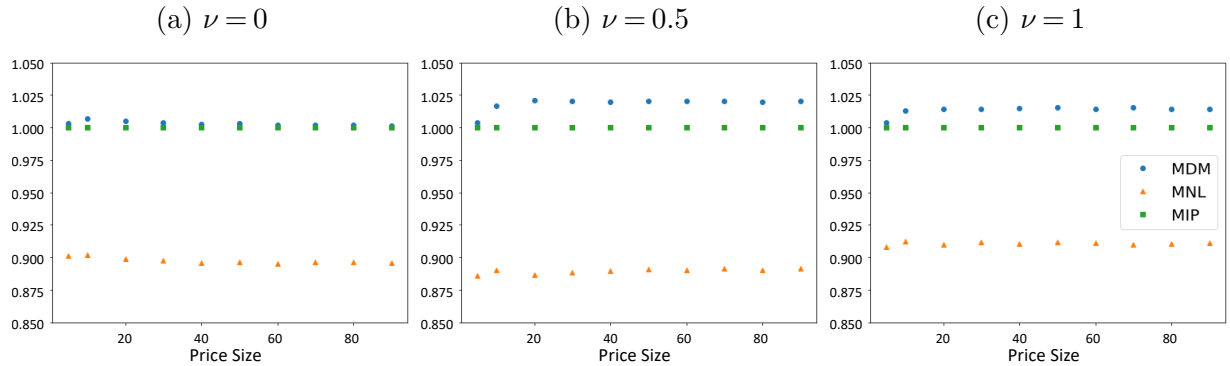


Table 1 Computation Time (seconds) Comparison

	MDM	MNL	MIP
Case 1($\nu = 0$)	0.4215	0.0995	181.8753
Case 2($\nu = 0.5$)	0.4289	0.0993	237.2261
Case 3($\nu = 1$)	0.4187	0.0996	326.5518

5.2. Case: Pricing with Automobile Data

We apply our method to a set of experimental data provided by an automobile manufacturer to demonstrate the benefits of the proposed method on practical pricing problems. We compare the method with two widely used approaches in practice: pricing with discrete choice models and pricing with elasticity models. We did not compare with the MIP approach as the valuation information is not provided and we only have aggregate market share data.

The manufacturer has developed an elaborate market simulator to evaluate the performance of different pricing proposals, taking into account the competitor's pricing strategy and the available outside options. In each simulation experiment, the company changes the price of one product to one of the four treatment levels, namely, -10%, -5%, 5%, and 10%. Then it simulates the corresponding market share of each product, including the outside option, using the simulator. There are 20 products in the pricing problem, and hence there are 81 experimental data in total, including the base price case. Base prices and cost, together with the brand information, are shown in Table 7 in Appendix E. There are three brands of products with 7 (and 7, 6) products in each brand A (and B, C, respectively). The current pricing solution yields a profit of 0.1822 million dollars, and the best solution from the 81 experiments yields a profit of 0.1844 million dollars.

The data include price and market share for each product as well as the outside market share. The underlying question is whether we can use the data to propose a better pricing solution for this problem. To evaluate the performance of the pricing solutions using various methods, we need to accurately predict the demand for given prices. Since we do not have access to the simulator to test

the performance of our recommended pricing solution, we apply various models—the discrete choice model, elasticity model, and SRCM model—to the data from the 81 experiments. We relegate the technical details of the estimation models (for MNL, NL, and elasticity models) using aggregate demand to Appendix C. It turns out that a price elasticity model can fit the data fairly well, and we will use it as the ground truth to perform counterfactual analysis for other pricing solutions.

In the following, we will first examine the prediction accuracy of the MDM approach vis-a-vis the method based on the MNL, NL, and elasticity model, then analyze the performance of the MDM pricing solution, assuming that the ground truth follows the price-elasticity model obtained.

5.2.1. Prediction Accuracy Since there are only 81 samples available, we validate the prediction accuracy of MDM in the following way. We use 80 data points (in-sample data) to calibrate the model via the estimation model (11), then predict demand under the remaining experiment (out-sample data) using the validation model (17). We repeat this experiment 50 times. In each experiment, the out-sample data are randomly selected. As a basis for comparison, we repeat the experiments, but now using (i) an MNL model, (ii) an NL model, and (iii) a price-elasticity model. We apply (39) and (43) to calibrate the MNL and NL models to get another set of predictions from the MNL and NL models, respectively. In particular, we use the brand information to categorize products into three different branches when applying the NL model. We further estimate the price elasticity coefficient at the base price via (34) and predict out-sample demand using the estimated elasticity coefficient.

We plot the boxplot of the prediction errors between the predicted demand and observed demand in out-sample under different methods in Figure 4, in which we can see that the elasticity-based approach predicts the demand fairly accurately, followed by our MDM approach. The latter provides a much better prediction than the MNL model and NL model. We evaluate the prediction error using two metrics. Specifically, we denote the profit at the predicted demand for the out-sample data as $\hat{\Pi}$ and the profit provided in the data as Π_0. We use the profit difference defined by $\frac{|\hat{\Pi}-\Pi_0|}{\Pi_0}$ and the mean square error (MSE) of the predicted demand to evaluate the prediction error in one validation experiment.

We report the prediction error using four estimation methods (MDM, MNL, NL, and the elasticity-based approach) in Table 2. We can see that the MDM method provides a better prediction than the MNL and NL models. It is surprising to see that the elasticity-based approach fits the data pretty well. In the following evaluation of the pricing solution, we will use the elasticity model as the ground truth to evaluate the performance of pricing solutions generated from different methods.

Figure 4 Predicted Demand for the 20 products in 50 Experiments

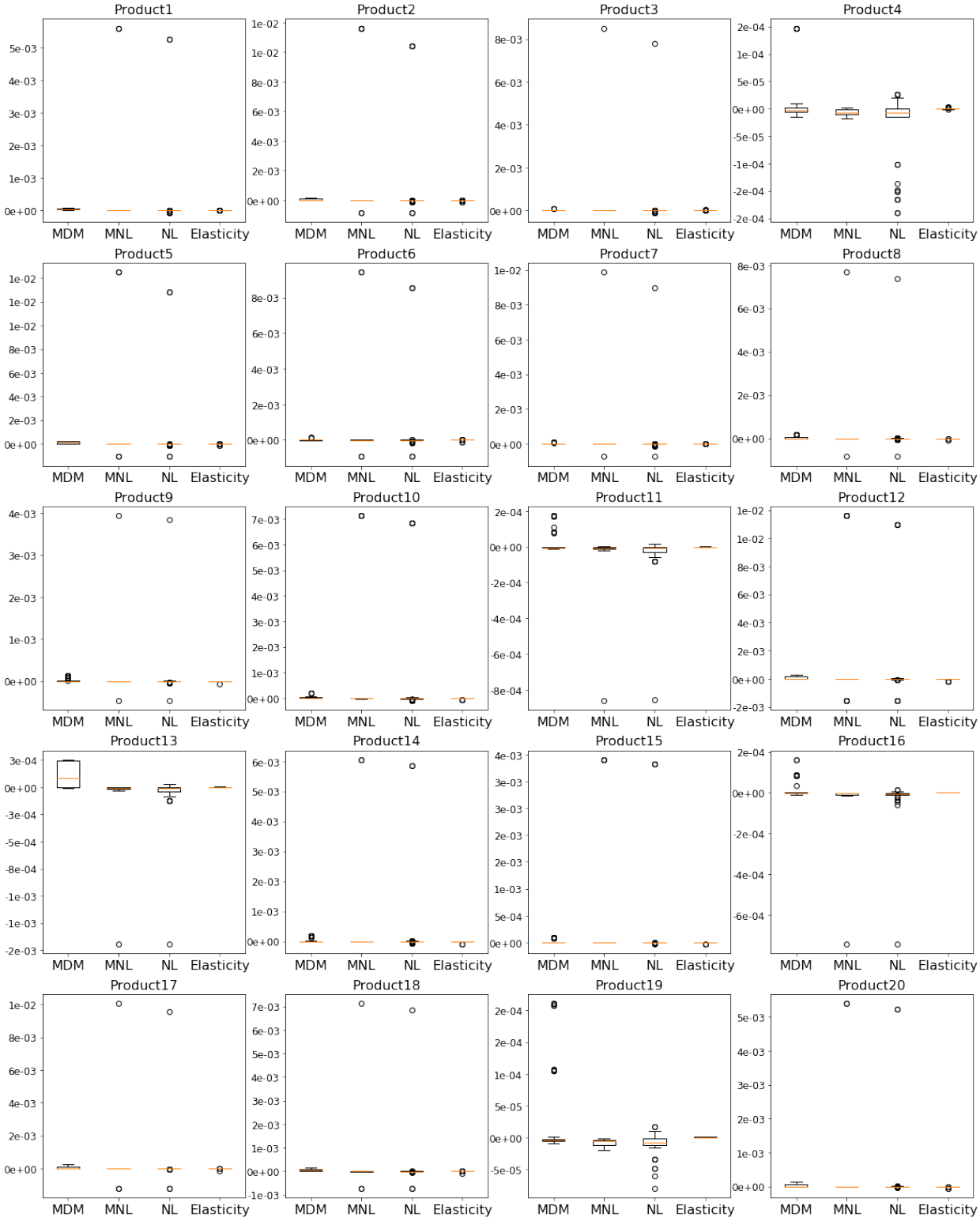


Table 2 Prediction Error in the Validation

$\frac{\lvert\hat{\Pi}-\Pi_0\rvert}{\Pi_0}$	MDM	MNL	NL	Elasticity
mean	0.0255	0.1652	0.1409	0.0012
std	0.0077	0.1974	0.1614	0.0018
MSE	**MDM**	**MNL**	**NL**	**Elasticity**
mean	3.4821E-08	4.4764E-06	3.6508E-06	4.9627E-10
std	2.0140E-08	5.3737E-06	4.2090E-06	6.9778E-10

5.2.2. Pricing Solution Building on the estimation result, we apply our MDM approach to get a new pricing proposal. We rule out significant changes from the current prices and impose a bound on the prices when doing the optimization. Specifically, suppose the price changes are required to be within 10% of the current prices—i.e., the pricing constraint set is supposed to be

$$\Omega_j = \left\{0.9p_j^c \leq p_j \leq 1.1p_j^c\right\},$$

where p_j^c denote the current price of product j. To incorporate these bounds on prices, we apply our MIP model (15) in Section 4.2 to get the optimal prices. We list the price obtained from different methods in Table 3. The profit generated from each pricing solution is evaluated under the elasticity model and is reported on the last row of Table 3. Note that we use the elasticity model as the ground truth; hence the true optimal pricing solution in the last column of Table 3 is obtained by solving the pricing with the elasticity model, as shown in (35). Also, we test the prices suggested by pricing with the fitted MNL and NL models (see (37) and (42), respectively). In particular, we relax the price constraints in Ω_j for $j = 1, \ldots, n$ when solving these two models to maintain model tractability. Unfortunately, even with the relaxed models, the MNL and NL approaches do not generate good pricing solutions. In general, we can see that the MDM method achieves 0.1951 million dollars in profit—a 7.08% improvement compared with the current pricing solution, and a 5.8% improvement compared with the best pricing solution from the 81 experiments. Compared with the MNL and NL models, it achieves 98.5% optimality versus 80% and 75.7% using the MNL and NL approaches, respectively. More interestingly, while we did not explicitly incorporate the brand feature in our model, the pricing strategy obtained by MDM suggests a minor adjustment around the base price for Brand A, and increasing the price for Brand B and Brand C. This shares a similar pattern with the true optimal pricing solution.

5.3. Case: Pricing with Fast-food Data

In this section, we consider a related application from a fast-food company that includes 53 products divided into 4 categories. The first category (Category A) has 14 types of entrées and 7 types of

Table 3 Optimal Prices (thousand)

Product	MDM	MNL	NL	Current Price	Experiment Best	True Opt
1	44.6076	39.3685	38.9086	45.5209	45.5209	44.7798
2	45.9235	40.4074	39.9475	46.9060	46.9060	47.0647
3	46.4188	39.0199	38.5600	45.0560	45.0560	44.5304
4	45.2370	38.1335	37.6737	43.8742	43.8742	44.5332
5	46.0395	40.4990	40.0391	47.0282	47.0282	47.3839
6	41.3047	36.7610	36.3011	42.0442	42.0442	43.0640
7	44.6091	39.3697	38.9099	45.5225	45.5225	45.1537
8	42.3051	40.7373	40.2186	39.4550	39.4550	42.7994
9	40.4121	39.3873	38.8686	37.9550	37.9550	41.0240
10	34.5448	35.0918	34.5730	33.1821	33.1821	35.9705
11	29.3931	29.7573	29.2386	27.2550	27.2550	29.7992
12	38.5948	37.5618	37.0431	35.9266	39.5193	39.0502
13	36.0233	35.4231	34.9044	33.5504	33.5504	36.4888
14	40.6655	39.3873	38.8686	37.9559	37.9559	41.1980
15	41.1127	39.0153	38.4802	39.7500	39.7500	41.8474
16	37.1717	35.6655	35.1304	35.8090	35.8090	37.8978
17	39.5902	37.7212	37.1860	38.2270	38.2270	41.0615
18	42.6127	40.29038	39.7552	41.2500	41.2500	43.9865
19	37.4239	35.8799	35.3447	36.0612	36.0612	38.3317
20	41.1127	39.0153	38.4802	39.7500	39.7500	42.0411
Profit (million)	**0.1951**	**0.1585**	**0.1500**	**0.1822**	**0.1844**	**0.1981**

small-sized Extra Value Meals (EVM). The second category (Category B) has 11 types of medium-sized EVM and 11 types of large-sized EVM. The third category (Category C) has 6 types of Happy Meals and the last category (Category D) has 4 types of sides. The prices of drinks are also provided as fixed parameters. Each consumer is assumed to choose at most one product within a category (or an outside option), but may choose more than one product across different categories. For each category, the company provided historical monthly sales data from April 2016 to March 2017, the prices of each product during this period, and an estimate of the outside market share. There are only four different sets of prices. Each product is labeled with nutrition information on calories, total fat, carbohydrates, and protein. Demographic information is also provided, including income and age.

Some business constraints required by the company are listed in Table 4, in which the base price refers to the price in the first period. While a majority of the constraints can be modeled in a straightforward manner using linear constraints, Constraint 2 involves integer variables.

Table 4 Price Constraints

1. Overall store price (weighted by quantity) increases by up to 2%.
2. Price change by increment of \$0.05
3. Price increase cap at min(\$0.30, round down the nearest whole number of 10% of original price). The absolute value of price decrease cap at min(\$0.20, round down the nearest whole number of 5% of original price).
4. EVM must be lower in price than sum of its components, including entrée, fries and drinks.
5. EVM must be higher in price than entrée + drinks.
6. Price per unit of fries must be decreasing with size.
7. Smaller-sized items must be cheaper than larger-sized items.
8. Small-sized EVM consists of entrée, small-sized fries and drinks.
9. Medium-sized EVM consists of entrée, medium-sized fries and drinks.
10. Large-sized EVM consists of entrée, large-sized fries and drinks.

We now apply the MDM approach to the price-demand data with four pricing experiments. The solution suggests keeping the price for small-sized EVM but increasing the price for medium and large-sized EVM. The adjustment of the entrée is around the base price. It also suggests a slight price increase for the items in Category C and the sides in Category D. With the adjustment, the profit is improved by 2.5% compared with the base price. We refer readers to Table 8 in Appendix E.2 for the price details of each product.

Note that the profit of the MDM pricing solution is evaluated under the calibrated MDM model. To justify its validity, we apply a validation procedure to test the predictive accuracy of the MDM approach. Since we have only four periods of data, we randomly choose one period as the out-sample data and calibrate the model using the other three data points. The prediction error is defined as the gap between the observed and the prediction and is reported in Table 5. From the table, we can see that the MDM model yields reasonably good predictions.

5.3.1. Explorations Based on Random Coefficient Logit Model The application to the fast-food data above demonstrates the good performance of the MDM approach in handling complicated practical constraints. We conduct more experiments in this section to understand how

Table 5 Summary of Prediction Error

	Category A	Category B	Category C	Category D
Error	9.997E-05	1.189E-05	3.146E-04	3.839E-04

well the MDM approach handles consumer heterogeneity. To this end, we compare next against a random coefficient logit choice model, based on the classical BLP approach (see Berry et al. (1995)).

Specifically, we apply the BLP algorithm to calibrate a random coefficient logit model from the existing data. Starting from the base price, we randomly generate a set of price increments that are multipliers of 0.05 and within the cap specified by Constraint 3 in Table 4. Together with the base price, we obtain a new set of prices. Notably, the generated prices might violate some constraints in Table 4. At each price, we apply the calibrated random coefficient logit model to obtain its demand. The analysis below is based on such a synthetic data set with 100 prices.

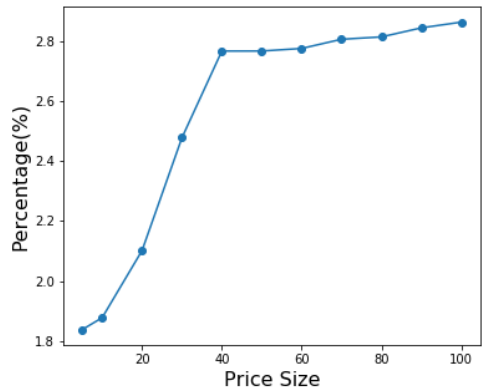
Predictive Accuracy We first investigate the predictive accuracy of the MDM method in a data set generated from the random coefficient logit model. To conduct the validation experiment, we use the first 80 data points as training data and the other 20 data points as test data. We report the mean and standard deviation of out-sample errors in Table 6. We can see that the out-sample errors are fairly small, which indicates that our method predicts the demand from a random coefficient logit model well.

Pricing Solution We apply the MDM method to the synthetic data with various price sizes and report the profit improvement over the base profit in Figure 5. Ten runs of simulations are conducted. Prices in each run are generated by a method similar to Algorithm 2, and hence they are nested across different simulations. The profit is evaluated under the calibrated random coefficient logit model, which is the ground truth model in the synthetic data. From the figure, we can see that with more price experiments, the MDM pricing solution tends to generate a higher profit while ensuring practical constraints. We further compare the MDM pricing solution with the best experimental price and the best feasible price in the experiment, respectively. The MDM pricing solution achieves an average improvement of 0.8% and 1.07% over the two pricing solutions in the experiment, respectively.

Table 6 Prediction Error (MSE) to Random Coefficient Logit Model

	Cat A	Cat B	Cat C	Cat D
Mean	1.2878E-06	1.2337E-05	4.1457E-06	1.3362E-04
Std	4.6056E-07	3.1391E-06	2.5990E-06	8.0316E-05

Figure 5 Profit Improvement over Base Profit with Various Price Sizes



In summary, the numerical studies indicate that the MDM approach can achieve a reasonably good pricing solution with both synthetic data and industry data. Applications to the fast-food data set further demonstrate that the framework works well for a heterogeneous consumer population and can handle complicated price constraints.

6. Concluding Remarks

We develop a data-driven approach for multi-product pricing problems using a representative consumer model in discrete choice. We characterize the conditions under which the pricing problem becomes tractable. In addition, the representative consumer model establishes a set of closed-form relationships between the prices and demand with an additive and separable perturbation function. The relationships provide conditions for identifying the perturbation functions up to a constant shift from the ground truth. This can be used to guide the search toward the optimal pricing solution, even if we have only a small number of experiments to test the effect of different pricing decisions. We use this approach on two sets of industry data to show that the framework produces reasonable price recommendations, even in the presence of consumer heterogeneity. This is a common concern in pricing with aggregate sales data. The empirical evidence suggests that our approach performs well even on small dataset.

This approach is related to recent work in online learning. Dong et al. (2018) propose an online method to learn the "best" customer utility function by fitting the sales data when the data are generated online. In a discrete choice context, the loss function proposed by Dong et al. (2018) corresponds to

$$l^r(\hat{\boldsymbol{x}}, \boldsymbol{p}, \boldsymbol{\theta}) = \min_{\boldsymbol{x}(\boldsymbol{p}) \in \mathcal{S}(\boldsymbol{\theta}, \boldsymbol{p})} ||\boldsymbol{x}(\boldsymbol{p}) - \hat{\boldsymbol{x}}||_2^2,$$

where $\hat{\boldsymbol{x}}$ is the observed demand (a potentially noisy estimate of the theoretical demand vector) at observed prices $\boldsymbol{p}$, and $\boldsymbol{x}(\boldsymbol{p})$ is the (unobservable) theoretical demand vector at price $\boldsymbol{p}$ where

$S(\boldsymbol{\theta}, \boldsymbol{p})$ specifies the optimality conditions the theoretical demand must satisfy at the observed pricing solution $\boldsymbol{p}$, with the parameter $\boldsymbol{\theta}$. While this problem is generally non-convex, Dong et al. (2018) show that under appropriate conditions on the optimal solution set $S(\boldsymbol{\theta}, \boldsymbol{p})$, it is possible to estimate parameters $\boldsymbol{\theta}^{(t)}$ in an online manner such that the average loss converges to the optimal solution:

$$\frac{1}{T}\sum_{t=1}^{T} l^r(\hat{\boldsymbol{x}}_t, \boldsymbol{p}_t, \boldsymbol{\theta}^{(t)}) \longrightarrow \min_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=1}^{T} l^r(\hat{\boldsymbol{x}}_t, \boldsymbol{p}_t, \boldsymbol{\theta}),$$

with a sublinear convergence rate. We compare the MDM method with other traditional approaches (offline and online) in Appendix F. In general, our method is able to produce better pricing solution using fewer pricing experiments, compared with current online methods.

This approach can also be strengthened and extended in various ways. For example, one of the limitations in the proposed approach is the assumption that the function $C(\boldsymbol{x})$ is additive and separable, so that the pricing problem can be solved using the proposed model. To extend this technique to handle a more general representative consumer model, we can use the following trick: Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ be i.i.d noise with mean 0. We can model the local structure of the function $\boldsymbol{C}(\boldsymbol{x})$ by looking at

$$\boldsymbol{C}(\boldsymbol{x}+\boldsymbol{\epsilon}) \approx \boldsymbol{C}(\boldsymbol{x}) + \sum_{i=1}^{n} \frac{\partial \boldsymbol{C}}{\partial x_i}\epsilon_i + \sum_{i,j} \frac{\partial^2 \boldsymbol{C}}{\partial x_i \partial x_j}\epsilon_i\epsilon_j,$$

and hence

$$\mathbf{E}\left[\boldsymbol{C}(\boldsymbol{x}+\tilde{\boldsymbol{\epsilon}})\right] \approx \boldsymbol{C}(\boldsymbol{x}) + \sum_{i} \frac{\partial^2 \boldsymbol{C}}{\partial x_i^2}\mathbf{E}\left[\tilde{\epsilon}_i^2\right].$$

Instead of assuming that the general function $\boldsymbol{C}(\cdot)$ is additive, we assume that the local structure of $\boldsymbol{C}(\cdot)$ at point $\boldsymbol{x}$ can be represented by an additive function, and perform pricing experiments around a small region of $\boldsymbol{x}$ to calibrate the corresponding demand function using the MDM approach. This allows us to optimize over the region and identify a search direction that will guide us to a better solution. An adaptive search procedure can be applied to heuristically generate samples to guide us in the search optimization. The main idea of the procedure is to first generate a relatively small-size sample from a large range of prices; estimate the model and optimize to get a set of optimal prices within the range; then generate another set of small-size samples around the obtained optimal price—but with a smaller interval for prices—to search for a better pricing solution. We test the value of the adaptive procedure on a synthetic data set generated by an NL model in Appendix G. The data-driven approach is able to generate a near-optimal pricing solution, in the case in which the underlying choice process follows an NL model that cannot be represented using additive and separable convex perturbation functions in a representative consumer model.

Acknowledgment

The research of the first author is partly supported by NTU start-up grant and MOE Academic Research Fund Tier 1 grant (No. 2021-T1-001-058). The research of the second author was partly supported by MOE Academic Research Fund Tier 2 grant T2MOE1706, 'On the Interplay of Choice, Robustness and Optimization in Transportation' and the SUTD-MIT International Design Center grant IDG21700101 on 'Design of the Last Mile Transportation System: What Does the Customer Really Want?'. The research of the third author is supported by MOE Academic Research Fund Tier 3 grant call (Award ref: MOE-2019-T3-1-010), and by A-STAR, under its IAF-ICP grant (Award ref: I2001E0059).

References

Akçay, Y., H. P. Natarajan, S. H. Xu. 2010. Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science* **56(8)** 1345–1361.

Allen, Roy, John Rehbeck. 2019. Identification with additively separable heterogeneity. *Econometrica* **87**(3) 1021–1054.

Alptekinoğlu, A., J. H. Semple. 2016. The exponomial choice model: A new alternative for assortment and price optimization. *Operations Research* **64(1)** 79–93.

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

Bagnoli, M., T. Bergstrom. 2005. Log-concave probability and its applications. *Economic Theory* **26(2)** 445–469.

Bell, M.G. 1995. Alternatives to dial's logit assignment algorithm. *Transportation Research Part B: Methodological* **29(4)** 287–295.

Ben-Akiva, Moshe, Steven R Lerman. 2018. *Discrete choice analysis: theory and application to travel demand*. Transportation Studies.

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

Bunch, David S. 1987. Maximum likelihood estimation of probabilistic choice models. *SIAM Journal on Scientific and Statistical Computing* **8**(1) 56–70.

Chen, Xi, Ilias Diakonikolas, Dimitris Paparas, Xiaorui Sun, Mihalis Yannakakis. 2018. The complexity of optimal multidimensional pricing for a unit-demand buyer. *Games and Economic Behavior* **110** 139–164.

Dong, Chaosheng, Yiran Chen, Bo Zeng. 2018. Generalized inverse optimization through online learning. *Advances in Neural Information Processing Systems*. 86–95.

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

Einav, L., T. Kuchler, J. Levin, N. Sundaresan. 2011. Learning from seller experiments in online markets. *NBER Working Paper No. 17385* .

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

Feng, Guiyun, Xiaobo Li, Zizhuo Wang. 2017. On the relation between several discrete choice models. *Operations research* **65**(6) 1516–1525.

Ferrante, Marco, Monica Saltalamacchia. 2014. The coupon collector's problem. *Materials matematics* 0001–35.

Fox, J. T. 2007. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *RAND Journal of Economics* **38(4)** 1002–1019.

Fudenberg, D., R. Iijima, T. Strzalecki. 2015. Stochastic choice and revealed perturbed utility. *Econometrica* **83(6)** 2371–2409.

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

Gallego, Guillermo, Huseyin Topaloglu, et al. 2019. *Revenue management and pricing analytics*, vol. 209. Springer.

Hanson, W., K. Martin. 1996. Optimizing multinomial logit profit functions. *Management Science* **42(7)** 992–1003.

Hanson, W., R. K. Martin. 1990. Optimal bundle pricing. *Management Science* **36(2)** 155–174.

Hofbauer, J., W. H. Sandholm. 2002. On the global convergence of stochastic fictitious play. *Econometrica* **70(6)** 2265–2294.

Huchette, Joey, Juan Pablo Vielma. 2017. Nonconvex piecewise linear functions: Advanced formulations and simple modeling tools. *arXiv preprint arXiv:1708.00050* .

Huh, W. T., H. Li. 2015. Technical note: Pricing under the nested attraction model with a multistage choice structure. *Operations Research* **63(4)** 840–850.

Keshavarz, Arezou, Yang Wang, Stephen Boyd. 2011. Imputing a convex objective function. *2011 IEEE International Symposium on Intelligent Control*. IEEE, 613–619.

Lewbel, Arthur. 1998. Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica* 105–121.

Li, G., P. Rusmevichientong, H. Toplagalou. 2015. The d-level nested logit model: Assortment and price optimization problems. *Operations Research* **63(2)** 325–341.

Li, H., W. T. Huh. 2011. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing and Service Operations Management* **13(4)** 549–564.

Li, Hongmin, Scott Webster. 2017. Optimal pricing of correlated product options under the paired combinatorial logit model. *Operations Research* **65**(5) 1215–1230.

Li, Hongmin, Scott Webster, Nicholas Mason, Karl Kempf. 2019. Product-line pricing under discrete mixed multinomial logit demand: Winner—2017 m&som practice-based research competition. *Manufacturing & Service Operations Management* **21**(1) 14–28.

Magnanti, Thomas L, Dan Stratila. 2012. Separable concave optimization approximately equals piecewise-linear optimization. *arXiv preprint arXiv:1201.3148* .

Manski, C. F. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* **3** 205–228.

Marandi, Ahmadreza, Virginie Lurkin. 2020. Static pricing problems under mixed multinomial logit demand. *arXiv preprint arXiv:2005.07482* .

Misener, R., C. Floudas. 2012. Global optimization of mixed-integer quadratically-constrained quadratic programs (miqcqp) through piecewise-linear and edge-concave relaxations. *Mathematical Programming* **136**(1) 155–182.

Mishra, V. K., K. Natarajan, D. Padmanabhan, C. P. Teo, X. Li. 2014. On theoretical and empirical aspects of marginal distribution choice models. *Management Science* **60(6)** 1511–1531.

Myerson, R. 1981. Optimal auction design. *Mathematics of Operations Research* **6(1)** 58–73.

Natarajan, Karthik, Miao Song, Chung-Piaw Teo. 2009. Persistency model and its applications in choice modeling. *Management Science* **55**(3) 453–469.

Shi, Xiaoxia, Matthew Shum, Wei Song. 2018. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica* **86**(2) 737–761.

Song, J. S., Z. L. Xue. 2007. Demand management and inventory control for substitutable products .

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

van de Geer, Ruben, Sandjai Bhulai. 2016. Optimal pricing under mixed logit choice. *arXiv preprint arXiv:1605.08340* .

Vershynin, Roman. 2018. *High-dimensional probability: An introduction with applications in data science*, vol. 47. Cambridge university press.

Zhang, H., P. Rusmevichientong, H. Topaloglu. 2018. Technical note - multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. *Operations Research* **66(6)** 549–564.

Appendix A: Proofs

Proof of Theorem 1

Proof: Denote $h_j(x) = -xC_j'(x), \forall j = 1,\ldots,n,\ h_0(x) = -(1-x)C_0'(1-x)$ and $g_j(x) = h_j'(x), \forall j = 0,\ldots,n$. Let μ denote the dual variable for the equality constraint in formulation (4). The optimality conditions of (4) yield

$$\begin{aligned} v_j - \alpha w_j + h_j'(x_j) + \mu &= 0,\ \forall j = 1,\ldots,n, \\ -h_0'(x_0) + \mu &= 0. \end{aligned} \tag{18}$$

Reorganize (18) to get

$$\begin{aligned} x_j &= g_j^{-1}(\alpha w_j - v_j - \mu),\ \forall j = 1,\ldots,n, \\ x_0 &= g_0^{-1}(\mu). \end{aligned} \tag{19}$$

Since at optimality, $\boldsymbol{x}$ also satisfies the feasibility condition $\sum_j x_j = 1$, we can calculate the optimal dual variable μ by solving the nonlinear equation

$$G(\mu) := \sum_{j=1}^{n} g_j^{-1}(\alpha w_j - v_j - \mu) + g_0^{-1}(\mu) = 1. \tag{20}$$

On the other hand, since $h_j(x)$ for $j = 1,\ldots,n$ is concave, $g_j(x)$ for $j = 1,\ldots,n$ is non-increasing in x. Similarly, the convexity of $h_0(x)$ implies that $g_0(x)$ is nondecreasing in x. Therefore, $G(\mu)$ is an increasing function in μ. Therefore, we can solve the nonlinear equation (20) by a line search method in the range

$$\left[\max\left\{\max_j\left\{-v_j + \alpha w_j - g_j(0)\right\}, g_0(0)\right\}, \min\left\{\min_j\left\{-v_j + \alpha w_j - g_j(1)\right\}, g_0(1)\right\}\right].$$

With the optimal dual variable μ, the optimal demand $\boldsymbol{x}$ is obtained from (19). Q.E.D.

Proof of Proposition 1

Proof: To prove the proposition, we first establish the following lemma.

Lemma 2. *Let $F_j(x)$ for $j = 0,\ldots,n$ be the marginal distributions. Then,*
(i) $xF_j^{-1}(1-x)$ is concave in x if and only if $\frac{1}{1-F_j(x)}$ is convex in x.
(ii) $xF_0^{-1}(x)$ is convex in x if and only if $\frac{1}{F_0(x)}$ is convex in x.

Proof of Lemma 2

For the smooth inverse function $F^{-1}(x)$, the derivative is given as $(F^{-1})'(x) = 1/F'(F^{-1}(x))$. Denote $h(x) := xF^{-1}(1-x)$. Then,

$$\begin{aligned} h'(x) &= F^{-1}(1-x) + x(F^{-1}(1-x))', \\ &= F^{-1}(1-x) - \frac{x}{F'(F^{-1}(1-x))}. \end{aligned} \tag{21}$$

Calculating the second derivatives,

$$\begin{aligned} h''(x) &= -\frac{2}{F'(F^{-1}(1-x))} - \frac{xF''(F^{-1}(1-x))}{(F'(F^{-1}(1-x)))^3}, \\ &= -\frac{2F'(F^{-1}(1-x))^2 + xF''(F^{-1}(1-x)))}{[F'(F^{-1}(1-x))]^3}. \end{aligned} \tag{22}$$

Let $y = F^{-1}(1-x)$, then $1-x = F(y)$. Since $F(\cdot)$ is cumulative distribution function, $F'(\cdot) \geq 0$. Therefore $h(x)$ is concave in $[0,1]$ if and only if

$$2F'(F^{-1}(1-x))^2 + xF''(F^{-1}(1-x)) = 2F'(y)^2 + (1-F(y))F''(y) \geq 0, \ \forall y. \tag{23}$$

On the other hand, we have

$$\left(\frac{1}{1-F(y)}\right)' = \frac{F'(y)}{(1-F(y))^2}, \tag{24}$$

$$\left(\frac{1}{1-F(y)}\right)'' = \frac{2F'(y)^2 + (1-F(y))F''(y)}{(1-F(y))^3}. \tag{25}$$

So $1/(1-F(y))$ is convex if and only if inequality (23) holds. Result (i) has been proved.

Similarly, denote $g(x) = xF^{-1}(x)$. Calculating the derivatives, we have

$$g'(x) = F^{-1}(x) + \frac{x}{F'(F^{-1}(x))}, \tag{26}$$

$$g''(x) = \frac{2F'(F^{-1}(x))^2 - xF''(F^{-1}(x))}{[F'(F^{-1}(x))]^3}. \tag{27}$$

Let $z = F^{-1}(x), x = F(z)$. Then function $g(x)$ is convex if and only if:

$$2F'(z)^2 - F(z)F''(z) \geq 0. \tag{28}$$

The function $\frac{1}{F(z)}$ is convex if and only if the second derivative is nonnegative, namely,

$$(\frac{1}{F(z)})'' = \frac{2F'(z)^2 - F(z)F''(z)}{F(z)^3} \geq 0,$$

which is equivalent to (28). Q.E.D.

Then we are ready to prove the proposition. When the tail distribution $\bar{F}_j(x)$ for $j = 1, \ldots, n$ is log-concave, we have

$$(\ln(1-F_j(y)))'' = -\frac{F_j'(y)^2 + (1-F_j(y))F_j''(y)}{(1-F_j(y))^2} \leq 0.$$

Hence,

$$0 \leq F_j'(y)^2 + (1-F_j(y))F_j''(y) \leq 2F_j'(y)^2 + (1-F_j(y))F_j''(y).$$

From the proof of Lemma 2, $xF_j^{-1}(1-x)$ is concave. (ii) can be shown using a similar procedure. The last statement follows from results in Bagnoli and Bergstrom (2005). The proposition holds due to the one to one correspondence between the marginal distribution $F_j(\cdot)$ and the perturbation function $C_j(\cdot)$ in (6). Q.E.D.

Proof of Theorem 3

From the analysis in the main text, we have $\bar{C}_j'(x_{tj}) = C_j'(x_{tj}) + \Delta$ for some constant Δ and $t \in \mathcal{N}$ and

$j=0,\ldots,n$ if $\bar{\boldsymbol{C}} \in \mathcal{M}$ solves (8). For any point $\boldsymbol{x}_1$ in the simplex, it is within ϵ neighborhood of a point in $\mathcal{N}$, denoted by $\boldsymbol{x}_2$, i.e., $\sqrt{\sum_{j=0}^{n}(x_{1j}-x_{2j})^2} \le \epsilon$.

$$\begin{aligned}
&|\bar{C}'_j(x_{1j}) - C'_j(x_{1j}) - \Delta| \\
=\ &\left|\bar{C}'_j(x_{1j}) - \bar{C}'_j(x_{2j}) + \bar{C}'_j(x_{2j}) - C'_j(x_{2j}) - \Delta + C'_j(x_{2j}) - C'_j(x_{1j})\right| \\
&\text{[add and subtract } \bar{C}'_j(x_{2j}) - C'_j(x_{2j}) \text{ and re-organize terms]} \\
\le\ &\left|\bar{C}'_j(x_{1j}) - \bar{C}'_j(x_{2j})\right| + \left|C'_j(x_{2j}) - C'_j(x_{1j})\right| \\
&\text{[Triangle inequality and } \bar{C}'_j(x_{tj}) = C'_j(x_{tj}) + \Delta \text{ for } t \in \mathcal{N}] \\
\le\ &D\,|x_{1j} - x_{2j}| + D\,|x_{2j} - x_{1j}| \\
&\text{[Lipschitz continuity]} \\
\le\ &D\epsilon + D\epsilon.
\end{aligned}$$

The last inequality holds as $|x_{2j}-x_{1j}| = \sqrt{(x_{1j}-x_{2j})^2} \le \sqrt{\sum_{j=0}^{n}(x_{1j}-x_{2j})^2} \le \epsilon$. Q.E.D.

Proof of Theorem 4

(9) implies that $p_{tj} = (-\bar{C}'_j(x_j^{\boldsymbol{C}}(\boldsymbol{p}_t)) + \bar{C}'_0(x_0^{\boldsymbol{C}}(\boldsymbol{p}_t)))$ for all $t=1,\ldots,T$. Following the same analysis as in Theorem 2 we can imply that $\Delta_{tj} = \Delta_t$ for all $j=0,\ldots,n$ and $t=1,\ldots,T$. For ease of exposition, we use x_{tj} to denote $x_j^{\boldsymbol{C}}(\boldsymbol{p}_t)$ when the context is clear.

Consider any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Denote the nearest center of the $\epsilon-$balls covering $\boldsymbol{x}_t$ as $\boldsymbol{x}_t^c$ for $t=1,2$. If $\boldsymbol{x}_1^c = \boldsymbol{x}_2^c$, which means $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in the same ball, hence $\sqrt{\sum_{j=0}^{n}(x_{1j}-x_{2j})^2} \le \epsilon$, which implies $|x_{1l}-x_{2l}| \le \epsilon$ for any $j=0,\ldots,n$. From the Lipschitz continuity of $\bar{\boldsymbol{C}}'$, we have $|y_{2l}-y_{1l}| \le D|x_{2l}-x_{1l}| \le D\epsilon$, where $y_{tl} = \bar{C}'_j(x_{tl})$. From Assumption 2 we can also derive $|y^*_{2l}-y^*_{1l}| \le D|x_{2l}-x_{1l}| \le D\epsilon$, where $y^*_{tl} = C'_j(x_{tl})$. Therefore,

$$\begin{aligned}
|\Delta_1 - \Delta_2| &= |y^*_{1l} + \Delta_1 - y^*_{2l} - \Delta_2 - y^*_{1l} + y^*_{2l}| \\
&\le |y^*_{1l} + \Delta_1 - y^*_{2l} - \Delta_2| + |-y^*_{1l} + y^*_{2l}| \\
&= |y_{1l} - y_{2l}| + |-y^*_{1l} + y^*_{2l}| \\
&\le 2D\epsilon.
\end{aligned}$$

Next we consider the case in which $\boldsymbol{x}_1^c$ and $\boldsymbol{x}_2^c$ are two different points. Note that with sufficiently many experiments (a large enough T), demand vectors close to every vector in $\mathcal{N}$ are included. Following the same analysis as in deriving Theorem 3, we can find $\boldsymbol{x}_3^c \in \mathcal{N}$ such that a demand vector within its $\epsilon-$ball is included in the T demand vectors in the experiment, and $x^c_{3l} = x^c_{1l}$ for some product l, and $x^c_{3l'} = x^c_{2l'}$ for another l'. Hence,

$$\begin{aligned}
|x_{3l} - x_{1l}| &= |x_{3l} - x^c_{3l} + x^c_{3l} - x^c_{1l} + x^c_{1l} - x_{1l}| \\
&\le |x_{3l} - x^c_{3l}| + |x^c_{3l} - x^c_{1l}| + |x^c_{1l} - x_{1l}| \le \epsilon + \epsilon. \\
&\quad \text{[triangle inequality]} \qquad\qquad \text{[from the definition of } \boldsymbol{x}_t^c \text{ and } x^c_{3l} = x^c_{1l}]
\end{aligned}$$

Similarly, we can derive $|x_{3l'} - x_{2l'}| \le 2\epsilon$. From the Lipschitz continuity of $\bar{\boldsymbol{C}}'$, we have $|y_{3l} - y_{1l}| \le D|x_{3l} - x_{1l}| \le 2D\epsilon$ and $|y_{3l'} - y_{2l'}| \le D|x_{3l'} - x_{2l'}| \le 2D\epsilon$, where $y_{tl} = \bar{C}'_j(x_{tl})$ for $t=1,2,3$. From Assumption 2 we

can also derive $|y^*_{3l} - y^*_{1l}| \le D|x_{3l} - x_{1l}| \le 2D\epsilon$ and $|y^*_{3l'} - y^*_{2l'}| \le D|x_{3l'} - x_{2l'}| \le 2D\epsilon$, where $y^*_{tl} = C'_j(x_{tl})$ for $t = 1, 2, 3$. Hence,

$$\begin{aligned}|\Delta_1 - \Delta_3| &= |y^*_{1l} + \Delta_1 - y^*_{3l} - \Delta_3 - y^*_{1l} + y^*_{3l}| \\ &\le |y^*_{1l} + \Delta_1 - y^*_{3l} - \Delta_3| + |-y^*_{1l} + y^*_{3l}| \\ &= |y_{1l} - y_{3l}| + |-y^*_{1l} + y^*_{3l}| \\ &\le 4D\epsilon.\end{aligned}$$

Similarly, we can derive $|\Delta_2 - \Delta_3| \le 4D\epsilon$. Hence

$$\begin{aligned}|\Delta_1 - \Delta_2| &= |\Delta_1 - \Delta_3 + \Delta_3 - \Delta_2| \\ &\le |\Delta_1 - \Delta_3| + |\Delta_3 - \Delta_2|. \\ &\le 8D\epsilon\end{aligned}$$

In summary, at any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the shift of $\bar{\boldsymbol{C}}$ from $\boldsymbol{C}$ is bounded by $O(\epsilon)$. Consider a demand vector obtained from the experiment that is closet to a point in $\mathcal{N}$ as $\boldsymbol{x}_1$, at which the deviation of $\bar{\boldsymbol{C}}'$ from $\boldsymbol{C}'$ is Δ. Then the analysis above implies that for all the demand vectors $\boldsymbol{x}_t, t = 1, \ldots, T$ from the experiment, we have

$$|\bar{C}'_j(x_{tj}) - C'_j(x_{tj}) - \Delta| \le O(\epsilon).$$

Following a similar analysis as in Theorem 3, we can prove the statement from the Lipschitz continuity. Q.E.D.

Proof of Theorem 5

Similar to the argument in Theorem 4, define an $\epsilon-$ball is collected if its center $\boldsymbol{x}_c$ is closest to $\hat{\boldsymbol{x}}_t$. Hence

$$\begin{aligned}&\sum_{j=1}^{n} \left(-C'_j(x_{cj}) + C'_0(x_{c0}) - (-\bar{C}'_j(x_{cj}) + \bar{C}'_0(x_{c0})\right)^2 \\ \le\ &\sum_{j=1}^{n} 2\left(-C'_j(\hat{x}_{tj}) + C'_0(\hat{x}_{t0}) - (-\bar{C}'_j(\hat{x}_{tj}) + \bar{C}'_0(\hat{x}_{t0}))\right)^2 + \\ &\sum_{j=1}^{n} 2\left(-C'_j(x_{cj}) + C'_j(\hat{x}_{tj}) + C'_0(x_{c0}) - C'_0(\hat{x}_{t0}) - (-\bar{C}'_j(x_{cj}) + \bar{C}'_j(\hat{x}_{tj}) + \bar{C}'_0(x_{c0}) - \bar{C}'_0(\hat{x}_{t0})\right)^2 \\ &\left[\text{add and subtract } C'_j(\hat{x}_{tj}), C'_0(\hat{x}_{t0}), \bar{C}'_j(\hat{x}_{tj}) \text{ and } \bar{C}'_0(\hat{x}_{t0}) \text{ in each term, respectively, and } (a+b)^2 \le 2a^2 + 2b^2\right] \\ \le\ &32n^2D^2\epsilon_S^2 + 4 \times (2nD^2\epsilon^2 + 2nD^2\epsilon^2) = 16nD^2(2n\epsilon_S^2 + \epsilon^2) \\ &\left[\text{From (14) and Lipschitz Continuity of } C'_j(\cdot) \text{ and } \bar{C}'_j(\cdot) \text{ for } j = 0, \ldots, n, \text{ and } (a+b)^2 \le 2a^2 + 2b^2\right].\end{aligned} \tag{29}$$

With sufficiently many T, which allows to collect each of the balls at least once as analyzed earlier, we can imply that $\sum_{j=1}^{n} \left(-C'_j(x_{cj}) + C'_0(x_{c0}) - (-\bar{C}'_j(x_{cj}) + \bar{C}'_0(x_{c0})\right)^2 \le 16nD^2(2n\epsilon_S^2 + \epsilon^2)$ holds for each ball covering the simplex. Note that for any $\boldsymbol{x}$ in the simplex, it must be within $\epsilon-$ball of the center in one ball. Hence we can derive the following inequality using the same argument.

$$\sum_{j=1}^{n} \left(-C'_j(x_j) + C'_0(x_0) - (-\bar{C}'_j(x_j) + \bar{C}'_0(x_0)\right)^2 \le 32nD^2(n\epsilon_S^2 + \epsilon^2) \tag{30}$$

Hence (10) is satisfied for all the $\boldsymbol{x}$ in the simplex, with $\eta = 32nD^2(n\epsilon_S^2 + \epsilon^2)$.

Therefore, from the argument in the main text, the price problem with an objective function

$$\sum_{j=1}^{n} -w_j x_j - \frac{1}{\alpha}\sum_{j=1}^{n} x_j(\bar{C}'_j(x_j) - \bar{C}'_0(x_0))$$

is within $4D\sqrt{2n(n\epsilon_S^2+\epsilon^2)}$-neighbourhood of the true objective function. Q.E.D.

Proof of Proposition 2

Proof: Suppose an estimate $\bar{C}_j'(\cdot)$ is a constant shift from $C_j'(\cdot)$ at each sample point, i.e., $\bar{C}_j'(\hat{x}_{tj}) = C_j'(\hat{x}_{tj}) + \Delta$ for any j and t. Denote $\bar{C}_j'(\hat{x}_{tj})$ as $\hat{y}_{tj}$ and $C_j'(\hat{x}_{tj})$ as y_{tj}^*. Consider an arbitrary demand $\boldsymbol{x}^*$ and assume x_j^* is located between $\hat{x}_{r_{t_j}^j,j}$ and $\hat{x}_{r_{t_j+1}^j,j}$, for $j=0,1,\ldots,n$, and $x_j^* = \lambda_j^*\hat{x}_{r_{t_j}^j,j} + (1-\lambda_j^*)\hat{x}_{r_{t_j+1}^j,j}$. The profit function at an arbitrary demand $\mathbf{x}^*$ can be represented as follows.

$$
\begin{aligned}
&-\sum_{j=1}^{n}\left(\lambda_j\hat{x}_{r_{t_j}^j,j}\hat{y}_{r_{t_j}^j,j} + (1-\lambda_j)\hat{x}_{r_{t_j+1}^j,j}\hat{y}_{r_{t_j+1}^j,j}\right) + \left(\lambda_0(1-\hat{x}_{r_{t_0}^0,0})\hat{y}_{r_{t_0}^0,0} + (1-\lambda_0)(1-\hat{x}_{r_{t_0+1}^0,0})\hat{y}_{r_{t_0+1}^0,0}\right)\\
=&-\sum_{j=1}^{n}\left(\lambda_j\hat{x}_{r_{t_j}^j,j}y^*_{r_{t_j}^j,j} + (1-\lambda_j)\hat{x}_{r_{t_j+1}^j,j}y^*_{r_{t_j+1}^j,j}\right) + \left(\lambda_0(1-\hat{x}_{r_{t_0}^0,0})y^*_{r_{t_0}^0,0} + (1-\lambda_0)(1-\hat{x}_{r_{t_0+1}^0,0})y^*_{r_{t_0+1}^0,0}\right)\\
&-\sum_{j=1}^{n}\left(\lambda_j\Delta\hat{x}_{r_{t_j}^j,j} + (1-\lambda_j)\Delta\hat{x}_{r_{t_j+1}^j,j}\right) + \left(\lambda_0\Delta(1-\hat{x}_{r_{t_0}^0,0}) + (1-\lambda_0)\Delta(1-\hat{x}_{r_{t_0+1}^0,0})\right).
\end{aligned}
$$

The profit deviation is

$$
\begin{aligned}
&-\sum_{j=1}^{n}\left(\lambda_j\Delta\hat{x}_{r_{t_j}^j,j} + (1-\lambda_j)\Delta\hat{x}_{r_{t_j+1}^j,j}\right) + \left(\lambda_0\Delta(1-\hat{x}_{r_{t_0}^0,0}) + (1-\lambda_0)\Delta(1-x_{r_{t_0+1}^0,0})\right)\\
=&-\Delta\left(\sum_{j=1}^{n}\left(\lambda_j\hat{x}_{r_{t_j}^j,j} + (1-\lambda_j)\hat{x}_{r_{t_j+1}^j,j}\right) + \left(\lambda_0(1-\hat{x}_{r_{t_0}^0,0}) + (1-\lambda_0)(1-\hat{x}_{r_{t_0+1}^0,0})\right)\right)\\
=&\ \Delta\left(\sum_{j=1}^{n}x_j^* - (1-x_0^*)\right) = 0.
\end{aligned}
$$

The uniqueness of optimal price can be proved similarly. For completeness, we write down the price of product j as follows.

$$
\begin{aligned}
p_j^* =& -\left(\lambda_j^*\hat{y}_{r_{t_j}^j,j} + (1-\lambda_j^*)\hat{y}_{r_{t_j+1}^j,j}\right) + \left(\lambda_0^*\hat{y}_{r_{t_0}^0,0} + (1-\lambda_0^*)\hat{y}_{r_{t_0+1}^0,0}\right)\\
=& -\left(\lambda_j^*y^*_{r_{t_j}^j,j} + (1-\lambda_j^*)y^*_{r_{t_j+1}^j,j}\right) + \left(\lambda_0^*y^*_{r_{t_0}^0,0} + (1-\lambda_0^*)y^*_{r_{t_0+1}^0,0}\right)\\
&-\left(\lambda_j^*\Delta + (1-\lambda_j^*)\Delta\right) + \lambda_0^*\Delta + (1-\lambda_0^*)\Delta\\
=& -\left(\lambda_j^*y^*_{r_{t_j}^j,j} + (1-\lambda_j^*)y^*_{r_{t_j+1}^j,j}\right) + \left(\lambda_0^*y^*_{r_{t_0}^0,0} + (1-\lambda_0^*)y^*_{r_{t_0+1}^0,0}\right),
\end{aligned}
\tag{31}
$$

which implies that the prices obtained under the estimator $\bar{C}_j'(\cdot)$ for $t=1,\ldots,T$ and $j=0,\ldots,n$ is the same as those from the true $C_j'(\cdot)$. Q.E.D.

Appendix B: Examples Mentioned in Introduction

The following example illustrates the effect of model misspecification on pricing problems.

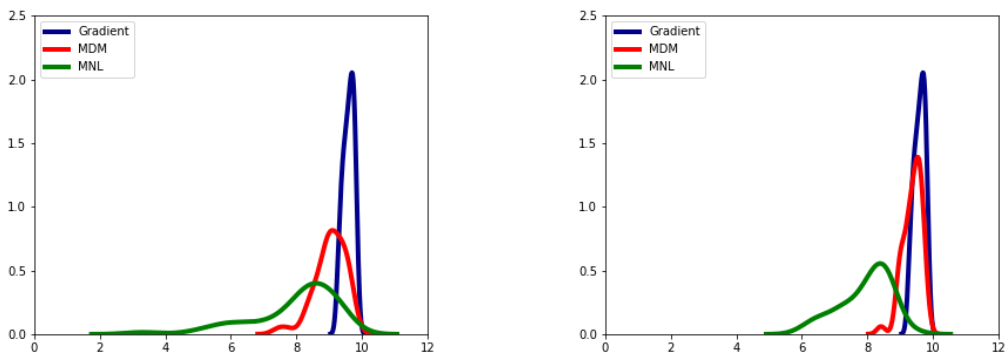
Example 1. We generate a set of aggregate demand data assuming the underlying choice model (ground truth) is a mixed logit choice model with 10 products and 500 consumer types, each with a different logit model. The synthetic data is generated as follows: we generate random sets of prices and calculate the corresponding demand of each product, using the mixed logit choice model. We then conduct an estimation with an MNL choice model (misspecified) and optimize the prices based on the estimated MNL model. The actual profit generated by the pricing solution is evaluated under the underlying mixed logit choice model. We also calculate the profit by a gradient descent method introduced in Li et al. (2019), which has been shown to perform well for the pricing problem with mixed logit choice models. We change the parameters in

the mixed logit choice model and repeat the above procedure to simulate the performance of the MNL-based approach vis-a-viz those obtained from the gradient descent approach.

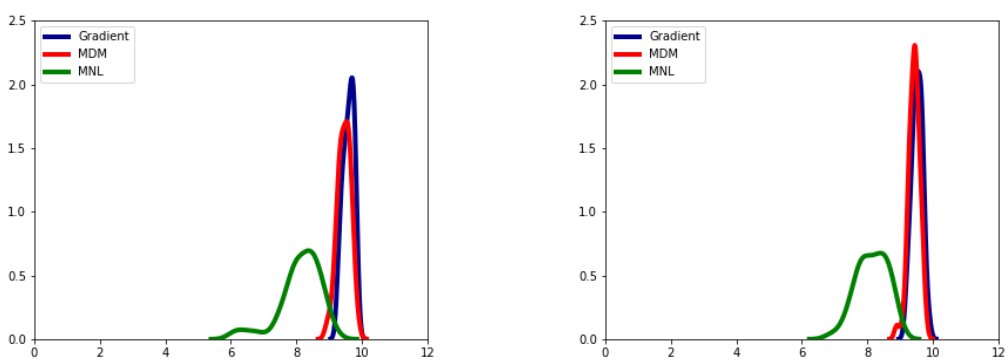
We plot the (normal) kernel density functions of the two sets of profits in Figure 6. With 20 pricing experiments, there is a big gap in the profits obtained from the MNL-based and gradient-based approaches. More importantly, this gap can not be eliminated even if we increase the number of pricing experiments to 100. In other words, there exists a genuine gap in performance if the underlying choice model is misspecified. The figure also provides the results from our proposed approach denoted by MDM. Applying the MDM approach and plotting the (normal) kernel density function of the profits in Figure 6, we observe that more experimental data result in a more accurate profit kernel density, indicating a closer fit of the underlying choice model. More importantly, the MDM approach makes minimal apriori assumptions on the form of choice models and uses only a parsimonious set of observations and data to obtain near-optimal prices in the experiments.

Figure 6 Profit Comparison for Mixed Logit Model

(a) Number of Price Experiment = 20 (b) Number of Price Experiment = 50



(c) Number of Price Experiment = 80 (d) Number of Price Experiment = 100



Example 2. We consider a set of six products and a population consisting of 1,000 consumers. For each customer, the utility of product i is drawn from a uniform distribution in $[0.5v_i, 1.5v_i]$ when $i = 1, 2$ and 3,

and a lognormal distribution with mean v_i and variance $v_i^2/12$ when $i = 4, 5$ and 6. The mean utility v_i is drawn randomly from $[3+i, 4+i]$l.

We randomly sample a subset of consumers (e.g., 100 consumers) to obtain their utility of the products and solve a standard MIP. As a comparison, we perform pricing experiments and aggregate the individual choices of all 1,000 customers in each experiment. We then apply our MDM and MNL approaches (as a benchmark) to the aggregate demand data to get the corresponding pricing solutions. We compare these approaches using four different data sets, with various numbers of price experiments (size = 5, 10, 20, 40). Figure 7 shows the average performance of the three approaches, over 50 independent simulations, normalizing the performance of the MIP approach to 1.

Note that under the distributional assumptions of the utility functions—a mix of a light-tailed distribution (uniform) and a heavy-tailed distribution (lognormal), the corresponding choice probabilities, to the best of our knowledge, cannot be explicitly characterized in closed form. On the other hand, the MIP approach, applicable for the pricing with given utility data, is restricted by the size of the consumer valuation set. In the numerical experiments, we use valuations from 100 customers, instead of the entire population, because of the long computational time it takes to solve the latter. Interestingly, using the aggregate demand information from 5 pricing experiments, MDM can match the performance of the MIP model (with 10% of the valuations in the population). MDM outperforms the MIP model when we have more pricing experiments. MNL consistently underperforms the MIP and MDM models in this setup, regardless of the number of pricing experiments.

Appendix C: Details on the Benchmarks

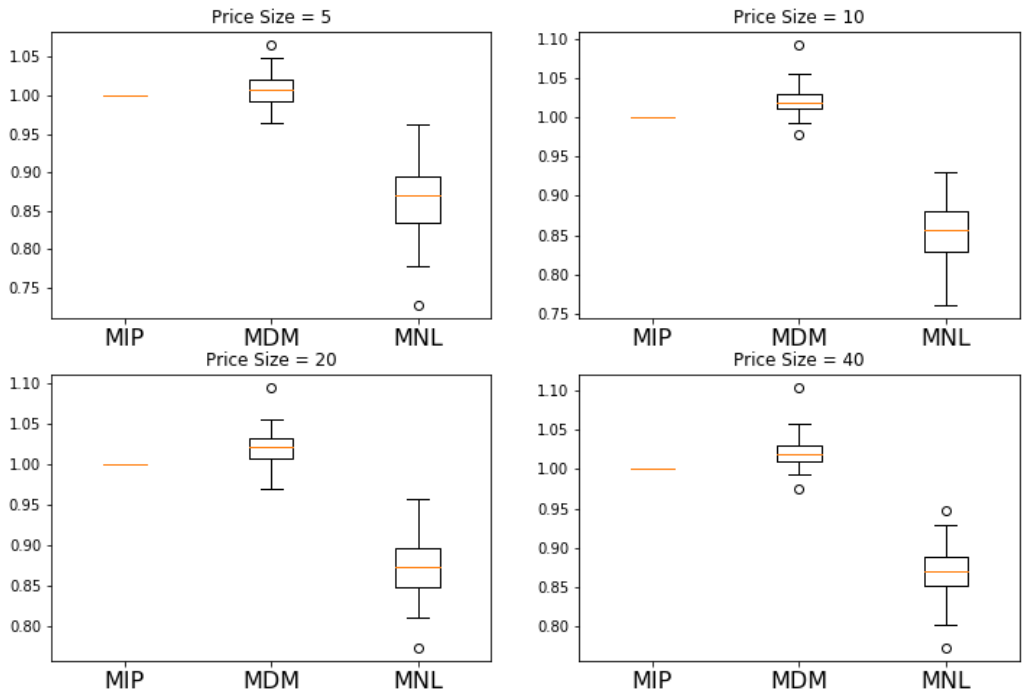
C.1. Pricing with Valuation Model (MIP Formula)

When a sample of consumer valuation of each product is given, Hanson and Martin (1990) proposed an elegant mixed-integer linear program to get the optimal prices. Specifically, the mixed-integer linear program is formulated as follows.

$$
\begin{aligned}
Z^I = \max_{\boldsymbol{p},\boldsymbol{y},\boldsymbol{\delta}} \; & \sum_{i=1}^{n} \sum_{w=1}^{T} \tfrac{\delta_{wi}}{T}, \\
s.t. \; & \delta_{wi} \geq p_i - M(1-y_{wi}), \forall i = 1,\ldots,n, w = 1,\ldots,T, \\
& \delta_{wi} \leq p_i, \forall i = 1,\ldots,n, w = 1,\ldots,T \\
& s_{wi} = v_i^{(w)} y_{wi} - \delta_{wi}, \forall i = 1\ldots,n, w = 1,\ldots,T, \\
& \sum_{i=1}^{N} s_{wi} \geq v_j^{(w)} - p_j, \forall j = 1,\ldots,n, w = 1,\ldots,T, \\
& \sum_{i=1}^{N} s_{wi} \geq \sum_{i=1}^{n} v_i^{(w)} y_{ki} - \delta_{ki}, \forall w, k = 1,\ldots,T, \\
& \sum_{i=1}^{N} y_{wi} \leq 1, \forall w = 1,\ldots,T, \\
& y_{wi} \in \{0,1\}, \forall i = 1,\ldots,n, w = 1,\ldots,T, \\
& s_{wi} \geq 0, \delta_{wi} \geq 0, \forall i = 1\ldots,n, w = 1,\ldots,T,
\end{aligned} \tag{32}
$$

where $M = \max_{i,w} v_i^w$. The first two constraints imply that if $y_{wi} = 1$, then $\delta_{ki} = p_i$. The third constraint and the non-negativity constraint of s_{wi} further imply that if $y_{wi} = 0$, then $\delta_{wi} = 0$ and $s_{wi} = 0$. Therefore, δ_{wi} is

Figure 7 Box-plot of Profit Ratio Comparison for Pricing with Utility Model



regarded as the unit revenue earned by the retailer, and s_{wi} represents consumer w's surplus if she chooses product i. The fourth constraint enforces the surplus maximizing constraint, i.e., the selected item generates the highest surplus among all the products. The fifth constraint is a redundant one, which helps tighten the formula (see Hanson and Martin (1990)). The sixth constraint restricts each customer to choose at most one product.

C.2. Pricing with Elasticity Model

Price elasticity of demand is a commonly used term in the economy to characterize demand sensitivity to price changes. It is defined as the ratio of the percentage change in demand to the percentage change in prices (see (33)).

$$\beta = \frac{\Delta Q/Q}{\Delta P/P}. \tag{33}$$

Given a set of data on prices and aggregate demand vectors $(p_{tj}, \hat{x}_{tj})$ for $t = 1, \ldots, T$ and $j = 1, \ldots, n$, the elasticity coefficient can be estimated by minimizing the deviation between theoretical and observed demand. In particular, the theoretical demand at a price p_{tj} based on the elasticity model is $Q_j + \sum_i \frac{Q_j}{P_i}\beta_{ji}(\hat{p}_{ti} - P_i)$ by definition, where β_{ji} is the price elasticity coefficient between product j and i, $\boldsymbol{P}$ denotes the base price, and $\boldsymbol{Q}$ is the corresponding demand for the base price. We use l_1 to measure the deviation for the sake of tractability. Then the elasticity estimation model is built to as follows.

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \sum_{j=0}^{n} \left| Q_j + \sum_i \frac{Q_j}{P_i}\beta_{ji}(\hat{p}_{ti} - P_i) - \hat{x}_{tj} \right|. \quad \text{(Elasticity Estimation)} \tag{34}$$

Using an estimated demand price elasticities $\hat{\boldsymbol{\beta}}$, we can solve the price optimization problem by a quadratic program (QP) as follows.

$$\max_{\boldsymbol{\Delta P}} \sum_{j=1}^{n} (Q_j + \sum_i \frac{Q_j}{P_i} \hat{\beta}_{ji} \Delta P_i)(P_j + \Delta P_j). \quad \text{(Pricing with Elasticity Model)} \tag{35}$$

C.3. Pricing with Multinomial Logit Choice Model

According to Natarajan et al. (2009), the optimal solution to the representative consumer model recreates MNL choice model by defining $C_j(x_j) = -\int_{1-x_j}^{1} F_j^{-1}(t)dt$, where $F_j(\epsilon) = 1 - e^{-\epsilon}$ for $\epsilon \geq 0$ for $j = 0, \ldots, n$. The corresponding choice probability at a price vector $\boldsymbol{p}$ is calculated as follows.

$$x_j(\boldsymbol{p}) = \frac{e^{v_j - \alpha p_j}}{1 + \sum_{k=1}^{n} e^{v_k - \alpha p_k}}, \ \forall j = 1, \ldots, n, \tag{36}$$

where v_j is the deterministic utility of product j (excluding the price) and α is the price-sensitivity parameter, in which without loss of generality, we assume $v_0 = p_0 = 0$ for the outside option.

We further show that when $\alpha = 1$ the pricing with MNL choice model can be recreated from (4) and (6) by defining $F_j(\epsilon) = 1 - e^{-\epsilon}$ for $\epsilon \geq 0$ for $j = 0, \ldots, n$. In particular, the corresponding pricing model is as follows.

$$\begin{aligned} \max_{\boldsymbol{x}} \ & \sum_{j=1}^{n} (v_j - w_j) x_j - \sum_{j=1}^{n} x_j \ln(x_j) + \left(\sum_{j=1}^{n} x_j \right) \ln \left(1 - \sum_{j=1}^{n} x_j \right) \\ \text{s.t.} \ & \sum_{j=1}^{n} x_j \leq 1, \\ & x_j \geq 0, \ \forall j = 1, \ldots, n. \end{aligned} \quad \text{(Pricing with MNL Model)} \tag{37}$$

It can be easily verified the tail function $\bar{F}_j(\epsilon)$ is log-concave in ϵ for each j, and the marginal distribution function of the outside option $F_0(\epsilon)$ is log-concave. From Proposition 1 and Theorem 1, the optimal pricing problem is a convex problem and the optimal pricing strategy is

$$p_j = v_j + \ln \left(1 - \sum_{j=1}^{n} x_j \right) - \ln x_j, \ \forall j = 1, \ldots, n. \tag{38}$$

That is exactly the optimal pricing function under MNL model shown in Proposition 2 in Song and Xue (2007).

As a special case of the representative consumer model, the MNL choice model can be estimated by minimizing deviation from the optimality condition (38) given a set of data on prices and aggregate and normalized demand vectors $(p_{tj}, \hat{x}_{tj})$ for $t = 1, \ldots, T$ and $j = 1, \ldots, n$. Similar ideas have been used by Berry et al. (1995) in estimating random coefficient logit models with aggregate demand data. For the sake of tractability, we use l_1 norm to obtain a LP based estimation model in (39).

$$\begin{aligned} \min_{\hat{\boldsymbol{v}}, \boldsymbol{\phi}} \ & \sum_{t,j} \phi_{tj} \\ \text{s.t.} \ & \phi_{tj} \geq \hat{v}_j + \ln(1 - \sum_{j=1}^{n} \hat{x}_{tj}) - \ln(\hat{x}_{tj}) - p_{tj}, \forall t = 1, \ldots T, j = 1, \ldots, n, \\ & \phi_{tj} \geq -(\hat{v}_j + \ln(1 - \sum_{j=1}^{n} \hat{x}_{tj}) - \ln(\hat{x}_{tj}) - p_{tj}), \forall t = 1, \ldots T, j = 1, \ldots, n, \\ & \hat{\mathbf{v}} \geq \mathbf{0}. \end{aligned} \quad \text{(MNL Estimation)} \tag{39}$$

C.4. Pricing with Nested Logit Choice Model

Consider an NL model with K nests. We denote each nest as a set N_k and use the parameter $\tau_k \in [0,1]$ to represent the dissimilarity among products in nest k. Suppose the deterministic utility of each product is v_j for $j \in N_k$ and $k = 1\ldots, K$. We show that by defining $C_j(x_j) = -\int_{1-x_j}^{1} F_j^{-1}(t)dt$, where $F_{jk}(\epsilon) = 1 - e^{-\epsilon}\left(\sum_{j\in N_k} e^{v_{jk}-p_{jk}}\right)^{\tau_k-1}$ for $\epsilon \geq (\tau_k - 1)\ln(\sum_{j\in N_k} e^{v_{jk}-p_{jk}})$, the pricing with representative consumer model (4) in the form of (6) recreates the pricing with NL model. In particular, from the optimality condition of the representative consumer model (4) with the aforementioned perturbation functions, we can get

$$p_j = v_j + (\tau_k - 1)\ln(\sum_{j\in N_k} e^{v_j - p_j}) - ln(x_j) - \ln\left(1 + \sum_{t=1}^{K}(\sum_{j\in N_t} e^{v_j-p_j})^{\tau_t}\right), \quad \text{for } j \in N_k. \tag{40}$$

The corresponding choice probability at price vector $\boldsymbol{p}$ is calculated as follows.

$$x_j = \frac{e^{v_j-p_j}\left(\sum_{l\in N_k} e^{v_l-p_l}\right)^{\tau_k-1}}{1+\sum_{t=1}^{K}\left(\sum_{l\in N_t} e^{v_l-p_l}\right)^{\tau_t}}, \quad \text{for } j \in N_k. \tag{41}$$

Its pricing problem can be modeled as a convex optimization as follows.

$$\begin{aligned}
\max_{\mathbf{x}} \ & \sum_{k=1}^{K}\sum_{j\in N_k}(v_j x_j - x_j\ln(x_j)) + \sum_{k=1}^{K}\left(\frac{1}{\tau_k}(\sum_{l\in N_k} x_l)\ln(1-\sum_{t=1}^{K}\sum_{i\in N_t} x_i) - \frac{1-\tau_k}{\tau_k}(\sum_{l\in N_k} x_l)\ln(\sum_{l\in N_k} x_l)\right) \\
\text{s.t.} \ & \sum_{k=1}^{K}\sum_{j\in N_k} x_j + x_0 = 1 \qquad \text{(Pricing with Nested Logit Model)} \\
& x_j \geq 0, \forall j \in N_k, k = 1,\ldots,K.
\end{aligned} \tag{42}$$

The formula recreates that in Theorem 1 in Li and Huh (2011).

Similar to the MNL choice model, given a set of data on prices and aggregate and normalized demand vectors $(p_{tj}, \hat{x}_{tj})$ for $t = 1,\ldots,T$ and $j = 1,\ldots,n$, we estimate the NL model with pre-specified nests by minimizing deviation from the optimality condition (40) measured by l_1 norm. Note that the R.H.S. function in (40) is nonlinear in the parameters to estimate. To get a tractable estimation model, we define $q_k := \frac{(\sum_{j\in N_k} e^{v_j-p_j})^{\tau_k}}{1+\sum_{t=1}^{K}(\sum_{j\in N_t} e^{v_j-p_j})^{\tau_t}}$, from which we can derive

$$1 - \sum_{k=1}^{K} q_k = \frac{1}{1+\sum_{k=1}^{K}(\sum_{j\in N_k} e^{v_j-p_j})^{\tau_k}}, \quad \frac{q_k}{1-\sum_{l=1}^{K} q_l} = (\sum_{j\in N_k} e^{v_j-p_j})^{\tau_k}, \quad q_k = \sum_{j\in N_k} x_j.$$

From these equations, we can rewrite the optimality condition (40) as follows.

$$p_j = v_j + \frac{1-\tau_k}{\tau_k}\left[\ln(1-\sum_{l=1}^{K} q_l) - \ln q_k\right] + \left[\ln(1-\sum_{l=1}^{K} q_l) - \ln x_j\right], \quad \text{for } j \in N_k.$$

Therefore, the estimation problem can be reformulated as a linear program as follows.

$$\begin{aligned}
\min_{\hat{\boldsymbol{v}},\boldsymbol{\tau}^{-1},\boldsymbol{\phi}} \ & \sum_{t,j}\phi_{tj} \\
\text{s.t.} \ & \phi_{tj} \geq \hat{v}_j + (\tau_k^{-1}-1)\left(\ln(1-\sum_{k=1}^{K}\hat{q}_{tk}) - \ln(\hat{q}_{tk})\right) + \ln(1-\sum_{k=1}^{K}\hat{q}_{tk}) - \ln(\hat{x}_{tj}) - p_{tj}, \forall t, j \in N_k \\
& \phi_{tj} \geq -\left(\hat{v}_j + (\tau_k^{-1}-1)\left(\ln(1-\sum_{k=1}^{K}\hat{q}_{tk}) - \ln(\hat{q}_{tk})\right) + \ln(1-\sum_{k=1}^{K}\hat{q}_{tk}) - \ln(\hat{x}_{tj}) - p_{tj}\right), \forall t, j \in N_k \\
& \boldsymbol{\tau}^{-1} \geq \mathbf{1}, \quad \hat{\mathbf{v}} \geq \mathbf{0}, \qquad \text{(Nested Logit Model Estimation)}
\end{aligned} \tag{43}$$

where $\hat{q}_{tk} = \sum_{j \in N_k} \hat{x}_{tj}$.

Appendix D: Numerical Settings for Pricing with Customer Valuation Sample

The procedure to simulate valuation data from mixing distributions is shown in Algorithm 1, with a proportion of ν valuations from log-normal distributions and the rest from uniform distributions. With a sample of valuations, a mixed-integer linear program proposed by Hanson and Martin (1990) (see (32)) can be applied to get a benchmark. Alternatively, we can apply the method in this paper to get another set of pricing

Output: 1,000 individuals' valuations of $n = 4$ products

Input: The proportion of products with log-normal distributed valuations ν.

(Valuation Data) The valuation data is generated in the following way.

1. Randomly generate a mean valuation vector $\bar{\boldsymbol{v}}$ with $\bar{v}_j$ uniformly distributed in $(3+j, 4+j)$ for $j = 1, \ldots, n$.
2. Sample 1,000 individuals' valuations of the first $\lfloor (1-\nu)n \rfloor$ products from a uniform distribution in $(0.5\bar{v}_j, 1.5\bar{v}_j)$, for $j = 1, \ldots, \lfloor (1-\nu)n \rfloor$ and the other products from a log-normal distribution with mean $\bar{v}_j$ and variance $\frac{1}{12}\bar{v}_j^2$ for $j = \lfloor (1-\nu)n \rfloor + 1, \ldots, n$.

Algorithm 1: Valuation Data Generation

solutions. In detail, we uniformly generate a set of prices (see Algorithm 2) and calculate the corresponding demand at each price based on the valuation data, which serves as the input data in MDM approach. We also apply an MNL model to fit the aggregate demand data and get the corresponding pricing solution as another benchmark. We evaluate the true profit by each pricing solution using the valuation data generated from Algorithm 1 under a utility maximization framework. Different price experiments are tested to see how the performance varies with the sample size. A detailed procedure to compare the three methods is shown in Algorithm 3.

Appendix E: Details on Industry Data

E.1. Base Price and Cost in the Automobile Data Set

The base price and cost for each product is shown in Table 7.

E.2. Prices from MDM and Experiment in the Fast-Food Data Set

The obtained pricing solution is reported in the MDM column in Table 8, where Entrée$_i$ (for $i = 1, \ldots, 14$) represents each entrées and S_i (M_i, L_i) to denote its corresponding small (medium, large)-sized EVM if there is any. The happy meal and sides are labeled by C and D, receptively. Column P_i (for $i = 1, \ldots, 4$) in the table reports the price in ith experiment in the data, where P_1 is used as the base price.

Output: Various sets of price experiment $(p_{tj}^k) \in \mathcal{P}_k$ for $j = 1, \ldots, n$, $t = 1, \ldots, T_k$, and $k = 0, 1, \ldots, 9$.

Input: Valuation data from Algorithm 1

(Price Experiment) Price data is generated in the following way.

1 Uniformly generate 100 sets of prices for the 10 products around 20% of the valuation sample mean.

2 Design ten nested subsets of prices whose price sample size varies from 5 to 10 and from 10 to 90 in increments of 10. Denote each subset as $\mathcal{P}_k$ for $k = 0, 1, \ldots, 9$, with a size of $T_k \in \{5, 10, \ldots, 90\}$. $\mathcal{P}_k$ is randomly sampled from $\mathcal{P}_{k+1}$, i.e., $\mathcal{P}_k \subset \mathcal{P}_{k+1}$.

Algorithm 2: Price Experiment Design

Result: Price from MIP,MDM and MNL

for $\nu \in \{0, 0.5, 1\}$ **do**

  Generate valuation data from Algorithm 1 and solve the MIP model (32) with 100 samples from the generated valuations;

  Generate prices from Algorithm 2 based on the generated valuations;

  for $k \in \{0, , \ldots, 9\}$ **do**

    (Input Generation) Gnerate demand data x_{tj}^k based on the generated valuations for each product $j = 1, \ldots, n$ at each price $\boldsymbol{p}_t^k \in \mathcal{P}_k$ under the utility maximization framework;

    (MDM) Apply estimation model 11 and optimization model (15) to data $(p_{tj}^k, x_{tj}^k)_{tj}$;

    (MNL) Apply estimation model 39 and optimization model (37) to data $(p_{tj}^k, x_{tj}^k)_{tj}$;

  end

end

Algorithm 3: Comparison of MIP, MDM and MNL Approaches

Appendix F: Numeral Studies of Pricing with MNL and Nested Logit Model

F.1. Numeral Studies of Pricing with MNL Model

Suppose the underlying choice model is an MNL choice model. We uniformly sample different sets of prices and generate their corresponding choice probabilities using the MNL choice model shown in (36) as the observed demand in the simulation data. We apply different methods to fit the data and optimize the set of prices. We evaluate the profit of the obtained prices using the underlying choice model (MNL) and get the profit. As a reference, we solve the pricing problem with the MNL model exactly by (37) to get the true optimal profit, assuming that we know the true MNL parameters. We compare the gap between the profits generated from different methods and true profits.

Table 7 Base Price and Cost in the Data Set

	Brand	Base Price(thousand)	Cost(thousand)
1	A	45.5209	34.1407
2	A	46.906	35.1795
3	A	45.056	33.792
4	A	43.8742	32.9057
5	A	47.0282	35.2711
6	A	42.0442	31.5331
7	A	45.5225	34.1419
8	B	39.455	35.5095
9	B	37.955	34.1595
10	B	33.1821	29.8639
11	B	27.255	24.5295
12	B	35.9266	32.334
13	B	33.5504	30.1953
14	B	37.9559	34.1595
15	C	39.750	33.7875
16	C	35.809	30.4376
17	C	38.227	32.4933
18	C	41.250	35.0625
19	C	36.0612	30.652
20	C	39.750	33.7875

We generate 50 sets of simulation data assuming the underlying choice model (ground truth) is an MNL choice model. Each ground truth is generated by uniformly sampling the parameter $\boldsymbol{v}$ within a certain range (in this example, the range is from 5 to 15). In each simulation, we generate 50 sets of prices and choice probabilities. Various methods are tested in the data set, including the MDM method, an online method using the loss function defined in Dong et al. (2018), and a price elasticity model.

Using MDM method, (11) is used to get pointwise estimates of $C_j(\cdot), \forall j = 0, \ldots, n$ and (15) is applied to optimize prices. This is denoted MDM(F). We have also implemented an online version of MDM, which we termed MDM(O), to facilitate comparison. We omit the details here, but suffice to say, it uses a notion of loss function similar to Dong et al. (2018). Finally, the estimation model of price elasticity coefficients from aggregate demand data and the pricing with an elasticity model are referred to (34) and (35), respectively.

We report the profit obtained from different approaches in Figure 8. The result in the blue dash line represents the profit from the MDM (F) method. It almost hits the ground truth (in red solid line), which indicates that the MDM (F) method indeed generates the true optimal price with only 50 samples. The MDM (O) method (in green long dash line) achieves a near-optimal pricing solution with $T = 50$, but it requires a longer computation time. However, using the online method with the original loss function in Dong et al.

Table 8 Price Solution by MDM versus Prices in the Experiments

	MDM	P_1	P_2	P_3	P_4		MDM	P_1	P_2	P_3	P_4		MDM	P_1	P_2	P_3	P_4
Entrée$_1$	3.10	3.15	3.15	3.16	3.15	S_1	5.90	5.90	5.92	5.96	5.95	C_1	7.65	7.65	7.67	7.71	7.70
Entrée$_2$	5.05	5.00	5.04	5.10	5.10	S_2	8.30	8.30	8.32	8.37	8.35	C_2	5.60	5.60	5.64	5.71	5.70
Entrée$_3$	6.30	6.30	6.34	6.41	6.40	S_3	10.05	10.05	10.09	10.17	10.15	C_3	8.00	7.95	8.01	8.11	8.10
Entrée$_4$	6.40	6.40	6.40	6.41	6.40	S_4	10.05	10.05	10.09	10.18	10.15	C_4	5.65	5.60	5.64	5.71	5.70
Entrée$_5$	6.25	6.15	6.25	6.40	6.40	S_5	9.70	9.70	9.74	9.82	9.80	C_5	5.65	5.60	5.64	5.71	5.70
Entrée$_6$	5.05	5.05	5.07	5.11	5.10	S_6	9.00	9.00	9.04	9.12	9.10	C_6	5.65	5.60	5.64	5.71	5.70
Entrée$_7$	5.85	5.85	5.87	5.91	5.90	S_7	8.90	8.90	8.94	9.03	9.00						
Entrée$_8$	8.05	8.05	8.09	8.16	8.15												
Entrée$_9$	7.50	7.50	7.52	7.57	7.55												
Entrée$_{10}$	6.30	6.25	6.31	6.42	6.40												
Entrée$_{11}$	4.90	4.95	4.95	4.95	4.95												
Entrée$_{12}$	3.00	3.00	3.00	3.01	3.00												
Entrée$_{13}$	3.40	3.35	3.39	3.46	3.45												
Entrée$_{14}$	13.80	13.70	13.78	13.92	13.90												
	MDM	P_1	P_2	P_3	P_4		MDM	P_1	P_2	P_3	P_4		MDM	P_1	P_2	P_3	P_4
M_1	6.60	6.55	6.59	6.66	6.65	L_1	7.25	7.20	7.24	7.31	7.30	D_1	1.75	1.75	1.75	1.75	1.75
M_2	9.00	8.95	8.99	9.07	9.05	L_2	9.65	9.60	9.64	9.72	9.70	D_2	3.25	3.20	3.22	3.25	3.25
M_3	10.70	10.65	10.71	10.81	10.80	L_3	11.35	11.30	11.36	11.46	11.45	D_3	2.80	2.75	2.77	2.80	2.80
M_4	10.65	10.70	10.74	10.82	10.80	L_4	11.40	11.35	11.39	11.47	11.45	D_4	2.15	2.15	2.15	2.15	2.15
M_5	10.40	10.35	10.39	10.47	10.45	L_5	11.05	11.00	11.04	11.12	11.10						
M_6	9.75	9.70	9.74	9.82	9.80	L_6	10.40	10.35	10.39	10.47	10.45						
M_7	9.70	9.65	9.69	9.77	9.75	L_7	10.35	10.30	10.34	10.42	10.40						
M_8	11.85	11.80	11.86	11.97	11.95	L_8	12.50	12.45	12.51	12.62	12.60						
M_9	11.35	11.30	11.34	11.42	11.40	L_9	12.00	11.95	11.99	12.07	12.05						
M_{10}	10.45	10.40	10.44	10.52	10.50	L_{10}	11.10	11.05	11.09	11.17	11.15						
M_{11}	9.00	9.05	9.05	9.05	9.05	L_{11}	9.60	9.65	9.65	9.65	9.65						

(2018), the performance in these 50 different instances varies a lot (see the result in purple round dot line). It achieves a near-optimal solution in some cases but also results in a bad solution in other cases. In addition, it is time-consuming. In other words, online learning under our new loss function demonstrates a faster convergence rate. One possible reason is that under the original definition, the search space is much larger. Both demand and the slopes of piecewise linear functions are decision variables. There could be multiple combinations to reach the same price. Therefore, it will require more iterations to achieve a good solution in general and the computation time for each iteration is also longer. Last, the elasticity-based model (in black square dot line) does not perform well in this case, achieving only 65.5% optimality on average. We also report the computation time in Table 9. The results demonstrate that the MDM (F) method can achieve a good pricing solution within seconds, even with a small data sample.

Figure 8 Profit Comparison under MNL model

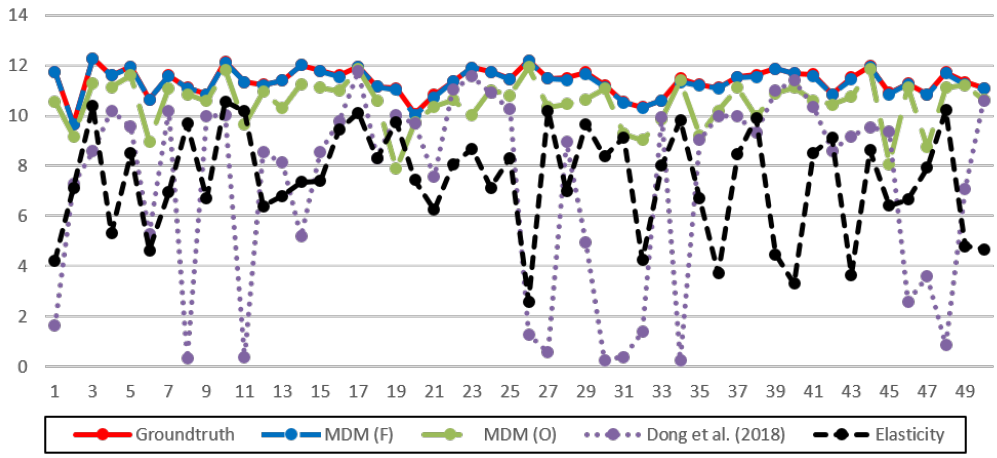


Table 9 Computation Time Comparison (MNL)

	MDM (F)		MDM (O)		Online (Dong et al. (2018))		Elasticity	
	Est	Opt	Est	Opt	Est	Opt	Est	Opt
Time (s)	0.60	0.25	0.28*50	0.89	15.62*50	0.33	2.0896	0.0127

F.2. Numeral Studies of Pricing with NL Model

We repeat the experiment, but this time the simulated data is generated from an NL model, whose pricing problem is known to be convex and hence we can get the true optimal profit. Assume that there are $N = 20$ products. We divide the products into $K = 4$ branches where each branch has 5 products. The choice probability under an NL model at price vector $\boldsymbol{p}$ is calculated from (41).

We randomly generate 50 $(\boldsymbol{v}, \boldsymbol{\tau})$ with the entries in $\boldsymbol{v}$ from 10 to 15 and the entries in $\boldsymbol{\tau}$ between 0.3 and 0.7. To generate a wide range of price data while at the same time ensuring the prices between products in one instance do not vary too much, for each $\boldsymbol{v}$ and $\boldsymbol{\tau}$, we generate the sales data as follows. Each price instance is generated after randomly selecting an integer number l from 1 to 10, and then each p_j is uniformly sampled within 10% range of $0.2lv_j$, i.e., $(0.2lv_j \times 0.9, 0.2lv_j \times 1.1)$. We generate 50 prices in each instance. For price vector $\boldsymbol{p}$, we use the choice probability calculated from (41) as the observed demand in the simulation data.

Similar to the case with the MNL model, we apply different methods to the simulation data and compare their performance. Again the performance is evaluated using the underlying choice model. The true optimal profit can be calculated by solving (42) under the true parameter $(\boldsymbol{v}, \boldsymbol{\tau})$.

We report the profits obtained from different methods in Figure 9 and the computation time in Table 10. The results are consistent with those in the MNL case, but the profit from the MDM (F) method generates a slightly worse solution than the true optimal one. It indicates that more sample is required if the underlying choice model is an NL model.

Figure 9 Profit Comparison under NL model

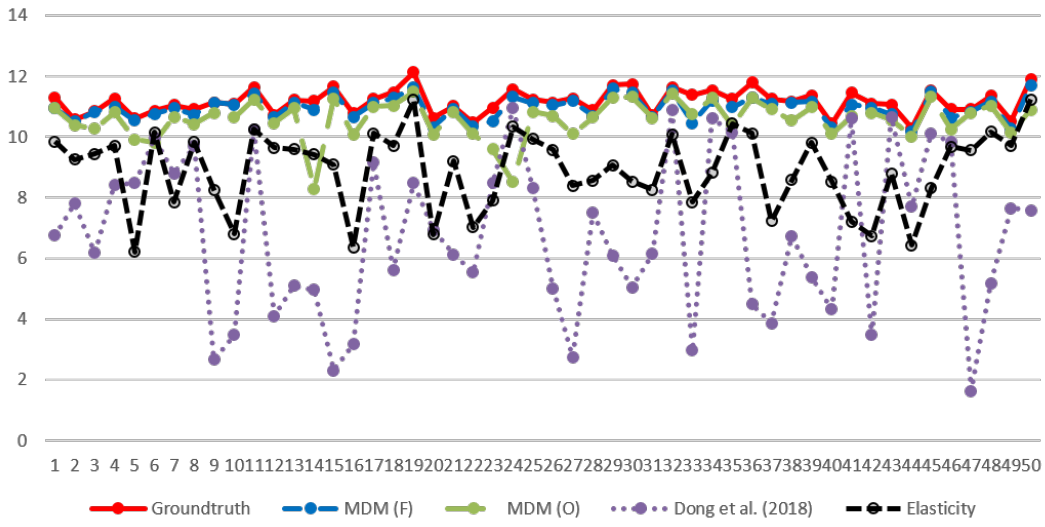


Table 10 Computation Time Comparison (NL)

	MDM (F)		MDM (O)		Online (Dong et al (2018))		Elasticity	
	Est	Opt	Est	Opt	Est	Opt	Est	Opt
Time (s)	0.71	0.38	0.48*50	27.21	103.39*50	0.63	2.0615	0.0125

Appendix G: The Value of Adaptive Procedure

Note that the perturbation function constructed to recreate the NL choice model is not separable (see Appendix C.4). Hence we apply the adaptive procedure mentioned in the concluding remark to resolve the pricing with an NL model. The details of the adaptive procedure are provided in Table 11.

The way to generate synthetic data is similar to that in Appendix F.2. We generate 200 prices in each iteration and apply the MDM method to get a pricing solution. Next, we apply the adaptive sample generating procedure in Table 11 to generate more samples. Ten iterations are tested for each instance. We repeat the procedure for 100 different instances which are generated by NL models with $\boldsymbol{v}$ sampling from 5 to 15 and $\boldsymbol{\tau}$ sampling from 0.3 to 0.7. We plot the normal kernel density estimation function of the 100 generated profits from the MDM method in each iteration in Figure 10. From the figure, we can see that after around 5 rounds of the adaptive sampling procedure (price sample size increases from 200 to 1000), the MDM method generates almost the same kernel density as the true optimal profit function. Here, similar to Appendix F.2, the true optimal profit can be calculated by solving (42) under the true parameter $(\boldsymbol{v}, \boldsymbol{\tau})$.

Finally, we would like to examine how well the piecewise linear function approximates the profit function in the proposed framework. The objective value of (15) returns the profit approximated by the piecewise linear function at the pricing solution by the MDM method. On the other hand, by evaluating the profit at the same price using the underlying choice model (NL model), we can get a true profit at the same price. The gap between the two profits, measured by the relative percentage, reflects the approximation error using the piecewise linear approximation. We summarize the approximation gap in each iteration on the simulated

Table 11 Adaptive Sample Generating Procedure

Adaptive Sample Generating Procedure
Step 1. Randomly generate a set of price in a reasonable and relatively large interval. (e.g., with the center of estimated deterministic utility v, from 0 to $2v$)
Step 2. Apply the estimation model model (11) and optimization (15) model to get an optimal price as the starting price, denoted as $\boldsymbol{p}^*$.
Step 3. Regenerate a set of prices $\mathcal{P}$ around $\boldsymbol{p}^*$, with a relatively small interval. (e.g., uniformly generate prices within the range $(\boldsymbol{p}^*(-5\% \sim -1\%), \boldsymbol{p}^*(1\% \sim 5\%)))$ and obtain the demand.
Step 4. Apply the estimation (11) and optimization (15) model to get a new optimal price $\hat{\boldsymbol{p}}^*$. Evaluate the profit generated by $\hat{\boldsymbol{p}}^*$, denote the profit as $\hat{\Pi}^*$. If $\hat{\Pi}^* \geq \Pi(\boldsymbol{p}^*)$, let $\boldsymbol{p}^* = \hat{\boldsymbol{p}}^*$, otherwise, keep $\boldsymbol{p}^*$. Go to Step 3.

Figure 10 Profit from the Iterative Method (NL Model)

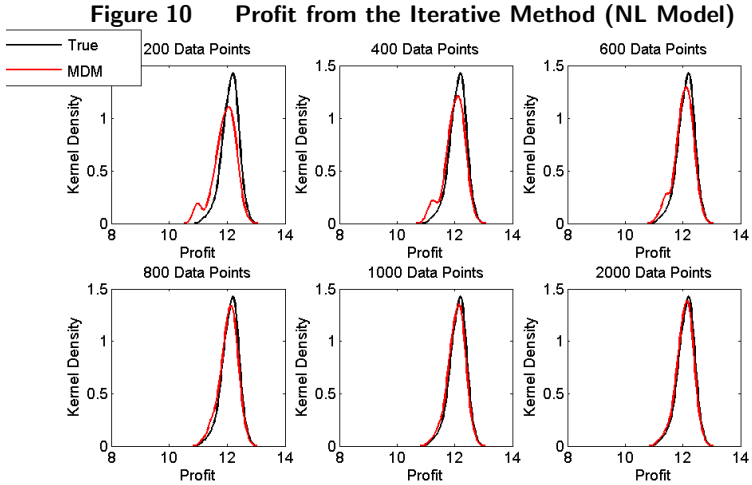


Table 12 Summary of Approximation Gap of Profit Function (NL)

Round	1	2	3	4	5	6	7	8	9	10
Mean	0.0338	0.0064	0.0138	0.0034	0.0028	0.0038	0.0024	0.0023	0.0028	0.0022
Std	0.0327	0.0063	0.0997	0.0030	0.0022	0.0085	0.0016	0.0015	0.0058	0.0013

100 different NL models in Table 12. From the table, we can see that the approximation gap is small (below 0.5%), especially after applying for the iterative procedure.