“Marginal Estimation + Price Optimization” for Multi-Product Pricing Problems

Zhenzhen Yan∗ Cong Cheng† Karthik Natarajan ‡ Chung-Piaw Teo§

February 7, 2018

Abstract

In this paper, we develop a data-driven approach for the multi-product pricing problem by exploiting properties of the representative consumer model in discrete choice. We relate a special case of the representative consumer model to a semiparametric choice model (called the Marginal Distribution Model) and show that the multi-product pricing problem in this case is convex when the marginal probability density functions are log-concave. In the special case with exponential marginal distributions, this recreates the convexity results of multi-product pricing under the Multinomial Logit model, while generalizing the result to other marginal distributions such as normal, logistic, extreme value or Laplace distributions.

Using this approach, we establish a set of closed-form relationships between prices and market shares for the products. While it is difficult to calibrate the shape of a general regularization function for the representative consumer model from data, we show that using a separable regularization function provides good estimates on the structure of the choice model and guides the algorithm in the search for optimal prices. In this way, we develop second order conic and linear programs to estimate the shape of the separable regularization function in the representative consumer model. Mixed integer linear programming models are used to find the optimal prices when side constraints are present and linear programs when side constraints are absent. This partially addresses the problem of model misspecification for pricing problems, since we do not explicitly assume the structural form in the consumer’s utility model. Extensive tests

∗National University of Singapore, Singapore. Email: a0109727@u.nus.edu
†Northeastern University, China. Email: chec_5588@163.com
‡Engineering System and Design, Singapore University of Technology and Design, Singapore. Email: karthik_natarajan@sutd.edu.sg
§National University of Singapore, Singapore. Email: bizteocp@nus.edu.sg
using both simulated data and industry data demonstrates clearly the benefits of this “Marginal Estimation + Price Optimization” approach.

1 Introduction

Pricing, as a strategy to shape the demand in a market, is commonly used in industry and has been extensively studied both by academicians and practitioners (see Talluri and van Ryzin (40)). The primary goal is to try and understand how demand is influenced by prices and to then optimize prices to maximize the profit. The growth of the Internet has led to a drastic increase in the capability of companies to make changes in prices to learn the demand model while doing price optimization. Einav et al. (15) analyze the targeted pricing and auction design variations with eBay sales data and find that of the 100 million listings on a given day, more than half will reappear as a separate listing, often with modified sales parameters (such as prices). In fact, with the boom in e-commerce activities, the practice of changing prices strategically to learn more about the customer’s behavior has been adopted by many companies. “Online retailers are turning to data to help them compete, and they have strategic price ranges that they play between”, says Meghan Heffernan, a spokeswoman for Savings.com. The availability of data from pricing experiments presents companies with an opportunity to increase profits if they can use it effectively. Hence, research on revenue maximization using data from pricing experiments has exploded in recent years.

In the multi-product setting, the difficulty in the pricing problem lies in capturing the impact of the prices on the cross-elasticity of demand, especially if the customer can only purchase one out of a set of products offered. Ferreira et al. (26) discuss the operational challenges faced by the online fashion retailer Rue La La which offers price discounts. They use regression trees and machine learning techniques to estimate the demand model with the use of integer programming to do price optimization. In another recent work, Bertsimas and Vayanos (10) study the exploration-exploitation tradeoff using dynamic pricing where the demand and price relationship is modeled through uncertainty sets that encode the seller’s beliefs about the demand curve parameters. Using mixed integer conic optimization, they develop an adaptive dynamic pricing mechanism to do price optimization. The demand and price relationship in their work is modeled through a linear function with additive error terms where the error terms are assumed to lie in a bounded uncertainty set. Fisher et al. (18) discuss a competition-based dynamic pricing model for the Chinese online retailer Yihaodian where pricing experiments are used to estimate price sensitivities and a best-response pricing algorithm is used to choose prices. In their work, a Nested Logit (NL) model is used to capture the customer’s choice behavior.
While this stream of research showcases the power of the “choice model estimation and optimization” framework in pricing products using data from experiments, a natural concern is the accuracy of the estimation model in capturing the customer’s purchase behavior, and the computational tractability of the pricing optimization model arising from the estimation technique. The critical step is to find a “proper” model to characterize the market demand, based on the available sales data. Our experience in working with data from a large automobile manufacturer is that the market demand model is often a “black box” to companies which predicts the sales for each product through a complex, time consuming market simulator using a pricing proposal (with all other features fixed) as the input to the black box. The market share of each product is then a reflection of the expected percentage of the total sales in the market that is earned by a particular product. There are essentially two approaches used to address this problem, reflecting the trade-off between flexibility and tractability in modeling the market shares as a function of the prices which we discuss below:

- A flexible model refers to a model which can mimic the mechanism of the black box well for different sets of input and output, by learning from data. Machine learning tools have become popular recently due to its modeling flexibility, compared to the traditional econometric methods. For instance, neural networks are well known for their capability to fit a black box system, given a large enough amount of training samples. However, the relationships between the input (prices) and output (sales/market share) described by the neural networks are complicated (essentially a composition of several nonlinear functions), and lack theoretical justification. Furthermore, the pricing optimization based on the market response function described by neural networks appears to be intractable.

- There is an alternative stream of research that focuses on identifying tractable instances of the pricing problem using discrete choice models. Most of these results are applicable to specific choice models, such as the Multinomial Logit (MNL) model (see Song and Xue [39], Dong et al. [14]) and its generalization to attraction models (see Keller et al. [27]), the NL model with identical price-sensitivity parameters for products within a nest (see Li and Huh [29]) and under some restrictions to non identical price-sensitivity parameters (see Gallego and Wang [21]), the d-level NL model (see Huh and Li [25], Li et al. [28]), the Paired Combinatorial Logit (see Li and Webster [30]), the Generalized Extreme Value (GEV) model (see Zhang et al. [42]) and the exponomial choice model (see Alptekinoğlu and Semple [4]).

Model misspecification is however a general concern in these models. For instance, in the estimation of NL model, the products in the different nests and the levels of the nesting structure
need to be pre-determined, and often involves a fair amount of ingenuity and understanding of the consumer market to arrive at an appropriate model. A maximum likelihood method is then typically used to estimate the parameters of the choice model and prices are set using the proposed choice model. A natural concern is what happens if the model is misspecified? We use next a small simulated example to illustrate the challenge of model misspecification for this class of pricing problems.

**Example 1.** We generate a set of sales data with the underlying choice model being a NL model with 4 branches and 5 products in each branch. For a given set of parameters in the NL model, we uniformly generate a set of prices and calculate its corresponding market shares using the assumed NL choice model (the details of the parameters are provided in Section 5.1.2). We estimate this data set using MNL choice model and optimize the prices based on the estimated MNL model. The true profit generated by the obtained pricing solution can be evaluated by calculating the choice probability using the assumed NL choice model. In contrast, we can also get the true optimal profit by solving the pricing with NL model, which is a convex optimization model as shown by Li and Huh [29]. We change the parameters in the NL model and repeat the procedure above to get a set of profits generated by the MNL based approach and the corresponding true optimal profits. We plot the kernel density functions of the two sets of profits in Figure 1.

When the price sample size is 200 points, there is a huge gap between the profits obtained from the MNL based approach and the true optimal profits. More importantly, this gap can not be eliminated by simply increasing the sample size to even 2000 points. In other words, there exists a systematic error if the underlying choice model is misspecified.

To resolve the problem of model misspecification, more sophisticated choice models such as the random coefficient logit has been proposed in the marketing and economics literature that relaxes the assumption that all consumers are identical and allows for consumer heterogeneity. Estimation of the random coefficient logit model however becomes much more complicated. Berry et al. [9] develop a two-step estimation method to estimate the random coefficient logit model from the aggregate sales data set. Solving the pricing problem with the random coefficient logit model is however an outstanding open challenge that still needs to be resolved. 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, in a study of the pricing problem in an bilevel retailing example, Harsha et al. [23] found that the following are common business rules that affect the pricing decisions: (a) Volume constraint - prices set must ensure that the sales target for certain products are met; (b) Price monotonicity constraint - prices in certain channels must be set at a discount of the prices offered in other
channels, for market positioning purposes; and (c) Price bounds - new prices must be confined to within certain range, as drastic changes in price levels may turn away customers. Harsha et al. (23) use attraction demand models to characterize customers’ demand and propose a general pricing model to handle these side constraints in the pricing problem. Mixed integer linear programs are proposed to solve these pricing problems.

In this paper, we restrict our attention to a recently proposed class of choice models which uses the properties of the marginal distributions of random utilities and identify conditions under which the pricing problem is computationally tractable. We show how this modeling perspective turns the estimation of choice model to that of estimation of marginal distribution functions, and develop a semiparametric method that partially addresses model misspecification and consumer heterogeneity considerations. Based on the estimation result, we propose a mixed integer programming based pricing model, which is able to incorporate different side constraints on the prices due to business strategies and reduces to a linear program in the absence of side constraints.

The rest of the paper is organized as follows: In Section 2 we provide a literature review of discrete choice models with a focus on pricing and estimation. In Section 3 we shows the connection between a special instance of the representative consumer model used in choice modeling, and the Marginal Distribution Model (MDM) introduced in Natarajan et al. (38). We also provide a robust interpretation of the model in this section. We use this connection to
obtain a general closed-form relationship between the prices and market shares, and show that
the pricing problem becomes convex and polynomial time solvable when the marginal density
functions are log-concave. This nicely extends some of the currently known results for the single
product pricing problem to the multi-product pricing problem. In Section 4, we develop the
“Marginal Estimation + Price Optimization” approach for pricing with aggregate sales data. We
discuss computational experiments to demonstrate the performance of the proposed approach
in Section 5 using both simulated data and industry data from an automobile manufacturer and
a fast food company.

2 Literature Review

There is now a significant amount of literature on the multi-product pricing problem using
discrete choice models of customer demand. In this section, we review some of the key results
that are relevant to the current paper.

2.1 Pricing with Choice Models

Myerson (34) provides a complete description of how a seller should choose a price to maximize
the expected profit when selling a single item to buyers whose valuations the seller is uncertain
about. Assuming that \( w \) is the cost of the product and \( F(p) \) is the seller’s assessment of the
probability that a buyer has a valuation less than or equal to a price \( p \), the optimal mechanism
is to sell the item at a fixed take-it-or-leave-it price \( p^* \) where:

\[
p^* = \arg \max_{p \geq 0} (p - w)(1 - F(p)). \tag{1}
\]

Alternatively, this problem can be reformulated in the choice probability (market share) variable
\( x = 1 - F(p) \) where the optimal market share \( x^* \) is the solution to:

\[
x^* = \arg \max_{x \in [0, 1]} (F^{-1}(1 - x) - w)x, \tag{2}
\]

and the optimal price is \( p^* = F^{-1}(1 - x^*) \). Log-concavity of the complementary distribution
function \( 1 - F(\cdot) \) implies quasi-concavity of the profit function in the price variable in (1) and
concavity of the profit function in the market share variable in (2) which makes this single
dimension optimization problem straightforward to solve (see Bagnoli and Bergstrom (6)). Ex-
amples of such distributions include the exponential, uniform, normal, logistic, extreme value,
Laplace distributions among others. Generalizing this result to multiple products is however
much more complicated.

Consider the multi-product pricing problem where the set of products offered by the seller is denoted as \( N = \{1, \ldots, N\} \) and \( \{0\} \) is the outside option. Let the price of product \( j \) be \( p_j \), the cost of product \( j \) be \( w_j \) where without loss of generality, the outside option parameters are assumed to be \( w_0 = p_0 = 0 \). The seller’s expected profit maximization problem is formulated as:

\[
\max_{\mathbf{p} \geq 0, p_0 = 0} \sum_{j=1}^{N} (p_j - w_j)x_j(\mathbf{p}),
\]

where \( \mathbf{p} = (p_0, p_1, \ldots, p_n) \) is the price vector and \( x_j(\mathbf{p}) \) is the market share of product \( j \) given the price vector \( \mathbf{p} \). In the simplest MNL choice model, the choice probabilities are given as:

\[
x_j(\mathbf{p}) = \frac{e^{v_j - \alpha p_j}}{1 + \sum_{k=1}^{N} e^{v_k - \alpha p_k}}, \quad \forall j = 1, \ldots, N,
\]

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

In model (4), the price-sensitivity parameter for each product is assumed to be identical. Then with probability \( x_0 = 1/(1 + \sum_{k \in N} e^{v_k - \alpha p_k}) \) the customer does not buy a product from the set. Hanson and Martin (22) showed that the multi-product profit function is unfortunately not a quasi-concave function of the price variables for the MNL model. Though a non-convex optimization problem, Akçay et al. (2) showed that the profit function is unimodal in the prices and demonstrated the efficiency of using the first order conditions in determining the optimal prices. An alternate approach to tackle this problem was developed by Song and Xue (39) and Dong et al. (14) who showed that pricing problem is a convex optimization problem for MNL in terms of the market share variables and reformulated the problem by optimizing over these variables rather than the prices. The pricing problem for MNL in terms of the choice probability variables is formulated as:

\[
\max_{\mathbf{x}} \sum_{j=1}^{N} (p_j(\mathbf{x}) - w_j)x_j
\]

s.t. \( \sum_{j=0}^{N} x_j = 1, \)

\[ x_j \geq 0, \quad \forall j = 0, 1 \ldots, N, \]
where the price as a function of choice probabilities is given as:

\[ p_j(x) = \frac{1}{\alpha} (v_j - \ln (x_j) + \ln (x_0)), \quad \forall j = 1, \ldots, N, \quad (6) \]

It is straightforward to verify that the objective function in (5) is concave in the choice probability variables when the inverse choice probability equation is given as (6) for the MNL model. The optimal prices for this model are known to have the equal markup property when the price sensitivities are identical.

Building on this technique, Li and Huh (29) studied the multi-product pricing problem under the NL model. In the NL model, the set of \( N \) products is divided into \( K \) nests where \( N_k \) denotes the number of products that belong to nest \( k \). Assuming that \( \tau_k \in [0,1] \) is a parameter describing the dissimilarity among products in nest \( k \), \( v_{jk} \) is the deterministic utility of product \( j \) in nest \( k \) (excluding the price), \( p_{jk} \) is the price of product \( j \) in nest \( k \) and \( \alpha_k \) is the price-sensitivity parameter for the products in nest \( k \), the customer’s probability of choosing product \( j \) in a nest \( k \) is given as:

\[
x_{jk}(p) = \frac{e^{v_{jk} - \alpha_k p_{jk}} \left( \sum_{l=1}^{N_k} e^{v_{lk} - \alpha_k p_{lk}} \right)^{\tau_k - 1}}{1 + \sum_{k=1}^{K} \sum_{l=1}^{N_k} e^{v_{lk} - \alpha_k p_{lk}}}, \quad \forall j = 1, \ldots, N, k = 1, \ldots, K.
\]

Li and Huh (29) showed that the multi-product pricing problem is a convex optimization problem in the market share variables in this case too. Gallego and Wang (21) showed that for product-differentiated price sensitivities, the convexity result for the NL model fails to hold and identified specific conditions on the price sensitivities and nest parameters for which the optimization problem remains convex in the choice probability variables.

Generalization of the pricing problem to a multi-level NL model has been recently studied in Li et al. (28) who developed an iterative algorithm to find a stationary point of the revenue function. The expected revenue is however not concave in the product prices for this case. Huh and Li (25) generalized these results to multi-stage attraction models and identified specific conditions under which the optimal prices can be found efficiently and characterize the optimal markup for each product. Li and Webster (30) studied the class of PCL models which allows for more general correlation structures as compared to the NL model and identify conditions under which an unique optimal price solution is computable. Zhang et al. (42) showed 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 (4) developed a
discrete choice model which they referred to as the exponomial choice model where the pricing problem was shown to be a convex optimization problem.

As these results indicate in comparison to the single product pricing problem for which convexity results are known for a large class of distributions, there are fewer convexity results known for the multi-product pricing problem and this is often obtained by careful analysis on a case by case basis. In this paper, we showcase an approach that can nicely exploit the generality of results from the single product case to the multi-product case.

2.2 Estimation of Choice Models

We review some of the key estimation techniques for discrete choice models in this section, starting with the MNL model. The most popular method used in the parametric estimation of choice models is the maximum likelihood method. The deterministic term in the utility specification captures part of the customer’s utility of product \( j \) and is represented as \( v_{ij} = \beta' z_{ij} \), where \( z_{ij} \) is the vector of product \( j \)’s attribute levels (excluding price) for customer \( i \) and \( \beta \) represents the customer’s partworth. Under the assumption of homogeneous partworth across all the customers with \( z_{i0} = 0, p_0 = 0 \), the maximum log-likelihood estimation of the parameters with the MNL choice model is given as:

\[
\max_{\beta, \alpha} \sum_{i=1}^{M} \sum_{j=0}^{N} y_{ij} \ln \left( \frac{e^{\beta' z_{ij} - \alpha p_j}}{1 + \sum_{k=1}^{N} e^{\beta' z_{ik} - \alpha p_k}} \right),
\]  

(7)

where \( M = \{1, \ldots, M\} \) is the set of customers and \( y_{ij} = 1 \) if customer \( i \) selects product \( j \) and 0 otherwise. The objective function in (7) is concave in the \( \beta \) and \( \alpha \) variables and the estimation problem is efficiently solvable (see McFadden (35)). One of the main restrictions of the MNL model is that it requires the error terms to be independent and identical. However, in practice, this assumption can be easily violated. The red bus-blue bus paradox is a classic example to illustrate this. To address this drawback, models such as the NL model have been proposed. Daganzo and Kusnic (13) and Mishra et al. (37) showed the maximum log-likelihood estimation of the NL model in the \( \beta \) and \( \alpha \) variables is solvable as a convex optimization problem when the parameters \( \tau_k \) are in \([0, 1]\) and fixed. The estimation problem is however non-convex when the \( \tau_k \) parameters also need to be estimated.

There has also been a significant interest in incorporating heterogeneity at the customer level in discrete choice models. One such model is the random coefficient logit choice model
(see McFadden and Train (36)) where estimation is typically carried out with simulation based optimization methods. However, the estimation problem for the mixed logit model for customer level data is non-convex. Berry et al. (9) consider aggregate sales data set and propose a two-step estimation method to address both partworth heterogeneity and price endogeneity. Under the homogeneous population assumption, the market share equals the individual choice probability and hence the estimation of the choice probability directly applies to the market share. In the case that the customers are heterogeneous in preference weights, the market share is the average of customers’ choice probability. In the first step, they propose to invert the market share function by solving a system of equations using iterative methods. In the second step, they estimate the distribution of the random coefficients using General Method of Moments (GMM) and use instrumental variables to address the price endogeneity issue. Another popular approach to estimation is to combine maximum likelihood estimation model with regularization technique adapted from machine learning to penalize deviations of customer specific partworths $\beta_i$ from the population average partworth $\beta$. The reader is referred to Evgeniou et al. (16) for a MNL model that incorporates heterogeneity at the partworth level using such an approach.

There has also been growing interest in semiparametric and nonparametric estimation methods. Building on the pioneering work of the maximum score estimator of Manski (32), Fox (19) showed that the semiparametric maximum-score estimator of choice models is consistent when using data only on subsets of choices. Farias et al. (17) proposed a nonparametric approach to estimate the choice model from a set of automobile sales transaction data. In their approach, the choice model is viewed as distributions over product’s rankings. Compared to parametric methods, which rely on the correctness of the assumed underlying choice model, semiparametric and nonparametric models make much fewer assumptions. Such approaches have however been primarily limited to the context of descriptive analytics rather than prescriptive analytics such as pricing problems.

The paper most closely related to our work is the class of choice models proposed by Nataraajan et al. (38) who develop a Marginal Distribution Model (MDM) where only the marginal distributions are specified but the joint distribution is not. By focusing on the joint distribution that maximizes customer’s expected utility, Mishra et al. (37) showed that MDM is able to recover a large number of widely used choice probabilities such as the MNL model and NL model by appropriately chosen marginal distributions. Mishra et al. (37) developed a scale heterogeneity version of this model that allows for customers to have different perception variances of the outside option by allowing for non-identical marginal distributions. Furthermore, they proposed a constrained maximum-likelihood estimation method for estimating the parameters.
of MDM with disaggregate data. Computational tests on simulated and real data on preferences of safety features in automobiles in Mishra et al. (37) illustrate that the model is suitable for capturing both product and consumer level heterogeneity. Ahipasaoglu et al. (1) extended this model to compute traffic equilibrium flows in transportation networks and illustrated the modeling flexibility that capturing the marginal distributions provides in this context. We build on this model in the paper and apply it to solve multi-product pricing problems.

3 Multi-Product Pricing with MDM

The multi-product pricing problem can be viewed as a bilevel optimization problem where in the outer step, the retailer sets the price for each product while in the inner step, the customers observe the prices and then make purchasing decisions. We use market shares as the metric to model the market response to price changes. A flexible model of the market shares is expected to fit the prices and sales data well in many cases but results in a more complicated estimation and pricing optimization procedure. Model misspecification in such a case might arise from a “wrong guess” of the underlying choice model generating the sales data. Our key contribution in this paper is to provide a model which can achieve modeling flexibility and computational tractability simultaneously.

The proposed model is built on a representative consumer who captures the stochastic choice behavior. Let \( v_{ij} \) denotes a consumer \( i \)'s deterministic valuation for product \( j \) and \( p_j \) denotes the price of product \( j \). Then \( v_{ij} - \alpha p_j \) represents consumer \( i \)'s surplus on product \( j \) where \( \alpha > 0 \) is a homogeneous price sensitivity parameter across products. We assume that \( x_{ij} \) denotes consumer \( i \)'s randomization strategy which is the probability of consumer \( i \) choosing product \( j \). The classical random utility framework uses a random noise \( \tilde{\epsilon}_{ij} \) to model choice behaviour, assuming that customer \( i \) will choose the product with the largest utility where the utilities are given by:

\[
U_{ij} := v_{ij} - \alpha p_j + \tilde{\epsilon}_{ij}.
\] (8)

The representative consumer model uses however a convex regularization term \( C(x) \) to serve as a reward to the consumer’s randomization strategy \( x \) as follows:

\[
\max_{x \in \Delta_N} \sum_{j=0}^{N} (v_{ij} - \alpha p_j) x_{ij} - C(x)
\] (9)
where the optimization is over the unit simplex defined as:

\[ \Delta_N = \left\{ \mathbf{x} \in \mathbb{R}_{+}^{N+1} \mid \sum_{j=0}^{N} x_{ij} = 1 \right\}, \tag{10} \]

Such a representative consumer model has been studied in Anderson et al. (3) and is known to recreate the MNL choice probabilities when \( C(\mathbf{x}) = \sum_i x_i \log x_i \) is the entropy function. Hofbauer and Sandholm (24) showed that for any random utility model when the error terms have a strictly positive density function, there exists a strictly convex regularization function \( C(\mathbf{x}) \), such that the solution to the representative consumer model provides the choice probability in the random utility model. However it is also known that the representative consumer model is not equivalent to the random utility model. For example, there is no random utility model which is equivalent to the representative consumer model when \( C(\mathbf{x}) = -\sum_i \log x_i \) (see Proposition 2.2 in Hofbauer and Sandholm (24)). Fudenberg et al. (20) recently studied a special case of the representative consumer model under the assumption that the regularization term is separable as follows:

\[
\max_{\mathbf{x} \in \Delta_N} \sum_{j=0}^{N} (v_{ij} - \alpha p_j) x_{ij} - \sum_{j=0}^{N} C_j(x_{ij}), \tag{11}
\]

where each \( C_j(\cdot) \) is a convex function and showed that in this case, the choice probabilities satisfy a weaker form of the IIA property. However they do not discuss applications of their model to pricing problem which is our focus in this paper.

In fact, for each differentiable convex function \( C_j(\cdot) : [0, 1] \rightarrow \mathbb{R} \), there exists a valid cumulative probability distribution \( F_j(\cdot) \) such that \( C_j(x_{ij}) = -\int_{1-x_{ij}}^{1} F_j^{-1}(t)dt \). In the following analysis, we focus on this particular representation of the separable convex function, turning our attention from a general convex function to a probability distribution function. This representation of the regularization function is motivated from the Marginal Distribution Model (MDM) proposed by Natarajan et al. (38). Their model is built on characterizing the choice probabilities for the extremal distribution in the set of all joint distributions with the given marginal distributions of the error terms that maximizes expected consumer utility (see Natarajan et al. (38) and Theorem 1 in Mishra et al. (37)). Specifically, we solve the following convex

\[ F_j(t) = 1 - C_j^{-1}(-t), \]

where \( C_j'(\cdot) \) denotes the derivative of function \( C_j(\cdot) \) and \( C_j^{-1}(\cdot) \) denotes the inverse of function \( C_j(\cdot) \). The convexity of \( C_j(\cdot) \) ensures the monotonicity of its first derivative, i.e. \( -C_j'(1-x_{ij}) \) is nondecreasing in \( x_{ij} \). Additionally as \( C_j(\cdot) \) is a function mapping from \([0,1]\) to \( \mathbb{R} \), the inverse of the derivative function maps to \([0,1]\), indicating \( F_j(\cdot) \) is a valid distribution function.
optimization problem to compute the choice probabilities:

$$\max_{x \in \Delta_N} \sum_{j=0}^{N} \left( (v_{ij} - \alpha p_j)x_{ij} + \int_{1-x_{ij}}^{1} F_j(t)dt \right),$$  \hspace{1cm} (12)$$

Natarajan et al. (38) have shown that for a fixed pricing solution $p$, MDM can recover the whole generalized extreme value (GEV) family by properly defining the marginal distribution functions. Mishra et al. (37) showed that all the choice probabilities in the relative interior of a simplex can be recreated by MDM under appropriate assumptions on the marginal distributions.

We note that there is an alternate interpretation of the choice probabilities in the MDM model from a robust optimization perspective. To this end, we show that a modeler who assumes that the utilities of the product lies in an uncertainty set and uses the worst-case realizations to estimate the choice probabilities can recreate the MDM choice probabilities under an appropriately choice of the uncertainty set. We consider a convex uncertainty set that is defined from the marginal distributions of the error terms as follows:

$$U(\delta) = \left\{ \epsilon \mid \sum_{j=0}^{N} \mathbb{E}F_j[\tilde{\epsilon}_{ij} - \epsilon_{ij}]^+ \leq \delta \right\}. \hspace{1cm} (13)$$

In this set, if $\delta_2 \geq \delta_1$, then clearly $U(\delta_1) \subseteq U(\delta_2)$. Consider a robust optimization problem of the following form:

$$\max_{x \in \Delta_N} \min_{\epsilon \in U, \delta \geq 0} \sum_{j=0}^{N} (v_{ij} + \epsilon_{ij} + \delta)x_{ij},$$  \hspace{1cm} (14)$$

where the modeler estimates the choice probabilities assuming that the error terms for each product and the budget is chosen by nature so as to put larger weights on the products chosen with lower probability. For larger values of $\delta$, the worst-case utility will be smaller while for smaller values of $\delta$, the worst-case utility will be larger. We incorporate the budget decision variable $\delta$ in to the objective function so as to capture this tradeoff and to ensure that the solution will not be too conservative. As we shall see next in the next theorem, the proposed robust choice model (14) is exactly equivalent to the convex optimization formulation of MDM.

**Theorem 1.** The optimal $x$ variables in the solution to the robust optimization problem (14) for the uncertainty set $U$ in (13) is exactly the choice probabilities in the convex optimization formulation of MDM in (12).

The proof of Theorem 1 is a direct application of convex optimality conditions and is provided in Appendix A. The result is closely related to the recent work of Fudenberg, Iijima and Strzalecki (20) who develop an additive perturbed nonlinear utility model which is related to
MDM. Furthermore, they show that the choice corresponds to an ambiguity averse preference of a modeler who is uncertain about the true utility and uses a regularization term in making the choices. In Theorem 1, we explicitly construct an uncertainty set and show that choice probabilities in MDM can be viewed as the solution to a robust optimization problem, thus providing an alternative perspective on discrete choice models, beyond random utility theory.

3.1 Pricing Model

We now consider the pricing model under MDM. Clearly, (12) is a convex optimization problem. We assume the error term is independent of the price of the products, i.e. \( F_j(\cdot), \forall j = 0, 1, \ldots, N \) is not a function of \( p \). The optimality condition of (12) then yields

\[
p_j^* = \frac{v_{ij} + F_j^{-1}(1 - x_j^*) - F_0^{-1}(1 - x_0^*)}{\alpha}, \quad \forall j = 1, 2, \ldots, N.
\]

This provides a wide class of closed-form relationship between prices and choice probabilities. With a slight abuse of notation, we use \( F_j(\cdot) \) to denote the marginal distribution function of the generalized random error term \( v_{ij} + \tilde{\epsilon}_{ij} \), and assume that the deterministic valuation \( v_{ij} = 0 \) in our model. In this case the relationship between price and market share can be written as

\[
p_j^* = \frac{F_j^{-1}(1 - x_j^*) - F_0^{-1}(1 - x_0^*)}{\alpha}, \quad \forall j = 1, 2, \ldots, N.
\]

and the choice probability vector is the solution to the convex optimization problem:

\[
x^* = \arg \max_{x \in \Delta_N} \sum_{j=0}^N \left( -\alpha p_j x_j + \int_{1-x_j}^1 F_j^{-1}(t) \, dt \right),
\]

We plug the closed-form solution between prices and market shares to the seller’s pricing model in (5) to obtain an equivalent reformulation of the multi-product pricing problem with the market share decision variables \( x \) as follows:

\[
\max_x - \sum_{j=1}^N w_j x_j + \frac{1}{\alpha} \sum_{j=1}^N x_j F_j^{-1}(1 - x_j) - \frac{1}{\alpha} (1 - x_0) F_0^{-1}(1 - x_0)
\]

\[
\text{s.t.} \quad \sum_{j=0}^N x_j = 1,
\]

\[
x_j \geq 0, \quad \forall j = 0, 1, \ldots, N.
\]

This brings us to the following result in Theorem 2, which provide conditions that guarantee that the pricing problem (16) is tractable (see Appendix A).

14
Theorem 2. Assume that the following two conditions hold:

C1. $x F_j^{-1}(1 - x)$ is a concave function in $x$ for each $j = 1, \ldots, N$.

C2. $x F_0^{-1}(x)$ is a convex function in $x$.

Then, the pricing problem (16) is a convex optimization problem in the market share $x$ variables and the optimal prices are computable in polynomial time. Moreover, if the optimal solution is $x^*$, then the optimal pricing strategy is

\[ p_j^* = \frac{F_j^{-1}(1 - x_j^*) - F_0^{-1}(1 - x_0^*)}{\alpha}, \quad \forall j = 1, 2, \ldots, N. \]  

(17)

In fact, Condition 1 and 2 are known to be fairly mild conditions as we see in Proposition 1 and Corollary 1.

Proposition 1. Let $F_j(x)$ for $j = 0, \ldots, N$ be the marginal distributions. Then,

(i) Function $x F_j^{-1}(1 - x)$ for $j = 1, \ldots, N$ is concave if and only if function $\frac{1}{1 - F_j(x)}$ for $j = 1, \ldots, N$ is convex.

(ii) Function $x F_0^{-1}(x)$ is convex if and only if function $\frac{1}{F_0(x)}$ is convex.

Let $F_j(x) = 1 - F_j(x)$ for $j = 1, \ldots, N$. Then, we have the following result.

Corollary 1. Conditions C1 and C2 hold if the marginal distributions satisfy the following conditions: (i) The tail distribution $F_j(\cdot)$ for $j = 1, \ldots, N$ is log-concave; (ii) The distribution $F_0(\cdot)$ is log-concave. Both the conditions (i) and (ii) are satisfied when the probability density function $f_j(\cdot)$ for $j = 0, \ldots, N$ is log-concave.

Since log-concavity is satisfied by many common probability distributions (see Bagnoli and Bergstrom (6)), such as the normal, logistic, exponential, extreme value and Laplace distributions, this result identifies a large class of problems for which the pricing problem is now tractable. We show in Appendix B that by defining appropriate marginal functions, the method is able to recover the pricing result for MNL.

By considering the optimality condition of the pricing model (16), we can further characterize the optimal markup of each product. The optimality condition implies

\[ -w_j + \frac{1}{\alpha} F_j^{-1}(1 - x_j^*) + \frac{1}{\alpha} x_j^* F_j^{-1}(1 - x_j^*) + \lambda = 0 \]

\[ \frac{1}{\alpha} F_0^{-1}(1 - x_0^*) - \frac{1}{\alpha} (1 - x_0^*) F_0^{-1}(1 - x_0^*) + \lambda = 0 \]  

(18)

Combining this with (17), we get the following characterization of the markup.
Proposition 2. The optimal markup of each product is characterized by the following:

\[ p_j^* - w_j = \frac{1}{\alpha} \frac{x_j^*}{F_j^{-1}(1-x_j^*))} + \frac{1}{\alpha} \frac{1-x_0^*}{F_0^{-1}(1-x_0^*)}, \]  

(19)

where \( x^* \) is the vector of choice probabilities from MDM.

This brings us to the following corollary which recovers the equal markup property of the MNL Model.

Corollary 2. If the utility of each product shares the same hazard rate \( \frac{1-F_j(U)}{F_j(U)} \), then the optimal markup of each product \( p_j^* - w_j \) is the same.

For exponential marginal distributions, the hazard rates are the same implying that the optimal markup is the same. For more general marginal distributions, however the optimal markup will not be the same under this choice model.

4 Data-Driven Price Optimization with MDM

A key part in optimizing the prices from data is to calibrate the consumer’s choice behavior. To mitigate the model misspecification issue, ideally we would try to estimate the convex regularization term, which would lead to the underlying representative consumer model. However, the regularization function is multi-dimensional, making it hard to estimate and furthermore there is no guarantee that the corresponding pricing problem is tractable. To achieve a balance between the estimation accuracy and price optimization tractability, we turn our estimation of the general convex function to that of a separable convex function. As shown in the previous section, this approximation can represent a family of common choice models used in practice and at the same time ensures the tractability of the price optimization problem. We denote this method as “Marginal Estimation + Price Optimization”. In this section, we will show how to estimate the marginal distribution function from aggregate sales data and then solve the pricing problem by solving a linear program (LP) and/or mixed integer linear program (MILP). We assume that the data is given as follows: There are \( M \) periods of sales data, which provides the prices and market shares of \( N \) products and the market share of the outside option. We denote the data as \( \{p_{tj}, x_{tj}\}_{t=1,...,M, j=0,...,N} \) where without loss of generality we assume that \( p_{t0} = 0 \) for all \( t \).
4.1 Estimating Marginals

The main idea of the estimation technique is to use the closed-form solution of the price in (17) in terms of the market shares to fit the sales data \( \{ p_{tj}, x_{tj} \}_{t=1,\ldots,M, j=0,\ldots,N} \). The goal is to estimate a set of valid marginal distribution functions such that the deviation of the fitted prices from the observed prices is minimized as follows:

\[
\min_{\alpha > 0, F_j^{-1} \text{increasing}} \sum_{j=1}^{N} \sum_{t=1}^{M} \left( \frac{F_j^{-1}(1-x_{tj})}{\alpha} - \frac{F_0^{-1}(1-x_{t0})}{\alpha} - p_{tj} \right)^2, \tag{20}
\]

This estimation model is inspired by one of the most commonly used nonparametric estimation methods in shape-constrained estimation. This method is favored due to several reasons. First, it is free of tuning parameters and requires fewer assumptions, which provides it with model identifying power (see Matzkin (33)). Secondly, incorporating the shape restrictions provides desirable sample properties of the estimator (see Beresteanu (7)) and achieves a high prediction accuracy under small sample sizes. Interested readers are referred to Chen (12) and Xu (41) for a detailed introduction of shape-constrained estimation. In this paper, we restrict ourselves to one special shape-constrained estimation model—additive isotonic regression, which was first proposed by Bacchetti (5). The estimation model considered in their work is as follows:

\[
\min_{c, f_l: \text{monotonic}} \sum_{i=1}^{M} (Y_i - c - f_1(X_i^1) - \cdots - f_L(X_i^L))^2, \tag{21}
\]

where \( (X_i^1, X_i^2, \ldots, X_i^L, Y_i) \) is the observed data for \( i = 1, \ldots, M \). Mammen and Yu (31) analyzed rates of convergence of the estimator of this model and demonstrated finite sample properties of the estimator through simulation experiments. It is clear that (20) is a straightforward application of (21), with two additive functions \( F_j^{-1}(1-x_{ij}) \) and \( -F_0^{-1}(1-x_{i0}) \) for each \( j = 1, \ldots, N \).

To solve the optimization problem, we modify (20) by defining \( y_j(x) := x F_j^{-1}(1-x)/\alpha \) for \( j = 1, \ldots, N \), \( y_0(x) := (1-x) F_0^{-1}(1-x)/\alpha \). Denote the estimated function value of \( y_j(\cdot) \) at point \( x_{tj} \) as \( y_{tj} \) for \( t = 1, \ldots, M, j = 0, \ldots, N \). Then \( F_j^{-1}(1-x_{tj})/\alpha = y_{tj}/x_{tj} \) and \( F_0^{-1}(1-x_{t0})/\alpha = y_{t0}/(1-x_{t0}) \). We add in the monotonicity shape constraints according to the natural monotonicity structure of \( F_j(\cdot) \). Then the estimation model can be written as follows:

\[
\min_{y \in \mathcal{M}} \sum_{j=1}^{N} \sum_{t=1}^{M} \left( \frac{y_{tj}}{x_{tj}} - \frac{y_{t0}}{1-x_{t0}} - p_{tj} \right)^2, \tag{22}
\]
where $\mathcal{M}$ is defined by a set of shape (monotonicity) constraints:

$$
\mathcal{M} = \left\{ y \bigg| \begin{array}{c}
\frac{y_{tj}}{x_{tj}} \leq \frac{y_{t'j}}{x_{t'j}} \text{ for all } (t, t') \text{ such that } x_{tj} \geq x_{t'j} \text{ for } j = 1, \ldots, N \\
\frac{y_{00}}{1-x_{00}} \leq \frac{y_{00}}{1-x_{00}} \text{ for all } (t, t') \text{ such that } x_{00} \geq x_{00}
\end{array} \right\}.
$$

In addition, to calibrate the marginal distribution functions with limited experiment data, we can add additional structure on the functional form. It has been shown in Corollary 1 that log-concavity of the density function implies conditions C1 and C2 which make the pricing problem computationally tractable and is satisfied by many common probability distributions. Hence we can also incorporate the convexity conditions from this in our estimation model. In summary, we add the following four constraints in our estimation model.

(i) $xF_j^{-1}(1-x)$ for $j = 1, \ldots, N$ is concave in the $x$ variable,

(ii) $(1-x)F_0^{-1}(1-x)$ is convex in the $x$ variable,

(iii) $F_j^{-1}(1-x)$ for $j = 1, \ldots, N$ is a monotone decreasing function in the $x$ variable,

(iv) $F_0^{-1}(1-x)$ is a monotone decreasing function in the $x$ variable.

We enforce these monotonicity and convexity conditions in our model as constraints. To do this, we sort for each product $j$, the market shares in the ascending order. We denote the corresponding indices of the sorted data set for a given $j$ as $s_j = (s_{j1}, \ldots, s_{jM})$. Then the constraints (i)-(iv) can be added to the estimation model (22) as linear constraints. Formally, the shape constraints we consider in our estimation problem are defined as

$$
\mathcal{S} = \left\{ y \bigg| \begin{array}{c}
(1 - \eta_{tj})y_{s_{t+1}j} + \eta_{tj}y_{s_{t-1}j} \leq y_{s_{t}j}, \quad \forall t = 2, \ldots, M-1, j = 1, \ldots, N \\
(1 - \eta_{0})y_{s_{t+1}0} + \eta_{0}y_{s_{t-1}0} \geq y_{s_{t}0}, \quad \forall t = 2, \ldots, M-1 \\
y_{s_{t}j} \leq y_{s_{t-1}j}, \quad \forall t = 2, \ldots, M, j = 1, \ldots, N \\
y_{s_{t}0} \leq \frac{y_{s_{t-1}0}}{1-x_{s_{t-1}0}}, \quad \forall t = 2, \ldots, M.
\end{array} \right\}
$$

where $\eta_{tj} = (x_{s_{t+1}j} - x_{s_{t}j})/(x_{s_{t+1}j} - x_{s_{t-1}j})$. Then the estimation problem can be modeled as a second order conic program (SOCP), which can be efficiently solved as follows:

$$
\min_{y, \phi} \sum_{t=1}^{M} \sum_{j=1}^{N} \phi_{tj} \\
\text{s.t.} \quad \left( \frac{y_{tj}}{x_{tj}} - \frac{y_{00}}{1-x_{00}} - p_{tj} \right)^2 \leq \phi_{tj}, \quad \forall t = 1, \ldots, M, j = 1, \ldots, N, \\
y \in \mathcal{S}
$$
It is also possible to use other estimators by minimizing the $L_1$ norm instead of $L_2$ norm, which turns the estimation problem to a linear program as follows:

$$\min_{y, \phi} \sum_{t=1}^{M} \sum_{j=1}^{N} \phi_{tj}$$

s.t.

$$\frac{y_{tj}}{x_{tj}} - \frac{y_{0j}}{1 - x_{0j}} - p_{tj} \leq \phi_{tj}, \quad \forall t = 1, \ldots, M, j = 1, \ldots, N,$$

$$-\left(\frac{y_{tj}}{x_{tj}} - \frac{y_{0j}}{1 - x_{0j}} - p_{tj}\right) \leq \phi_{tj}, \quad \forall t = 1, \ldots, M, j = 1, \ldots, N,$$

$$y \in S$$

The novelty of the estimation method under this choice model is that we estimate the marginal distribution functions rather than joint distributions by taking advantage of closed-form relationship between prices and market shares. The fitting of the marginal distribution functions from the aggregate sales data leads us to an appropriate choice model for the pricing problem, which partially overcomes the model misspecification issue that might arise by assuming a fixed structural form of the choice model as we show in the numerical experiments.

4.2 Optimizing Prices

With a set of fitted values $y_{tj}$, we now consider the problem of optimizing prices. Using the definition of $y_j(\cdot)$ from the previous section, the objective function in the pricing optimization model (16) is essentially given by the summation of all the $y_j(\cdot)$ functions. Therefore, it is natural to consider a piecewise linear approximation of the objective function. Define piecewise linear function value $PF(x; s; f)$ as follows:

$$PF(x; s; f) = \left\{ y = \sum_{t=1}^{M} \lambda_t f_t \mid \begin{array}{l}
\lambda_1 \leq z_1, \quad \lambda_t \leq z_{t-1} + z_t, \forall t = 2, \ldots, M - 1, \\
\lambda_M \leq z_{M-1}, \quad z_t \in \{0, 1\}, \forall t = 1, \ldots, M - 1, \\
\sum_{t=1}^{M-1} z_t = 1, \quad x = \sum_{t=1}^{M} \lambda_t s_t, \\
\end{array} \right\},$$

where the first argument $x$ denotes the independent variable at which we need to compute the function value. The second argument $s$ is the ordered input market share data and the third argument $f$ indicates the corresponding function values. 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 $f_t$ values as the $x$ value is a convex combination of the $s_t$ values. Specifically, if $x = \lambda_t s_t + \lambda_{t+1} s_{t+1}$, with $\lambda_t + \lambda_{t+1} = 1$, $\lambda_t, \lambda_{t+1} \geq 0$, then $y = \lambda_t f_t + \lambda_{t+1} f_{t+1}$. The binary
variable $z$ is introduced to indicate the interval in which the point is located in. Thus, $PF(x; s; f)$ is a singleton, whose value denotes the piecewise approximation of the $y$ value at $x$. Define $·/\cdot$ as the element wise division between two vectors. Then the price optimization problem can be formulated in the following manner:

$$
\Pi := \max_{x, \delta, F_1} - \sum_{j=1}^{N} w_j x_j + \sum_{j=1}^{N} \delta_j - \delta_0 \\
\text{s.t.} \quad \delta_j \in PF(x_j; x_s; y_s), \quad \forall j = 1, \ldots, N, \\
\delta_0 \in PF(x_0; x_s; y_s) \\
F_{I_j} \in PF(x_j; x_s; y_s \cdot x_s), \quad \forall j = 1, \ldots, N, \\
F_{I_0} \in PF(x_0; x_s; y_s \cdot (1 - x_s)) \\
F_{I_j} - F_{I_0} \in \Omega_j, \quad \forall j = 1, \ldots, N, \\
\sum_{j=0}^{N} x_j = 1 \\
x_j \leq x_{s,j}, \quad \forall j = 0, \ldots, N, \\
x_j \geq x_{s,j}, \quad \forall j = 0, \ldots, N, \\
x \geq 0,
$$

where the first constraint provides the piecewise linear approximation of the function $x_j F_j^{-1}(1 - x_j)/\alpha$ at $x_j$, the second constraint provides the approximation of $(1 - x_0) F_0^{-1}(1 - x_0)/\alpha$ at $x_0$. Similarly, function inverse value $F_{I_j}, j = 0, \ldots, N$ provides the approximation of $F_j^{-1}(1 - x_j)/\alpha$ at $x_j$. Hence the optimal price $p_j$ can be represented as $F_{I_j} - F_{I_0}$. We encapsulate all the price constraints in $\Omega_j, j = 1, \ldots, N$. For example, this may include bound constraints on the price $p_j$, e.g. $u_j \leq p_j \leq \bar{u}_j$. 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 $\Omega_j$ is described through linear and integer constraints, this problem is solvable as a mixed integer linear program.

Figure 2 provides a small example to illustrate the intuition of the piecewise linear approximation method (with $N = 2, M = 10$). The estimation model provides a pointwise estimation of

$$
\frac{x_j F_j^{-1}(1 - x_j)}{\alpha}, \forall j = 1, \ldots, N \quad \text{and} \quad \frac{(1 - x_0) F_0^{-1}(1 - x_0)}{\alpha}
$$

at each sample point. For any market share $x^*$, we can use a convex combination of the function values

$$
\frac{x_j F_j^{-1}(1 - x_j)}{\alpha} \quad (\text{resp.} \quad \frac{(1 - x_0) F_0^{-1}(1 - x_0)}{\alpha})
$$
Figure 2: Illustration of piecewise linear approximation of profit function

at two adjacent sample points containing $x_j^*$ to approximate

$$\frac{x_j^* F_j^{-1}(1-x_j)}{\alpha} \quad (resp. \quad \frac{(1-x_j^*) F_0^{-1}(1-x_j)}{\alpha}),$$

hence the profit value at $x^*$ can be represented in a linear form $\sum_{j=1}^N \delta_j^* - \delta_0^*$. The binary variable $z$ is introduced to indicate which interval $x_j^*$ locates in. Similarly, $\frac{F_j^{-1}(1-x_j^*)}{\alpha}$ is approximated by the same convex combination of the value $\frac{F_j^{-1}(1-x_j)}{\alpha}$ at the two two adjacent sample points containing $x_j^*$, which is denoted as $FI$ in (26). Then $FI_j - FI_0$ provides the approximated price $p_j, \forall j = 1, \ldots, N$ at the given market share $x^*$.

Finally, it is worth observing that if $\Omega_j$ only includes nonnegativity constraints, the optimization model can be further simplified as a linear program by taking advantage of the structure of the marginal distribution functions. Notice that for a concave function, a linear approximation of the function value at point $x_j$ can be calculated as $\min y_{s_j} + (y_{s_{j+1}} - y_{s_j})(x_{s_{j+1}} - x_{s_j})(x_j - x_{s_j})$ and for a convex function, the corresponding approximated value at point $x_0$ is obtained from $\max y_{s_0} + (y_{s_{1}} - y_{s_0})(x_{s_{1}} - x_{s_0})(x_0 - x_{s_0})$. Therefore, the optimization problem
can be modeled as the following linear programming problem:

\[
\Pi := \max_{x, \delta} \left( -\sum_{j=1}^{N} w_j x_j + \sum_{j=1}^{N} \delta_j - \delta_0 \right)
\]

s.t. \[
\delta_j \leq y_{s_{t,j}} + \frac{y_{s_{t+1,j}} - y_{s_{t,j}}}{x_{s_{t+1,j}} - x_{s_{t,j}}} (x_j - x_{s_{t,j}}), \quad \forall t = 1, \ldots, M - 1, j = 1, \ldots, N,
\]
\[
\delta_0 \geq y_{s_{0}} + \frac{y_{s_{1},0} - y_{s_{0}}}{x_{s_{1},0} - x_{s_{0}}} (x_0 - x_{s_{0}}), \quad \forall t = 1, \ldots, M - 1,
\]
\[
\sum_{j=0}^{N} x_j = 1
\]
\[
x_j \leq x_{s_{t,j}}, \quad \forall j = 0, \ldots, N,
\]
\[
x_j \geq x_{s_{t,j}}, \quad \forall j = 0, \ldots, N,
\]
\[
x \geq 0.
\]

**4.3 Uniqueness of the Optimal Prices and Profit Functions**

Note that in both (24) and (25), there are \(M(N + 1)\) variables \(y_{t,j}\) to be estimated including the outside option, but there are only \(MN\) constraints from the optimality conditions. There could hence be multiple solutions that can satisfy these set of conditions, leading to issues in model identification. This is because all the marginal distributions \(F_j(\cdot)\)'s can be scaled and shifted by a constant, without affecting the market share attained by each product. We show next that all the optimal solutions to this problem will generate the same optimal market shares and optimal prices when the samples are suitably generated and the sample size goes to infinity. This provides a theoretical justification on the uniqueness of the optimal prices and profit under this approach.

To see this, we denote the optimal solution in the estimation model which leads to the true underlying marginal distribution function as \(y^*_{t,j}, \forall j = 1, \ldots, N; t = 1, \ldots, M\). Notice that every \(\mathbf{y}\) satisfying \(p_{t,j} = \frac{y_{t,j}}{x_{t,j}} - \frac{y_{0,n}}{1-x_{s_{0}}^n}\) is a feasible solution to the estimation model. We can without loss of generality write the solutions as \(y_{t,j} = y^*_{t,j} + \delta_{t,j}\) for \(j = 0, 1, \ldots, N\) and \(t = 1, \ldots, M\), where \(\frac{\delta_{t,j}}{x_{s_{t,j}}} = \frac{\delta_{t,0}}{1-x_{s_{0}}^n}\), for any \(j = 1, \ldots, N\) and \(t = 1, \ldots, M\). Consider an arbitrary market share \(\mathbf{x}^*\) and assume \(x^*_j\) is located between \(x_{s_{t,j},j}\) and \(x_{s_{t+1,j},j}\), for \(j = 0, 1, \ldots, N\), and
\[ x_j^* = \lambda_j^* x_{s_j^*} + (1 - \lambda_j^*) x_{s_{j+1}} \]

Then the price of product \( j \) is given as:

\[
p_j^* = \lambda_j^* \frac{y_{t_j^*}}{x_{s_j^*} + 1} + (1 - \lambda_j^*) \frac{y_{t_{j+1}}}{x_{s_{j+1}}} - \left( \lambda_j^* \frac{y_{t_j^*}}{1 - x_{s_j^*} + 1} + (1 - \lambda_j^*) \frac{y_{t_{j+1}}}{1 - x_{s_{j+1}} + 1} \right)
\]

\[
= \lambda_j^* \frac{y_{t_j^*}}{x_{s_j^*}} + (1 - \lambda_j^*) \frac{y_{t_{j+1}}}{x_{s_{j+1}}} - \left( \lambda_j^* \frac{y_{t_j^*}}{1 - x_{s_j^*}} + (1 - \lambda_j^*) \frac{y_{t_{j+1}}}{1 - x_{s_{j+1}}} \right)
\]

\[
\lambda_j^* \frac{\delta_{s_j^*}}{x_{s_j^*}} + (1 - \lambda_j^*) \frac{\delta_{s_{j+1}}}{x_{s_{j+1}}} - \left( \lambda_j^* \frac{\delta_{s_j^*}}{1 - x_{s_j^*}} + (1 - \lambda_j^*) \frac{\delta_{s_{j+1}}}{1 - x_{s_{j+1}}} \right)
\]

\[
 \lambda_j^* \Delta_{s_j^*} + (1 - \lambda_j^*) \Delta_{s_{j+1}} - \left( \lambda_j^* \Delta_{s_j^*} + (1 - \lambda_j^*) \Delta_{s_{j+1}} \right)
\]

\[
(28)
\]

The last two terms indicate the price deviation from the true price under \( x^* \). Let \( \frac{\delta_{s_j}}{x_{s_j}} = \Delta \), for any \( j = 1, \ldots, N \). The price deviation can be rewritten as

\[
\lambda_j^* \left( \frac{\delta_{s_j^*}}{x_{s_j^*}} - \frac{\delta_{s_{j+1}}}{x_{s_{j+1}}} \right) + (1 - \lambda_j^*) \left( \frac{\delta_{s_{j+1}}}{x_{s_{j+1}}} - \frac{\delta_{s_{j+2}}}{x_{s_{j+2}}} \right)
\]

\[
\left( \lambda_j^* \frac{\delta_{s_j^*}}{1 - x_{s_j^*}} + (1 - \lambda_j^*) \frac{\delta_{s_{j+1}}}{1 - x_{s_{j+1}}} \right)
\]

\[
= \lambda_j^* \Delta_{s_j^*} + (1 - \lambda_j^*) \Delta_{s_{j+1}} - \left( \lambda_j^* \Delta_{s_j^*} + (1 - \lambda_j^*) \Delta_{s_{j+1}} \right)
\]

\[
(29)
\]

We now claim that under the following assumption, the price deviation in (29) goes to 0.

**Assumption (*)**

1. The pricing experiments are generated such that for each market share outcome \( x \), there is another experiment with market share \( x' \) such that \( \max_{j=0, \ldots, N} |x_j - x'_j| \leq \epsilon \).

2. The inverse of the marginal distribution function has a bounded first derivative within the range of data:

\[
\max_j \max_{x \in [\min_j x_j, \max_j x_j]} |(F_j^{-1})'(1 - x)| \leq D \text{ for some } D > 0.
\]

The assumptions hold when the data sample is carefully generated and the sample size is sufficiently large with strictly positive market shares for each experiment. We first show in Lemma [1] that under these assumptions, the deviations of \( y \) in different experiments are bounded.

**Lemma 1.** Under Assumption (*), \( |\Delta_i - \Delta_i| \leq D \epsilon \) for any \( i \neq t \).

Then we are ready to claim
Proposition 3. Under Assumption (*), in the case $\epsilon \to 0$, any optimal solution $y$ to (24) or (25) will generate the same optimal market shares $x^*$, optimal price $p^*$ and the optimal profit $\Pi^*$ in the pricing problem (26).

Proposition 3 guarantees the uniqueness of our approach. Although there might be multiple solutions in the proposed estimation model, they all lead to the same optimal market shares and optimal prices when the experiment is suitably designed and the sample size goes to infinity.

4.4 Approximating the General Representative Consumer Model

One of the limitations in the proposed approach is the assumption that the function $C(x)$ is separable so that the pricing problem can be solved using the MDM approach. To extend this technique to handle general representative consumer model, we use the following trick: Let $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ be i.i.d noise with mean 0. We can model the local structure of the function $C(x)$ by looking at

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

and hence

$$E[C(x + \epsilon)] \approx C(x) + \sum_{i} \frac{\partial^2 C}{\partial x_i^2} E[\epsilon_i^2].$$

Instead of assuming that the general function $C(\cdot)$ is separable, we assume that the local structure of $C(\cdot)$ at the point $x$ can be represented by a separable function, and perform pricing experiment around a small region of $x$ to calibrate the corresponding choice function using the MDM approach. This allows us to optimize over the region, and find a search direction that will guide us to a better solution.

We propose an adaptive search procedure 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. The details of the adaptive sample generating procedure are shown in Table 1. The proposed approach also allows us to do validation by dividing the data into two sets - in-sample data and out-sample data. Apply (25) to the in-sample data to get pointwise estimates of the utility’s marginal distribution functions. With these estimates, we solve the following validation model.
Table 1: 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 estimation (25) and optimization (26) model to get an optimal price as the starting price, denoted as \( p^* \).

Step 3. Regenerate a set of prices \( p^* \), with a relatively small interval. (e.g. uniformly generate prices within the range \((p^*(-5\% \sim -1\%), p^*(1\% \sim 5\%))\) and obtain the market shares.

Step 4. Apply estimation (25) and optimization (26) model to get a new optimal price \( \hat{p}^* \). Evaluate the profit generated by \( \hat{p}^* \), denote the profit as \( \hat{\Pi}^* \). If \( \hat{\Pi}^* \geq \Pi(p^*) \), let \( p^* = \hat{p}^* \), otherwise, keep \( p^* \). Go to Step 3.

\[
\Pi := \min_{x, F, \phi} \sum_{j=1}^{N} \phi_j \\
\text{s.t. } FI_j - FI_0 - p^*_j \leq \phi_j, \quad \forall j = 1, \ldots, N, \\
-(FI_j - FI_0 - p^*_j) \leq \phi_j, \quad \forall j = 1, \ldots, N, \\
FI_j \in PF(x_j; x_s; y_s \cdot /x_s), \quad \forall j = 1, \ldots, N, \quad (30) \\
FI_0 \in PF(x_0; x_s; y_s \cdot /y_s) \\
\sum_{j=0}^{N} x_j = 1 \\
x_j \geq 0, \quad \forall j = 0, \ldots, N.
\]

The main idea of the prediction model (30) is to find the best piecewise linear approximation of the estimated points such that the fitted price is closest to the observed price in \( L_1 \) norm.

5 Computational Experiments

In this section, we provide numerical experiments to demonstrate the performance of the “Marginal Estimation + Price Optimization” framework. In the first example, we present a simulation experiment with the underlying choice model to be MNL and NL choice models to showcase the performance of our framework. In the second example, we use experiment data set provided by an automobile company to optimize the prices. Using our approach, the profit improves by close to 7.54% compared to the base price, which shows the practical usefulness of our method. In the last example, we apply this framework to approximate the optimal prices.
obtained from a mixed logit approach using data from a fast food company. As we show, the “marginal estimation+price optimization” approach using only price and market share data (without knowledge of the mixed logit model) provides a close approximation to a quadratic programming model which precisely knows the sensitivities arising from a mixed logit model.

5.1 Performance with underlying MNL and NL

5.1.1 Pricing with MNL Model

In Section 3, we saw that MDM can recover the MNL choice model under the assumption that the marginal distribution function is independent of the prices using an exponential distribution. In this section, we provide computational evidence to illustrate this. Suppose the underlying choice model is MNL choice model. We observe the prices and shares of 20 products to apply the estimation and pricing method to fit the data and optimize the set of prices. We evaluate the profit of the obtained prices using the true choice model and get the profit performance. On the other hand, we can solve the pricing problem with MNL model exactly to get the true optimal profit, assuming that we know the true MNL parameters. We compare the gap between the two profits. We repeated this with 100 different MNL choice models and compare the empirical kernel density function of the 100 profits under the two approaches. Figure 3 lists the comparison between true optimal profit and the profit obtained from our data-driven approach. The samples are increased based on the adaptive sample generating approach. i.e., we use 5 random samples to fit the model in each iteration. Interestingly, the approach converges to the true optimal prices when the sample size increases to around 50 experiments (i.e., after 10 iterations), without assuming that the underlying choice model is MNL!

5.1.2 Pricing with Nested Logit Model

We repeat the experiment, but this time the simulated data is generated based on the Nested Logit choice model, whose pricing problem is known to be a convex optimization problem and hence we are able to get the true optimal profit. Assume there are $N = 20$ products. We divide the products into $K = 4$ branches where each branch has 5 products. Consider the customers utility model of product $j$ in branch $k$ as follows:

$$v_{jk} - p_{jk} + \tilde{\epsilon}_{jk},$$

with the marginal distribution function of the error term $\tilde{\epsilon}_{jk}$ to be $F_{jk}(\epsilon) = 1 - e^{-\epsilon} \left( \sum_{j=1}^{N_k} e^{v_{jk} - p_{jk}} \right)^{\tau_k - 1}$ for $\epsilon \geq (\tau_k - 1) \ln(\sum_{j=1}^{N_k} e^{v_{jk} - p_{jk}})$, where $\tau_k \in [0, 1]$ is the parameter representing the dissimilar-
ity among products in branch $k$. Note that for MDM, this choice of the marginal distribution recreates the Nested Logit choice model.

We randomly generate 100 $(v, \tau)$ with the entries in $v$ from 10 to 15 and the entries in $\tau$ 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 does not vary too much, for each $v$ and $\tau$, we generate the sales data in the following way: Each price instance is generated after randomly selecting an integer number $l$ from 1 to 10, and then each $p_i$ is uniformly sampled within 10% range of $0.2lv_i$, i.e. $(0.2lv_i \times 0.9, 0.2lv_i \times 1.1)$. For price vector $p$, the corresponding choice probability is calculated as follows:

$$x_{jk} = \frac{e^{v_{jk} - p_{jk}} \left( \sum_{l=1}^{N_k} e^{v_{lk} - p_{lk}} \right)^{\tau_k - 1}}{1 + \sum_{t=1}^{K} \left( \sum_{l=1}^{N_k} e^{v_{lt} - p_{lt}} \right)^{\tau_t}}. \quad (31)$$

We generate 200 scenarios of such sales data $(p, x)$ for each $(v, \tau)$.

In the following, we apply our “Marginal Estimation + Price Optimization” model to the simulated data to see its performance. As a benchmark, we use MNL model to estimate the data and further get an optimal price by solving the pricing with MNL model. We evaluate each pricing proposal according to the generated true profit, which can be calculated by applying $(31)$ to get its corresponding true market share. Note that the estimation of MNL model is solvable with a maximum likelihood estimation model using individual choice data. However, since we
only have aggregate share information, we instead adopt a least error estimation model\(^2\) using our marginal estimation model - we minimize the weighted absolute value of the error term in the MNL estimation model, which is formulated as follows:

\[
\begin{align*}
\text{Objective:} & \quad \min_{\hat{\mathbf{v}}, \mathbf{\phi}} \sum_{t,j} \phi_{tj} \\
\text{Subject to:} & \quad \phi_{tj} \geq \frac{1}{p_{tj}}(\hat{v}_j + \ln(1 - \sum_{j=1}^{N} x_{tj}) - \ln(x_{tj}) - p_{tj}), \forall t = 1, \ldots, M, j = 1, \ldots, N, \\
& \quad \phi_{tj} \geq -\frac{1}{p_{tj}}(\hat{v}_j + \ln(1 - \sum_{j=1}^{N} x_{tj}) - \ln(x_{tj}) - p_{tj}), \forall t = 1, \ldots, M, j = 1, \ldots, N, \\
& \quad \hat{v} \geq 0.
\end{align*}
\]

Once the \(\hat{v}\) variables are estimated, we compute the optimal price \(p^*_\text{MNL}\). According to the true values of \(v\) and \(\tau\), we can calculate the true choice probability using (31). The performance of \(p^*_\text{MNL}\) is evaluated based on the true profit generated. Denote the corresponding profit under \(p^*_\text{MNL}\) as \(\Pi^*_\text{MNL}\). Next we apply the estimation model (25) and optimization model (27) introduced in Section 4 to estimate the data and obtain another price proposal, denoted as \(p^*_\text{MDM}\), whose corresponding profit under the true model is denoted as \(\Pi^*_\text{MDM}\). We compare these two different methods using the performance metric as the deviation from the true optimal profit, denoted as \(\epsilon_\text{G}\). The true optimal profit can be calculated by solving (??) under the true parameter \((v, \tau)\). Then we can define our performance metric as follows:

\[
\epsilon_\text{G} = \frac{\Pi^* - \Pi^*_G}{\Pi^*},
\]

where \(G\) refers to MNL and MDM respectively. Notice we have generated 100 different sets of \((v, \tau)\). The average deviation is shown in Table 2. With 200 sample data, the prices from MNL model can only achieve a profit level that is 18.3\% of the true optimal one while the “Marginal Estimation + Price Optimization” model only deviates from the true optimal profit by 2.83\%. The result, on the one hand, demonstrates the model specification performance of our data-driven pricing framework. On the other hand, it shows us the importance of the calibration of marginal distribution functions in optimizing the price from sales data since we did not get a good price if we start by assuming a wrong choice model (MNL versus NL).

We apply the adaptive sample generating procedure introduced in Section 4.4 to test the performance as sample size increases. Applying the procedure 10 times, in other words, we enlarge the sample size by 10 times. As the iterations increase, more experiment sample is generated

\(^2\)See [8] for a discussion on choice model estimation using aggregate share data.
Table 2: Mean deviation of different methods

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon_{MDM}$</th>
<th>$\epsilon_{MNL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0283</td>
<td>0.1832</td>
</tr>
</tbody>
</table>

and we plot the kernel density estimation function of the generated profits from MDM model in each iteration in Figure 4. From the figure we can see more experiment data results in a more accurate profit kernel density, indicating a better estimation of the underlying choice model. After around 5 rounds of the adaptive sampling procedure (sample size increases from 200 to 1000), MDM almost generates the same kernel density as the true one which performs much better than MNL based approach (see Figure 1).

Figure 4: Kernel density function estimate of the profits under MDM versus the true optimal profit.

To demonstrate the importance of the adaptive procedure, we compare the performance using the adaptive procedure to one where the prices were generated randomly. As mentioned above, the adaptive sampling procedure converges within 5 rounds, i.e., with 1000 experiments. We compare the price obtained this way, to one obtained by solving the problem directly using a random sample size of 1000. The comparison is shown in Figure 5, from which we see the adaptive procedure indeed helps to generate better solutions using the same number of samples, and this is consistent for both MNL and MDM models.
5.2 Application on Automobile Data

In this section, we apply our method to a set of data provided by an automobile manufacturer. The company has developed an elaborate market share simulation model to evaluate the performance of different pricing proposals, taking into account competitor’s pricing strategy and outside options available. In each simulation experiment, the model changes the price of one product to one of 4 treatment levels: $-10\%$, $-5\%$, $5\%$, $10\%$. There are 20 products in the pricing problem, and hence there are 81 experiment data in total including the base price case. The base prices and cost information are shown in Table 3. The current pricing solution yields a profit of $182.2444$, whereas the best solution from the 80 set of prices obtained from local perturbation returns a profit of $185.4426$.

The experiment data set contains price and market share for each product as well as the outside market share. The underlying question is whether we can use these observations to propose a better pricing decision for this problem?

We first demonstrate the estimation method in Section 4.1 can provide a good estimation of the underlying choice model in the given data set. We divide the data set into two parts, in-sample data and out-sample data. Since there are only 81 samples available, we do the validation in the following way. We use 80 data points (in-sample data) to calibrate the model and predict
Table 3: Base Price and Cost in the Data Set

<table>
<thead>
<tr>
<th></th>
<th>Brand</th>
<th>Base Price (thousand)</th>
<th>Cost (thousand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>45.5209</td>
<td>34.1407</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>46.906</td>
<td>35.1795</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>45.056</td>
<td>33.792</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>43.8742</td>
<td>32.9057</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>47.0282</td>
<td>35.2711</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>42.0442</td>
<td>31.5331</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>45.5225</td>
<td>34.1419</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>39.455</td>
<td>35.5095</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>37.955</td>
<td>34.1595</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>33.1821</td>
<td>29.8639</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>27.255</td>
<td>24.5295</td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>35.9266</td>
<td>32.334</td>
</tr>
<tr>
<td>13</td>
<td>B</td>
<td>33.5504</td>
<td>30.1953</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>37.9559</td>
<td>34.1595</td>
</tr>
<tr>
<td>15</td>
<td>C</td>
<td>39.750</td>
<td>33.7875</td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>35.809</td>
<td>30.4376</td>
</tr>
<tr>
<td>17</td>
<td>C</td>
<td>38.227</td>
<td>32.4933</td>
</tr>
<tr>
<td>18</td>
<td>C</td>
<td>41.250</td>
<td>35.0625</td>
</tr>
<tr>
<td>19</td>
<td>C</td>
<td>36.0612</td>
<td>30.652</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
<td>39.750</td>
<td>33.7875</td>
</tr>
</tbody>
</table>
Table 4: Prediction Error in the Cross Validation

| \(|\hat{\Pi} - \Pi_0| / \Pi_0\) | MNL | MDM |
|-------------------------------|-----|-----|
| mean                          | 16.5% | 1.46% |
| std                           | 0.2098 | 0.0112 |

the market share under the remaining experiment (out-sample data) using the validation model \((30)\). We repeat this experiment 50 times. In each experiment, the out-sample data is randomly generated. As a basis for comparison, we repeat the experiments, assuming now the underlying choice model of the data set is an MNL model. We next apply \((32)\) to calibrate the model and apply the validation procedure described above to get another 50 values of prediction. We plot the predicted market share under different methods in Figure 6, from which we can see the MDM indeed help to predict the market share more accurately, compared to the MNL model. We also evaluate the prediction error from the profit perspective. Denote the profit of the out-sample data point under the predicted market share as \(\hat{\Pi}\) and the true profit in the data as \(\Pi_0\). We use the difference \(\frac{|\hat{\Pi} - \Pi_0|}{\Pi_0}\) to evaluate the prediction error in one validation experiment.

Figure 6: Predicted market share in 50 experiments

The comparison of the prediction error using the two estimation methods (MNL and MDM) is shown in Table 4, from which we can see the MDM based estimation model provides a much better estimation of the underlying choice model in a given data set.
Table 5: Optimal market share and prices

<table>
<thead>
<tr>
<th>Product</th>
<th>Optimal market share</th>
<th>Optimal Price (thousand)</th>
<th>Current Price (thousand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00113018</td>
<td>41.0182</td>
<td>45.5209</td>
</tr>
<tr>
<td>2</td>
<td>0.00172837</td>
<td>44.6084</td>
<td>46.906</td>
</tr>
<tr>
<td>3</td>
<td>0.00168842</td>
<td>40.6064</td>
<td>45.056</td>
</tr>
<tr>
<td>4</td>
<td>0.00222398</td>
<td>41.7288</td>
<td>43.8742</td>
</tr>
<tr>
<td>5</td>
<td>0.00214648</td>
<td>44.728</td>
<td>47.0282</td>
</tr>
<tr>
<td>6</td>
<td>0.00177148</td>
<td>39.9887</td>
<td>42.0442</td>
</tr>
<tr>
<td>7</td>
<td>0.00194276</td>
<td>41.0301</td>
<td>45.5225</td>
</tr>
<tr>
<td>8</td>
<td>0.000810995</td>
<td>43.3973</td>
<td>39.455</td>
</tr>
<tr>
<td>9</td>
<td>0.000573912</td>
<td>39.8781</td>
<td>37.955</td>
</tr>
<tr>
<td>10</td>
<td>0.0014093</td>
<td>34.8571</td>
<td>33.1821</td>
</tr>
<tr>
<td>11</td>
<td>0.00103533</td>
<td>29.9785</td>
<td>27.255</td>
</tr>
<tr>
<td>12</td>
<td>0.00157825</td>
<td>39.5193</td>
<td>35.9266</td>
</tr>
<tr>
<td>13</td>
<td>0.00160609</td>
<td>36.3603</td>
<td>33.5504</td>
</tr>
<tr>
<td>14</td>
<td>0.000891123</td>
<td>39.8749</td>
<td>37.9559</td>
</tr>
<tr>
<td>15</td>
<td>0.000573276</td>
<td>39.7495</td>
<td>39.750</td>
</tr>
<tr>
<td>16</td>
<td>0.00122113</td>
<td>35.8082</td>
<td>35.809</td>
</tr>
<tr>
<td>17</td>
<td>0.00148839</td>
<td>40.1564</td>
<td>38.227</td>
</tr>
<tr>
<td>18</td>
<td>0.000859814</td>
<td>43.3357</td>
<td>41.250</td>
</tr>
<tr>
<td>19</td>
<td>0.00131773</td>
<td>37.8818</td>
<td>36.0612</td>
</tr>
<tr>
<td>20</td>
<td>0.000907923</td>
<td>39.7442</td>
<td>39.750</td>
</tr>
<tr>
<td>Outside</td>
<td>0.973095</td>
<td>Profit: 0.1959796</td>
<td>Profit: 0.1822444</td>
</tr>
</tbody>
</table>

Building on the estimation result, we apply our “Marginal Estimation + Price Optimization” framework 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 prices changes are required to be within 10% of the current prices, i.e., the pricing constraint set is supposed to be

$$\Omega_j = \{0.9p^c_j \leq p_j \leq 1.1p^c_j\},$$

where $p^c_j$ denote the current price of product $j$. To incorporate these bounds on prices, we apply our MIP model (26) in Section 4.2 to get the optimal prices. The optimization model provides a new price proposal shown in Table 5 which achieves $195.9796$ in profit, a 7.54% improvement compared to the current pricing solution. More interestingly, our approach suggests a natural strategy for this pricing problem - the company needs to decrease aggressively the prices for Brand A products, increase aggressively the prices for Brand B, while maintaining the current price levels for products in Brand C!
5.3 Application on Fast Food Data

In this section, we consider a related application from a fast food company which includes 53 products divided into 4 categories. The first category has 14 types of entrees and 7 types of small sized Extra Value Meals (EVM). The second category has 11 types of medium sized EVM and 11 types of large sized EVM respectively. The third category has 6 types of happy meal while the last category has 4 types of sides. Each consumer is assumed to choose at most one product within a category (or an outside option), but may choose more than one product from among different categories. For each category, the company provided historical monthly sales data from April 2016 to March 2017, the prices of the products during this time period and also an estimate of the outside market share.

The current approach used by the company is to estimate a mixed logit choice model using the classical BLP approach (see Berry, Levinsohn, and Pakes (9)) with customer demographic data and to calculate the price elasticity of the demand from this model. Once the demand price elasticities are estimated, it is possible to build a price optimization model from the base prices by solving a quadratic program (QP) as follows:

\[
\max_{\Delta P} \sum_i \left( Q_i + \sum_j \frac{Q_j}{P_j} \beta_{ij} \Delta P_j \right) (P_i + \Delta P_i),
\]

where \(\beta_{ij}\) are the price elasticity coefficients, \(P\) denote the base price and \(Q\) is the corresponding market share for the base price. Note that the QP model in general is not a convex optimization problem. In addition to the nonconvexity of the objective function in the QP model, complicated side business constraints often need to be included. We list some of the main business constraints in Table 6. While a majority of the constraints can be modeled in a straightforward manner using linear constraints, constraints 2 and 10 can be modeled using linear constraints with integer variables.

We now use the “marginal estimation + price optimization” model to solve this problem assuming that only price-market share data is available. We conduct the pricing experiment in the following way. We uniformly generate new prices within the range of 3% of the base price to get 12 experiment prices at each iteration. The market share for any new experiment price is calculated from the base prices and base market shares in conjunction with the price elasticity provided by the BLP model. We perform 10 iterations of the algorithm using the constraints in Table 6 and calculate the profit based on the obtained market share. Note that under the “marginal estimation + price optimization”, the explicit knowledge of the price sensitivities is not assumed to be known unlike the QP model. Assuming only linear constraints,
Table 6: Price Constraints

1. Overall store price (weighted by qty) increase by 2%
2. Price change by increment of $0.05
3. Price increase cap at min($0.30, round down nearest whole number of 10% of original price). Absolute price decrease cap at min($0.20, round down nearest whole number of 5% of original price).
4. EVM must be lower in price than sum of its ALC components
5. BVM must be lower in price than sum of its ALC components.
6. EVM must be higher in price than Burger + Coke.
7. BVM must be higher in price than Burger + Hashbrown.
8. Small EVM to Medium EVM must have a $0.65 upcharge.
9. Medium EVM to Large EVM must have a $0.65 upcharge.
10. Ranking of prices of entrees same as ranking of prices of Small/Medium/Large EVM.
11. Price per unit (oz./kJ/piece) of Drinks/Fries/Nuggets must be decreasing with size/quantity.
12. Smaller size items must be cheaper than larger size items.
14. Medium EVM consist of Entree, Medium Fries, Medium Drinks.
15. Large EVM consist of Entree, Large Fries, Large Drinks

we compare the profits generated from the QP and the experiments with the base profit. The profit comparison result is show in Figure 7. Our approach achieves 5.13% improvement from the base profit whereas the QP approach achieves 5.16% improvement, indicating our approach performs fairly well even without knowing the price sensitivities in the pricing model directly. We next consider incorporating all the constraints where the pricing solution suggested by our approach is provided in Figure 8. Compared to the base price, we basically suggest to increase the prices for entree and decrease the prices for small sized EVM. We can keep the the current

Figure 7: Profits in each iteration with only linear constraints
pricing decision for the medium sized and large sized EVM. Lastly, we increase the price for sides a little bit. The profit in each iteration is shown in Figure 8 from which we can see the iterative approach almost converges after 5 iterations and we achieve profit improvement around 5.1% from the base price.

Figure 8: Price solution by MDM versus base price

Figure 9: Profits in each iteration with all the constraints
6 Conclusion

We develop a “Marginal Estimation + Price Optimization” framework for multi-product pricing problem. This framework provides a novel approach to multi-product pricing problem, and exploits the properties of marginal distributions of random utilities in discrete choice models. This marginal distribution based perspective makes the pricing problem convex for a large family of distributions—as long as the marginal density functions are log-concave. On the other hand, the marginals can be estimated from data using a second order conic program or a linear program, which are both computationally efficiently. With the estimated marginals, the pricing problem can be solved using a linear program or a mixed-integer linear program (when there are side constraints on prices). The advantage in computations makes our framework attractive in the data-driven settings. Numerous computational results also justify the model specification performance of this framework. Interesting application of this technique on two sets of industry data indicates that the framework produces reasonable recommendation, even in the presence of consumer heterogeneity, which is a common concern with pricing using aggregate sales data.

It is worthwhile to mention that although we have presented a data-driven approach to solve the multi-product pricing problem, we believe that such technique has implications to more general optimization problems with choice behavior. In the case of product assortment, the technique can be extended to handle the estimation of marginals when the companies can experiment with the type of assortments offered. In transport planning, the technique can be extended to deal with road toll pricing and how route choices are affected. We leave these issues for future research.

Acknowledgment

The research of the first and last author are supported by the Institute of Operations Research and Analytics, NUS. The research of the third 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?’.

References


38


Appendix A

Theorem 1

Proof: The robust optimization problem in [14] is reformulated as:

$$\max_{x \in \Delta_N} \min_{\epsilon} \sum_{j=0}^{N} (v_{ij} x_{ij} + \epsilon_j x_{ij} + E_F [\tilde{\epsilon}_ij - \epsilon_j])^+, \quad (34)$$

since at optimality we must have $\delta = \sum_j E_F [\tilde{\epsilon}_ij - \epsilon_j]^+$. For each feasible $x \in \Delta_N$, the inner minimization problem is separable in the $\epsilon_{ij}$ variables. Hence, (34) can be reformulated as:

$$\max_{x \in \Delta_N} \sum_{j=0}^{N} \left( v_{ij} x_{ij} + \min_{\epsilon_{ij}} \left( \epsilon_{ij} x_{ij} + E_F [\tilde{\epsilon}_ij - \epsilon_j] \right) \right). \quad (35)$$

The $j$th inner problem is a convex minimization problem in one variable with the optimal value $\hat{\epsilon}_{ij}$ satisfying:

$$x_{ij} - 1 + F_j(\hat{\epsilon}_{ij}) = 0,$$

where:

$$\hat{\epsilon}_{ij} = F_j^{-1}(1 - x_{ij}).$$

Note that for $x_{ij} = 0$, the $j$th inner problem has objective 0 with the optimal $\hat{\epsilon}_{ij} = \infty$ and for $x_{ij} = 1$, the $j$th inner problem has objective $E_F [\tilde{\epsilon}_ij]$ with the optimal $\hat{\epsilon}_{ij} = -\infty$. Plugging back into the objective function, we get the $j$th inner minimization problem is equivalent to:

$$\hat{\epsilon}_{ij} x_{ij} + E_F [\tilde{\epsilon}_ij - \hat{\epsilon}_{ij}]^+ = x_{ij} F_j^{-1}(1 - x_{ij}) + \int_{F_j^{-1}(1 - x_{ij})}^{\infty} (\epsilon_j - F_j^{-1}(1 - x_{ij})) dF_j(\epsilon_j),$$

$$= \int_{F_j^{-1}(1 - x_{ij})}^{\infty} \epsilon_j dF_j(\epsilon_j),$$

$$= \int_{1-x_{ij}}^{F_j^{-1}(t)} dt, \quad (36)$$

which provides the desired result.

Theorem 2

Proof: Denote $h_j(x) = x F_j^{-1}(1 - x), \forall j = 1, \ldots, N$, $h_0(x) = (1 - x) F_0^{-1}(1 - x)$ and $g_j(x) = h'_j(x), \forall j = 0, \ldots, N$. Let $\mu$ denote the dual variable for the equality constraint in formulation
The optimality conditions of (16) yields

\[ v_j - \alpha w_j + h'_j(x_j) + \mu = 0, \forall j = 1, \ldots, N, \]
\[ -h'_0(x_0) + \mu = 0. \]

(37)

Reorganize (37) to get

\[ x_j = g^{-1}_j(\alpha w_j - v_j - \mu), \forall j = 1, \ldots, N, \]
\[ x_0 = g^{-1}_0(\mu). \]

(38)

Since at optimality, \( x \) also satisfies the feasibility condition \( \sum_j x_j = 1 \), we can calculate the optimal dual variable \( \mu \) by solving the nonlinear equation

\[ G(\mu) := \sum_{j=1}^{N} g^{-1}_j(\alpha w_j - v_j - \mu) + g^{-1}_0(\mu) = 1. \]

(39)

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 \( \mu \). Therefore, we can solve the nonlinear equation (39) by a line search method in the range

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

With the optimal dual variable \( \mu \), the optimal market share \( x \) is obtained from (38).

\[ \square \]

**Proposition 1**

**Proof:** For the smooth inverse function \( F^{-1}(x) \), the derivative is given as \( (F^{-1})'(x) = 1/F'(F^{-1}(x)) \).

Denote \( h(x) := x F^{-1}(1-x) \). Then,

\[ h'(x) = F^{-1}(1-x) + x(F^{-1}(1-x))', \]
\[ = F^{-1}(1-x) - \frac{x}{F'(F^{-1}(1-x))}. \]

(40)

Calculating the second derivatives,

\[ h''(x) = -\frac{2}{F'(F^{-1}(1-x))} - \frac{x F''(F^{-1}(1-x))}{(F'(F^{-1}(1-x)))^3}, \]
\[ = -\frac{2 F'(F^{-1}(1-x))^2 + x F''(F^{-1}(1-x))}{[F'(F^{-1}(1-x))]^3}. \]

(41)
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

$$2F'(F^{-1}(1-x))^2 + xF''(F^{-1}(1-x)) = 2F'(y)^2 + (1 - F(y))F''(y) \geq 0, \forall y.$$  \hspace{1cm} (42)

On the other hand, we have:

$$
\left( \frac{1}{1 - F(y)} \right)' = \frac{F'(y)}{(1 - F(y))^2},
\left( \frac{1}{1 - F(y)} \right)'' = \frac{2F'(y)^2 + (1 - F(y))F''(y)}{(1 - F(y))^3}.
\hspace{1cm} (43)
$$

So $1/(1 - F(y))$ is convex if and only if inequality (42) holds. Result (i) has been proved.

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

$$
g'(x) = F^{-1}(x) + \frac{x}{F'(F^{-1}(x))},
g''(x) = \frac{2F'(F^{-1}(x))^2 - xF''(F^{-1}(x))}{[F'(F^{-1}(x))]^3}.
\hspace{1cm} (45)
$$

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

$$2F'(z)^2 - F(z)F''(z) \geq 0.
\hspace{1cm} (46)
$$

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

$$\left( \frac{1}{F(z)} \right)'' = \frac{2F'(z)^2 - F(z)F''(z)}{F(z)^3} \geq 0,
\hspace{1cm} (47)
$$

which is equivalent to (47).

**Corollary**

**Proof:** When the tail distribution $\bar{F}_j(x)$ for $j = 1, \ldots, N$ is log-concave, we have:

$$
(\ln(1 - F_j(y)))'' = -\frac{F_j'(y)^2 + (1 - F_j(y))F_j''(y)}{(1 - F_j(y))^2} \leq 0.
\hspace{1cm} (48)
$$

Hence:

$$0 \leq F_j'(y)^2 + (1 - F_j(y))F_j''(y) \leq 2F_j'(y)^2 + (1 - F_j(y))F_j''(y).
\hspace{1cm} (49)
$$

From the proof of Proposition 1, $xF_j^{-1}(1-x)$ is concave. (ii) can be shown using a similar procedure. The last statement follows from results in Bagnoli and Bergstrom (6).

$$\square$$
Lemma 1

Proof: Consider an arbitrary market share sample \( \mathbf{x}_i \), for any small number \( \epsilon \), there exists another sample \( \mathbf{x}_{i'} \) in a \( (L_{\infty}-\text{norm}) \) ball centered by \( \mathbf{x}_i \) with the radius \( \epsilon \). As the sum of all the products’ market share is 1, then there must exist \( j \) and \( j' \) such that \( x_{i,j} > x_{i',j} \) and \( x_{i,j'} < x_{i',j'} \).

WLOG, we assume neither \( j \) or \( j' \) is outside option. Monotonicity constraint implies

\[
\frac{y_{i,j}}{x_{i,j}} \leq \frac{y_{i',j}}{x_{i',j}}, \quad \frac{y_{i,j'}}{x_{i,j'}} \geq \frac{y_{i',j'}}{x_{i',j'}}
\]

We can also write them as

\[
\frac{y_{i,j}'}{x_{i,j}} + \Delta_i \leq \frac{y_{i',j}'}{x_{i',j}} + \Delta_i', \quad \frac{y_{i,j'}}{x_{i,j'}} - \Delta_i \geq \frac{y_{i',j'}}{x_{i',j'}} + \Delta_i'
\]

Therefore, \( \frac{y_{i,j}'}{x_{i,j}} - \frac{y_{i,j'}}{x_{i,j'}} \leq \Delta_i - \Delta_i' \leq \frac{y_{i',j}}{x_{i,j}} - \frac{y_{i',j'}}{x_{i',j'}} \). Let \( f_j(x) = F_j^{-1}(1 - x) \). Notice \( \frac{y_{i,j}'}{x_{i,j}} - \frac{y_{i,j'}}{x_{i,j'}} = f'(\phi x_{i,j} + (1 - \phi)x_{i',j'})(x_{i',j} - x_{i,j'}) \) for some \( \phi \in (0, 1) \). According to Assumption (*), \( |f'(\phi x_{i,j} + (1 - \phi)x_{i',j'})| \leq D \). Hence \( |\frac{y_{i,j}'}{x_{i,j}} - \frac{y_{i,j'}}{x_{i,j'}} - D|x_{i',j} - x_{i,j'}|, \) which implies \( \Delta_i - \Delta_i' \geq -D|x_{i',j} - x_{i,j'}| \). Similarly, we can get \( \Delta_i - \Delta_i' \leq D|x_{i',j} - x_{i,j'}| \). This implies \(-D\epsilon \leq \Delta_i - \Delta_i' \leq D\epsilon \). Therefore, \( \Delta_i - \Delta_i' \) goes to 0 when \( \epsilon \) goes to 0. 

\( \square \)

Proposition 3

Proof: Lemma 1 and (29) imply that multiple solution in the estimation model generates the same optimal price. Following a similar vein to the analysis of the price, the profit function under an arbitrary market share \( \mathbf{x}^* \) can be represented as

\[
\sum_{j=1}^{N} \left( \lambda_j y_{s_{t,j}} + (1 - \lambda_j) y_{s_{t+1,j}} \right) - \left( \lambda_0 y_{s_{t,0}} + (1 - \lambda_0) y_{s_{t+1,0}} \right)
\]

\[
= \sum_{j=1}^{N} \left( \lambda_j (y_{s_{t,j}} + \delta_{s_{t,j}}) + (1 - \lambda_j) (y_{s_{t+1,j}} + \delta_{s_{t+1,j}}) \right)
\]

\[
- \lambda_0 (y_{s_{t,0}} + \delta_{s_{t,0}}) + (1 - \lambda_0) (y_{s_{t+1,0}} + \delta_{s_{t+1,0}})
\]

The profit deviation is

\[
\sum_{j=1}^{N} \left( \lambda_j \delta_{s_{t,j}} + (1 - \lambda_j) \delta_{s_{t+1,j}} \right) - \left( \lambda_0 \delta_{s_{t,0}} + (1 - \lambda_0) \delta_{s_{t+1,0}} \right)
\]

\[
= \sum_{j=1}^{N} \left( \lambda_j \Delta_{s_{t,j}} x_{s_{t,j}} + (1 - \lambda_j) \Delta_{s_{t+1,j}} x_{s_{t+1,j}} \right)
\]

\[
- \left( \lambda_0 \Delta_{s_{t,0}} + (1 - \lambda_0) \Delta_{s_{t+1,0}} \right)
\]

The last equation holds since \( \frac{\delta_{s_{t,j}}}{x_{s_{t,0}}} = \frac{\delta_{s_{t,0}}}{x_{s_{t+1,0}}} = \Delta_i \). Under our assumption, \( \Delta_{s_{t,j}} = \Delta_{s_{t+1,j}} := \Delta \).
for any \( j = 0, 1, \ldots, N \), hence the profit deviation term becomes
\[
\Delta \left( \sum_{j=1}^{N} \left( \lambda_j x_{s t_j, j} + (1 - \lambda_j)x_{s t_{j+1}, j} \right) - \left( \lambda_0(1 - x_{s t_0, 0}) + (1 - \lambda_0)(1 - x_{s t_{0+1}, 0}) \right) \right)
\]
\[= \Delta \left( \sum_{j=1}^{N} x^*_j - (1 - x^*_0) \right) = 0 \]

In other words, although there are multiple \( y \) from the estimation model, the profit under each market share remains the same. Hence the optimization will provides the true market share.

\[\square\]

**Appendix B**

**Pricing with Multinomial Logit Choice Model**

Define \( F_j(\epsilon) = 1 - e^{-\epsilon} \) for \( \epsilon \geq 0 \) for \( j = 0, \ldots, N \). For \( \alpha = 1 \), from (16), the pricing problem under MDM with the exponential marginal distributions can be written as

\[
\max_{x} \quad \sum_{j=1}^{N} (v_j - w_j)x_j - \sum_{j=1}^{N} x_j \ln(x_j) + \left( \sum_{j=1}^{N} x_j \right) \ln \left( 1 - \sum_{j=1}^{N} x_j \right)
\]

s.t.
\[
\sum_{j=1}^{N} x_j \leq 1,
\]
\[
x_j \geq 0, \quad \forall j = 1, \ldots, N.
\]

It can be easily verified the tail function \( \bar{F}_j(\epsilon) \) is log-concave on \( \epsilon \) for each \( j \), and the marginal distribution function of the outside option \( F_0(\epsilon) \) is log-concave. From Corollary 1 and Theorem 2, the optimal pricing problem is a convex problem and the optimal pricing strategy is
\[
p_j = v_j + \ln \left( 1 - \sum_{j=1}^{N} x_j \right) - \ln x_j, \quad \forall j = 1, \ldots, N,
\]

That is exactly the optimal pricing function under MNL model shown in Proposition 2 in Song and Xue [39].