Product-line Pricing under Discrete Mixed Multinomial Logit Demand

Hongmin Li • Scott Webster

W.P. Carey School of Business, Arizona State University
Tempe, Arizona 85287, USA
hongmin.li@asu.edu, scott.webster@asu.edu

Nicholas Mason • Karl Kempf

Decision Engineering, Intel Corporation
Chandler, AZ 85226, USA
nicholas.mason@intel.com, karl.g.kempf@intel.com

Abstract

We study a product-line price optimization problem with demand given by a discrete mixed multinomial logit (MMNL) model. The market is divided into a finite number of market segments, with product demand in each segment governed by the multinomial logit (MNL) model. We show that the concavity property with respect to the choice probability vector shown for the MNL model breaks down under MMNL even for entirely symmetric price sensitivities. In addition, the equal markup property identified for the MNL model no longer holds under MMNL, suggesting that heterogeneity in customer population justifies non-equal-markup pricing. In this paper, we characterize the profit function under MMNL as the sum of a set of quasiconcave functions and present efficient optimization algorithms. We demonstrate the application of our methods using data from Intel Corporation. Our results show that the optimal prices exploit segment differences through redistribution of sales and profit among customer segments. In addition, we use our methods to generate the profit-share efficient frontier, helping the firm to set prices that effectively balance these two measures. The tools developed in this research function in three important ways at Intel: (1) They provide a new alternative for market share prediction among Intel products for different customer segments, adding to the suite of independent demand forecasting tools; (2) they optimize product prices based on segment-specific customer preferences revealed through sales data, a capability that Intel's current pricing tools include only heuristically; (3) they quantify the tradeoff between profit and market share, adding a new decision support capability to the company.

Keywords: *Pricing, Revenue Management, Mixed Multinomial Logit, Latent Class Model*

1 Introduction

Increasingly diversified market preferences have driven firms to offer multiple substitutable products that differ in various dimensions such as features and prices. The resulting product proliferation increases the complexity of many business decisions, one of which is pricing. In practice, two major hurdles exist in pricing: (1) the frequent updating of the product line as a firm introduces new products and retires old products (i.e., product prices need to be adjusted each time such an event occurs); (2) the heterogeneity in the customer population (i.e., different types of customers use the products differently and consequently value them differently). As such, a decision tool is in order to systematically optimize prices, accounting for past sales and price information and adjusting to changes in the product line as well as heterogeneity in the customer population.

We use an example in the computer microprocessor industry to illustrate the challenges of pricing due to heterogeneities in customer preferences. Consider the three microprocessor stock-keeping-units (SKUs) given in Table 1. Each SKU is defined by the unique combination of feature designs including the number of cores (number of processors run in parallel), frequency (speed of each core), TDP (an index of how much electric power the processor consumes), as well as price. Consider three different customers. Customer 1 needs a microprocessor used in a data center performing web search; customer 2 needs a microprocessor for a server performing scientific simulation studies; customer 3 needs a microprocessor for a simple database server at a small enterprise.

Table 1: Microprocessor SKUs.

SKU	Cores	Frequency	TDP	Price
1	8	2.9	135	$2100
2	4	3.2	95	$1400
3	4	2.2	60	$550

Since web search is a process that can be distributed to various cores, having more cores allows many jobs to be processed simultaneously, increasing the total number of jobs finished. This makes a high number of cores more valuable to customer 1 than a high frequency. Although a high power consumption index is unfavorable, in general, it can be compensated through the necessary cooling infrastructure in the case of larger data centers. Therefore, for customer 1, SKU 1 is more likely to be most favored. In the case of customer 2, it may be that the computing workload of the simulation studies is not easy to parallelize or distribute to multiple cores. Therefore, additional cores may not be as useful as a high frequency. In this case, SKU 2 will be most valuable. For customer 3, a low-end SKU may suffice in computation. Furthermore, a low TDP will make maintenance and cooling costs low. As a result, SKU 3 may be best for customer 3. As we observe in this example, different types of customers place different emphasis on features and price. This affects their SKU choices, and consequently, should influence the firm's pricing decision.

The multinomial logit (MNL) choice model is widely used for modeling demand of multiple differentiated products. It is based on random utility maximization (customers' utility for a product follows a given random distribution and each customer chooses the product that yields the highest utility) and has been empirically proven to perform well in various industries such as transportation (McFadden, 1974), telephone services (Train et al., 1987) and coffee purchases (Guadagni and Little, 1983). A key advantage of the MNL model is attributed to its flexibility in incorporating customer characteristics in addition to attributes of the choice alternatives. In other words, the choice prediction in the MNL model may depend on both the alternatives and the customer type. For example, the utility of product i for a customer of type k can be modeled as $u_{ik} = a_{ik} - b_{ik}p_i + \epsilon_i$ where p_i is the price of product i, b_{ik} signifies type k customer's price sensitivity toward product i, a_{ik} represents price-independent attractiveness, and ϵ_i signifies noise. While pricing of differentiated products under the MNL and its derived models has attracted much attention from researchers, the state of art theoretical development has focused mainly on incorporating heterogeneity across the choice alternatives and very little on heterogeneity in the customer population. That is, extant pricing models under the MNL demand focus on the case for which $a_{ik} = a_i$ and $b_{ik} = b_i$ for all k and neglect heterogeneity across different types of customers, which fails to take advantage of the the MNL model's capability for making customer-specific choice predictions.

In this paper, we take an initial step in filling this void and address the pricing problem under a form of logit choice model that incorporates customer segment information and apply it to a practical setting of pricing microprocessors at Intel Corporation. In particular, we consider a discrete mixed multinomial logit (MMNL) demand model, which aligns with the setting of a market that can be decomposed into a finite number of market segments, each with its own set of product utility parameters reflecting the unique emphasis that this segment of customers place on the features and price (see Kamakura and Russell 1989 and

Train 2003). We refer to this model as the MMNL model and we refer to the MNL model without customer-specific consideration as the basic MNL model or simply the MNL model in the remainder of the paper.

We show that incorporating segment-specific preference parameters breaks the concavity of the profit function with respect to the choice probability vector identified in Li and Huh (2011), as well as the analysis in Gallego and Wang (2014). We characterize the total profit as the sum of a set of quasiconcave functions with respect to a vector of market shares for a particular segment and show that it is in general not a concave or quasiconcave function. We present an example in which the profit changes from a concave to a non-concave (and non-quasiconcave) function as the parameter values shift.

A salient result of profit optimization under the basic MNL model is that the optimal markups for all products are equal (Dong et al., 2009), which is controversial as "equal markup" for differentiated products is not commonly observed in practice (Alptekinoğlu and Semple, 2016). In contrast to the basic MNL model, we show that the equal markup property does not hold under the MMNL model even with symmetric price sensitivities across products and segments, which helps reconcile the divergence of theoretical prediction and observed practice.

In the case that the model parameters are such that the total profit is quasiconcave (which typically occurs when the segment differences are sufficiently small), we propose an efficient algorithm for finding the optimal prices under the MMNL model to account for the impact of customer heterogeneity on prices. When the total profit is not quasiconcave, we propose a gradient-descent approach to search for stationary point solutions and by randomizing the starting price vector we find the stationary point solution that is most likely to be the global optimum. In practice, management may be concerned about both profit and market share. We generalize our approach to generate the efficient frontier of optimal profit by total market share, thereby helping management to strike a balance between these two measures of interest. We apply this model to Intel products, and illustrate how the model can be used in practice to improve decision making and how the optimal pricing strategies derived through our analysis compare with the current practice. Our results show that the optimal prices exploit segment differences through redistribution of sales and profit among customer segments. In addition, we derive the profit-market share efficient frontier and locate the current practice relative to this frontier. This provides insights for decision markers on how to balance profit and market share considerations to best achieve Intel's objectives.

Our contributions are both theoretical and practical. The MMNL model, as shown by McFadden and Train (2000), can approximate any discrete choice model consistent with random utility maximization (RUM) to any degree of precision. Thus, from a theoretical perspective, results regarding MMNL pricing problems reflect the character of a general discrete choice model; the solution approach we propose for solving optimal prices under MMNL is further generalizable to other discrete choice models consistent with RUM. From a practical perspective, we present a systematic approach for modeling demand and managing the pricing decision of a dynamically evolving product line that takes into consideration sales history and differences in the various constituents of the customer population. The decision tools derived from this research serve multiple objectives at Intel: (1) They provide a new alternative for market share prediction among Intel products for different customer segments, adding to the company's suite of independent demand forecasting tools; (2) they optimize product prices based on segment-specific customer preferences revealed through sales data, a capability that Intel's current pricing tools include only heuristically; (3) they quantify the tradeoff between profit and market share, adding a new decision support capability to the company. Given the wide applicability of the logit family models, the analysis and the solution approach extend to other industries with similar problems.

The rest of the paper is organized as follows. We review prior related work in Section 2. In Section 3, we characterize the profit function under the discrete MMNL model and derive the solution approach. Then, we apply the solution approach to Intel's microprocessor products in Section 4. We conclude in Section 5.

2 Prior Research

The use of MNL and its derived models has been extensive in empirical and analytical research. MMNL can be viewed as an MNL model with coefficients varying by customer. In addition to allowing taste variation among customers and correlation among unobserved utilities, MMNL overcomes the *Independence-from-Irrelevant-Alternatives* (i.e., the ratio of choice probabilities of any two alternatives is independent of other alternatives) restriction of the MNL model and admits more general substitution patterns among choice alternatives (Train 2003). Its application first appeared in Boyd and Mellman (1980) and Cardell and Dunbar (1980) for predicting consumer choice in the automobile industry. The model has since been widely studied and applied (see McFadden and Train 2000, Train 2003, and references therein). A powerful result established by McFadden and Train (2000) states that any random utility based discrete choice model can be approximated by an MMNL with arbitrary precision, i.e., MMNL can accurately represent any random utility maximization discrete choice model. Kamakura and Russell (1989) propose a discrete MMNL model for preference segmentation and use household-level data to estimate intrinsic brand utilities and price sensitivities. This is the same model we investigate in this paper, which is also referred to as the "latent class model" in the literature (see Greene and Hensher 2003). As explained by Train (2003), *"this specification is useful if there are M segments in the population, each of which has its own choice behavior or preferences."* Such interpretation allows us to apply the discrete MMNL to many practical choice settings in which customers can be segmented as in the case of Intel. More broadly, given the connection between MMNL and any general discrete choice model established by McFadden and Train (2000), and that a continuous MMNL can always be approximated by a discrete MMNL, theoretical results and solution approaches derived in our paper suggest a path to solving the pricing problem under any discrete choice model that is consistent with RUM.

A number of researchers have studied the product assortment problem under MMNL demand (e.g., Blanchet et al. 2013, Feldman and Topaloglu 2015, Rusmevichientong et al. 2014), however, the literature on pricing under MMNL is sparse. To our knowledge, Hanson and Martin (1996) is the only paper that considers product-line pricing under MMNL demand. They provide a path-following algorithm to find a global maximum if, among other conditions, there exists a unique global maximum over a set of rescaled utility functions (i.e., utility reduced by factor $t \in [0, 1]$). However, for a given problem, the authors note that it is not possible to test whether there is a unique global maximum and have not explored any structural property of the profit function. Aydin and Porteus (2008) consider profit optimization under the MNL model and give an example of a special case of their demand model that they describe as "demand with multiple customer segments" (i.e., demand model D3). In spite of this label, this model is fundamentally different from MMNL. In particular, their measure of product attraction is a weighted sum of product attraction in each segment. As a consequence, the share of a product is not the weighted sum of product shares in each segment (as it is in MMNL), but is the attraction of the product divided by the sum of attraction of all products. Aksoy-Pierson et al. (2013) study price competition under the

MMNL model in which each firm owns a single product and show that a unique pure Nash equilibrium exists if the market concentration is not high, i.e., within each market segment, each firm only captures a percentage share that is below a certain level.

Most relevant pricing models for the family of logit choices have been for MNL and nested logit (NL) demand and this stream of literature is well developed. The MNL model, essentially a special case of the MMNL, exhibits nice properties in the context of price optimization. Researchers have shown that a unique price solution exists and can be found by the first-order condition (Aydin and Porteus 2008, Hopp and Xu 2005, Maddah and Bish 2007, Akçay et al. 2010). Although the profit is not concave in prices (Hanson and Martin 1996), Dong et al. (2009) and Song and Xue (2007) show that it is concave in the choice probability vector. An interesting result arising from the MNL pricing model is the equal-markup property: if price sensitivities are symmetric, the optimal markups are the same across products (see, Dong et al. 2009).

The NL model generalizes the MNL model by allowing choices to be grouped into different nests (McFadden 1978, McFadden 1980). Pricing problems under the NL model have drawn attention in recent years (e.g., Kök and Xu 2011, Li and Huh 2011, Wang 2012, Gallego and Wang 2014, Rayfield et al. 2015, Li et al. 2015, Kouvelis et al. 2015). Li and Huh (2011) prove concavity of the profit with respect to the choice probability vector in a NL model. Gallego and Wang (2014) consider both assortment and pricing under NL demand allowing product-specific price sensitivities. They show that a unique solution can be ensured with a restrictive condition on price sensitivities and nest-specific similarity parameters. Huh and Li (2015) extend to a similar condition in a multi-level NL model. Such uniqueness conditions share a common theme – the degree of asymmetry in the price-sensitivity parameters needs to be limited.

In contrast to the MNL and NL models, we show in this paper that the profit function under the MMNL model is far less well-behaved. The concavity property with respect to the choice probability vector shown in Dong et al. (2009), Song and Xue (2007), and Li and Huh (2011) breaks down under MMNL even for entirely symmetric price sensitivities across *all* products and *all* segments. With the MMNL model, the equal markup property identified for the MNL model no longer holds, which suggests that heterogeneity in customer population may help justify non-equal-markup pricing. In addition, the analysis that leads to a unique solution under the MNL or NL model shown in the extant literature does not carry through to the MMNL model. In this paper, we characterize the profit function under the MMNL model as the sum of a set of quasiconcave functions and demonstrate with examples how the profit function might shift from concave to nonconcave functions as the model primitives change. Moreover, we present efficient optimization algorithms for the pricing problem and demonstrate the application of our methods using data from Intel Corporation.

With the possible exception of price optimization methods applied to perishable products–a classical revenue management setting—there is a gap between the academic literature on pricing and pricing in practice (Phillips 2005, p. 5), e.g., cost-plus pricing is common in practice (Eek and Berggren 2007, Noble and Gruca 1999). For business-to-business durable goods, the setting we consider, the literature on price optimization methods in practice is sparse but emerging. Silva-Risso and Ionova (2008) and Silva-Risso et al. (2008) describe methods and a decision support system to guide the design of price discounts at Chrysler using a two-level NL model. As in our application, their system generates an efficient frontier. However, they generate the efficient frontier of a product's net margin by share (i.e., for given reduction in list price, the mix of discounting through rebate and/or subsidized financing to maximize share) whereas we generate the efficient frontier of total profit by share. Mookherjee et al. (2016) describe a price optimization decision support system at Ingram Micro, a large distributor of technology products. Similar to our approach, the system uses an MNL model to predict how a segment of consumers respond to price. However, in Mookherjee et al. (2016), as well as in Silva-Risso and Ionova (2008) and Silva-Risso et al. (2008), product price can vary by customer segment or individual customer. In contrast, we require the same product prices across customer segments (e.g., modeled via MMNL demand), which is a harder problem.

3 Analysis of the Price Optimization Problem

The MMNL model is derived from MNL choice models with utility parameters drawn from a mixing distribution. While the MMNL model allows for a continuous mixing distribution, we limit our discussion to discrete distributions as it aligns with the setting of a market that can be decomposed into a finite number of market segments, each with its own set of product utility parameters. This discrete MMNL model is also referred to as the "latent class model" (Greene and Hensher, 2003).

While the theoretical importance of the MMNL model is clearly stated by McFadden and Train (2000) (in that it can approximate any RUM choice model with arbitrary precision), the practical importance needs emphasis. The basic MNL model embeds all market heterogeneity in the random Gumbel term, which essentially means that the known information of all customers is the same. MMNL, in contrast, explicitly models the known differences among customers, which can be more realistic and useful. In particular, consider customers making a selection of one of n product choices and a no-purchase alternative. The market is comprised of m customer segments with utility $u_{ik} = a_{ik} - b_{ik}p_i + \epsilon_i$ for product i and segment k. The product purchase probabilities within each segment are given by the MNL model:

$$q_{ik} = \frac{e^{a_{ik}-b_{ik}p_i}}{1+\sum\limits_{j=1}^{n} e^{a_{jk}-b_{jk}p_j}} = q_{0k}A_{ik}e^{-b_{ik}p_i} \tag{1}$$

where q_{ik} is the probability that a customer in segment k chooses product i, p_i is the price of product i, a_{ik} is the price-independent preference value for product i in segment k (referred to as "preference value" hereafter), $A_{ik} = e^{a_{ik}}$ is the price-independent "attraction" of product i in segment k, and the no-purchase probability among segment k customers is

$$q_{0k} = \frac{1}{1+\sum\limits_{j=1}^{n} A_{jk}e^{-b_{jk}p_j}}. \tag{2}$$

The probability that a randomly selected customer belongs to segment k is w_k with $\sum_{k=1}^{m} w_k = 1$, and thus the purchase probability of product i and the no-purchase probability are

$$\begin{aligned} q_i &= \sum_{k=1}^{m} w_k q_{ik} = \sum_{k=1}^{m} w_k \frac{e^{a_{ik}-b_{ik}p_i}}{1+\sum\limits_{j=1}^{n} e^{a_{jk}-b_{jk}p_j}} \quad \text{and} \\ q_0 &= \sum_{k=1}^{m} w_k q_{0k} = \sum_{k=1}^{m} w_k \frac{1}{1+\sum\limits_{j=1}^{n} e^{a_{jk}-b_{jk}p_j}} = 1 - \sum_{i=1}^{n} q_i. \end{aligned}$$

Let the marginal cost of product i be c_i. The profit as a function of price vector $\mathbf{p} = (p_1, ..., p_n)$ is

$$\pi(\mathbf{p}) = \sum_{i=1}^{n} (p_i - c_i)\, q_i = \sum_{i=1}^{n} (p_i - c_i) \left(\sum_{k=1}^{m} w_k q_{ik} \right) = \sum_{k=1}^{m} w_k \sum_{i=1}^{n} (p_i - c_i)\, q_{ik} = \sum_{k=1}^{m} w_k r_k(\mathbf{p})$$

where $r_k(\mathbf{p}) = \sum_{j=1}^{n} (p_j - c_j)\, q_{jk}$ is the profit contribution from a segment k customer. Taking derivatives of the total profit with respect to prices yields

$$\frac{\partial \pi}{\partial p_i} = q_i + \sum_{j=1}^{n}(p_j - c_j)\sum_{k=1}^{m} w_k b_{ik} q_{ik} q_{jk} - (p_i - c_i)\sum_{k=1}^{m} w_k b_{ik} q_{ik} \ ,$$

which leads to the following first order necessary condition for optimality:

$$p_i - c_i = \frac{1}{\sum_k \frac{w_k q_{ik}}{q_i} b_{ik}} + \frac{\sum_k \frac{w_k q_{ik}}{q_i} b_{ik} r_k}{\sum_k \frac{w_k q_{ik}}{q_i} b_{ik}} \ . \tag{3}$$

This condition reveals a property of the optimal markup that contrasts with the basic MNL model. The following lemma shows that, even with symmetric price sensitivities across *all* products and *all* segments, the optimal mark-up is in general not equal across products. In particular, customer heterogeneity in preference value (i.e., difference in a_{ik} values across segments), justifies differentiated markups across products (see Lemma 2 and its proof in the online appendix for details). Recall that the basic MNL model prescribes equal-markup pricing for symmetric price sensitivities regardless of preference value differences among products. This changes under the MMNL model because different segments of customers value the same product differently. Interestingly, the order of the optimal markup for different products does not necessarily follow the same sequence as the product preference value. That is, even if $a_{ik} > a_{jk}$ for all k, it is not necessarily true that the optimal markup of product i is greater than that of product j. Rather, the sequence of the optimal markup depends on how each product's preference value differs across segments, as illustrated in the following two-segment case.

Lemma 1. *Let there be two segments, i.e., $k \in \{A, B\}$ and assume $b_{ik} = b$ for all i, k. Let $p_i^*, i = 1, \ldots, n$ satisfy* (3)*. Then $p_i^* - c_i \geq p_j^* - c_j$ if and only if*

$$\left[(a_{iA} - a_{iB}) - (a_{jA} - a_{jB})\right](r_A - r_B) \geq 0 \quad \textit{for } i \neq j.$$

The above finding is not intuitive at first glance and we elaborate with a hypothetical "steak versus tofu" scenario. Steak and tofu are both protein-rich menu options and are often considered substitutes or competing items. Let there be two customer segments A and B. The restaurant cannot charge different prices for the same product to different segments of customers, as in our problem setting. If preference values do not differ across segments (i.e., $a_{ik} = a_i$ for all i and k), then the segments become degenerate and the MMNL model reduces to MNL; in this case condition (3) reduces to $p_i - c_i = \frac{1}{b} + \pi(\mathbf{p})$, i.e., steak and tofu have the same markup. However, suppose that customers in segment A have higher preference values than customers in segment B (i.e., $a_{iA} > a_{iB}$), and thus for any price vector, we have $r_A > r_B$. Furthermore, suppose that customers have a higher preference value for steak ($i = 1$) than tofu ($i = 2$) (i.e., $a_{1k} > a_{2k}$), and that tofu is only attractive to segment A that is more conscious about cholesterol intake. In this example, the difference in tofu preference values across the two segments is larger than steak (i.e., $a_{2A} - a_{2B} > a_{1A} - a_{1B}$). The large difference in preference values for tofu allows the restaurant to increase profit by setting a higher markup for tofu, focusing on the high-valuation segment A of cholesterol-conscious customers and effectively pricing the low-valuation segment B out of the market. This is not the case for steak where the difference in segment preference values is smaller; the restaurant maximizes profit by setting the price of steak to appeal to both high- and low-valuation segments resulting in a lower markup compared to tofu. Therefore, the pricing strategy for tofu is of a niche product strategy whereas that for steak is a high-volume product strategy. In summary, differentiated markups in the MMNL model is due to segment differentiation. In the online appendix, we provide a two-product-two-segment numerical example for further illustration.

For more than two segments and/or asymmetric b_{ik} values, the condition of markup sequence comparison becomes intractable but it suffices to say that in general the sequence of the optimal markup does not necessarily follow the sequence of preference value.

For asymmetric price sensitivities, Gallego and Wang (2014) define $p_i - c_i - \frac{1}{b_i}$ as the "adjusted markup" and build an analysis upon the fact that, at optimality, the adjusted markup is the same across products under the NL model for which the basic MNL model is a special case. However, as we have observed in equation (3), this adjusted markup becomes product dependent under MMNL. As a result, the analysis used in Gallego and Wang (2014) does not carry through to the MMNL model.

The profit function $\pi(\mathbf{p})$ is not quasiconcave in $\mathbf{p}$, even for the special case of the basic MNL (Hanson and Martin 1996). However, for the basic MNL model, the profit as a function of the quantity vector $\mathbf{q}$ is concave. This is shown in Dong et al. (2009) and Song and Xue (2007) for symmetric price sensitivities and in Li and Huh (2011) for more general price sensitivities. Unfortunately, as illustrated by the example in Figure 1, such concavity property breaks down under MMNL. In this example, there is a single product and two

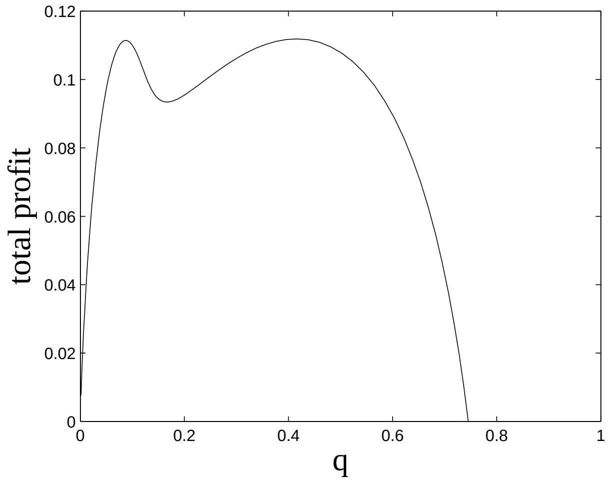


Figure 1: Profit by Market Share (not Concave or Quasiconcave in Market Share)

customer segments with parameter values given by $b_{11} = 1, b_{12} = 10, A_{11} = 1, A_{12} = 10, w_1 = 0.4, w_2 = 0.6$. The horizontal axis in the figure is the market share of the product. We note that, the profit is not even quasiconcave in market share.

The discussion above indicates that the profit function under the MMNL model is not as well behaved as the basic MNL or NL models. Since the analytical approaches used for other logit models do not apply, we explore a new approach to characterize the profit function under MMNL, while taking advantage of the profit concavity with respect to market share of the basic MNL model.

3.1 Characterizing the Profit Function

Let $\mathbf{q} = (\mathbf{q}^1, ..., \mathbf{q}^m)$ where $\mathbf{q}^k = (q_{1k}, ..., q_{nk})$ is the purchase probability vector of segment k. Inverting (1), we have

$$p_i = g_{ik}(\mathbf{q}^k) = \log \left(\frac{A_{ik}\left(1 - \sum_{j=1}^{n} q_{jk}\right)}{q_{ik}} \right)^{1/b_{ik}} \quad \text{for any } k \tag{4}$$

and total profit as a function of

$$\mathbf{q} \in \Omega = \left\{ q_{ik} \middle| \sum_{i=1}^{n} q_{ik} \leq 1,\ q_{ik} \geq 0,\ g_{ik}(\mathbf{q}^k) \geq c_i\ ,\ g_{i1}\left(\mathbf{q}^1\right) = \ldots = g_{im}\left(\mathbf{q}^m\right)\ \forall i, k \right\}$$

is

$$\Pi(\mathbf{q}) = \sum_{i=1}^{n} \left(g_{ik}\left(\mathbf{q}^k\right) - c_i\right) \sum_{l=1}^{m} w_l q_{il} = \sum_{k=1}^{m} w_k \sum_{i=1}^{n} \left(g_{ik}\left(\mathbf{q}^k\right) - c_i\right) q_{ik} = \sum_{k=1}^{m} w_k R_k\left(\mathbf{q}^k\right)$$

where $R_k(\mathbf{q}^k) = \sum_{i=1}^{n} \left(g_{ik}\left(\mathbf{q}^k\right) - c_i\right) q_{ik}$ is the profit contribution from a segment k customer. Note that, for any segment k, $R_k(\mathbf{q}^k)$ is concave in $\mathbf{q}^k$ (i.e., the profit function with basic MNL demand is concave in the quantity vector, as noted above). We remark that the condition $g_{ik}(\mathbf{q}^k) \geq c_i$ in the definition of the set Ω is equivalent to $\left(e^{b_{ik}c_i}/A_{ik}\right) q_{ik} + \sum_{j=1}^{n} q_{jk} \leq 1$, which is linear and excludes prices that lead to negative markup of a product as these cannot be optimal. To see this, assume that product i is priced below cost c_i. Then by raising p_i to c_i while keeping other prices unchanged, the total profit strictly improves (product i's profit increases from negative to zero and profit of all other products improves due to increased quantity). Because $\Pi(\mathbf{q})$ is a weighted sum of concave functions, $\Pi(\mathbf{q})$ is concave in $\mathbf{q}$ (Boyd and Vandenberghe 2004, page 79), suggesting that $\Pi(\mathbf{q})$ may exhibit attractive properties for optimization. However, to assure feasible $\mathbf{q}$, we require

$$g_{i1}\left(\mathbf{q}^1\right) = \ldots = g_{im}\left(\mathbf{q}^m\right) \text{ for all } i \tag{5}$$

(i.e., the price of product i is constant across segments). From (4) and (5), it follows that for any k,

$$\left(\frac{A_{ik}q_{0k}}{q_{ik}}\right)^{1/b_{ik}} = \left(\frac{A_{i1}q_{01}}{q_{i1}}\right)^{1/b_{i1}}$$

and thus

$$q_{ik} = A_{ik}q_{0k}\left(\frac{q_{i1}}{A_{i1}q_{01}}\right)^{b_{ik}/b_{i1}} .$$

Hence,

$$1 - q_{0k} = \sum_{j=1}^{n} q_{jk} = q_{0k} \sum_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}q_{01}}\right)^{b_{jk}/b_{j1}} .$$

This yields the relationship

$$1 + \sum_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}q_{01}}\right)^{b_{jk}/b_{j1}} = \frac{1}{q_{0k}} .$$

Therefore, we can express q_{ik} as a function of the vector $\mathbf{q}^1 = (q_{11}, q_{21}, \ldots, q_{n1})$ for all i, k:

$$q_{ik} = \frac{A_{ik}\left(\frac{q_{i1}}{A_{i1}q_{01}}\right)^{b_{ik}/b_{i1}}}{1+\sum\limits_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}q_{01}}\right)^{b_{jk}/b_{j1}}} = \frac{A_{ik}\left(\frac{q_{i1}}{A_{i1}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{ik}/b_{i1}}}{1+\sum\limits_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{jk}/b_{j1}}}. \tag{6}$$

Define

$$f_k(\mathbf{q}^1) := \left(\frac{A_{1k}\left(\frac{q_{11}}{A_{11}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{1k}/b_{11}}}{1+\sum\limits_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{jk}/b_{j1}}}, \ldots, \frac{A_{nk}\left(\frac{q_{n1}}{A_{n1}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{nk}/b_{n1}}}{1+\sum\limits_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}\left(1-\sum\limits_{l=1}^{n} q_{l1}\right)}\right)^{b_{jk}/b_{j1}}} \right),$$

$$f\left(\mathbf{q}^1\right) := \left(f_1\left(\mathbf{q}^1\right), \ldots, f_m\left(\mathbf{q}^1\right)\right) ,$$

$$\hat{R}_k\left(\mathbf{q}^1\right) := R_k\left(f_k\left(\mathbf{q}^1\right)\right) , \text{ and} \tag{7}$$

$$\hat{\Pi}\left(\mathbf{q}^1\right) := \sum_{k=1}^{m} w_k \hat{R}_k\left(\mathbf{q}^1\right) = \sum_{k=1}^{m} w_k R_k\left(f_k\left(\mathbf{q}^1\right)\right) = \Pi\left(f_1\left(\mathbf{q}^1\right), \ldots, f_m\left(\mathbf{q}^1\right)\right) = \Pi\left(f\left(\mathbf{q}^1\right)\right). \tag{8}$$

Thus, the $n \times m$-dimensional profit maximization problem

$$\max_{\mathbf{q}\in\Omega} \Pi\left(\mathbf{q}\right)$$

is equivalent to the following n-dimensional optimization problem:

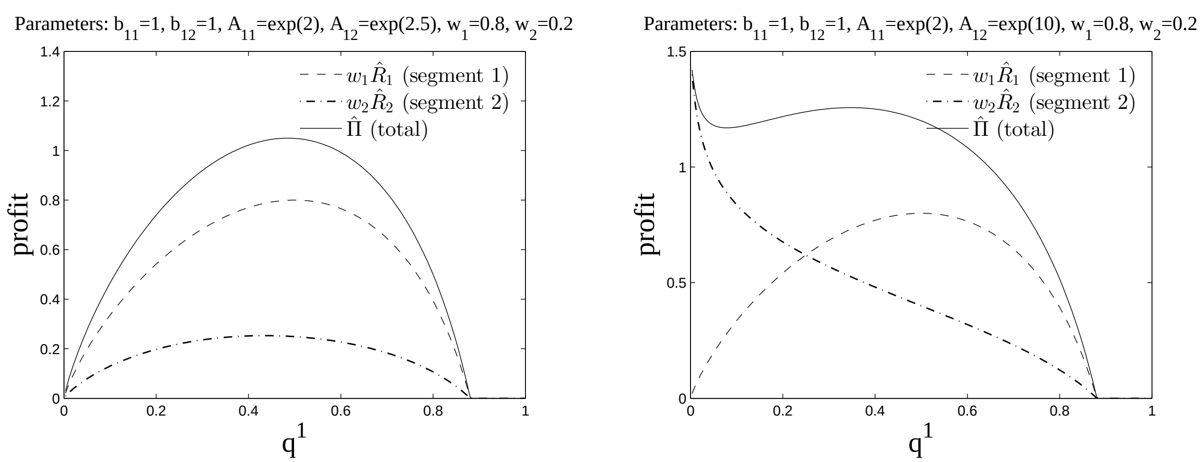
$$\max_{\mathbf{q}^1\in\Omega_1} \hat{\Pi}\left(\mathbf{q}^1\right)$$

where $\Omega_1 = \{\mathbf{q}^1 | \sum_{i=1}^{n} q_{i1} \leq 1,\ q_{i1} \geq 0,\ g_{i1}(\mathbf{q}^1) \geq c_i\ \forall i\}$. In other words, we have transformed the total profit to a function of the segment 1 quantity vector $\mathbf{q}^1$.

To address the question of whether the profit function $\hat{\Pi}\left(\mathbf{q}^1\right)$ defined over convex set Ω_1 is well-behaved (e.g., quasiconcave), we begin by considering the simple case of a single product with symmetric price sensitivities across segments. For this case, each segment is distinguished by a unique value of the price-independent attraction parameter A_{1k}. We show that the introduction of multiple segments in this simple setting (i.e., via the introduction of distinct price-independent attraction parameters) causes the concave structure of the MNL profit function to break down. The following proposition shows that $\hat{\Pi}(\mathbf{q}^1)$ is concave if the variation in A_{1k} values is within a certain range. After the proposition, we provide an example that illustrates how the profit function shifts from concave to nonconcave as variation in A_{1k} values increases.

Proposition 1. *For a single-product MMNL model, if* $\frac{\max_k A_{1k}}{\min_k A_{1k}} \leq 2$ *and* $b_{1k} = b$ *for all* k*, then* $\hat{\Pi}(\mathbf{q}^1)$ *is concave on* Ω_1*.*

Figure 2 illustrates functions $\hat{R}_1(\mathbf{q}^1)$, $\hat{R}_2(\mathbf{q}^1)$, and $\hat{\Pi}(\mathbf{q}^1)$ for a pair of problem instances with one product and two segments. Parameter values are identical except for the value of A_{12}; $A_{12}/A_{11} \approx 1.6$ in Figure 2(a) and $A_{12}/A_{11} \approx 2981$ in Figure 2(b). Figure 2(b) shows that even with symmetric price sensitivities, when the variation in A_{1k} values is sufficiently large, the profit function is not quasiconcave and uniqueness of the optimal solution is not guaranteed. Recall that the uniqueness conditions of the optimal prices under the NL



(a) Concave Profit

(b) Non-concave (and non-quasiconcave) Profit

Figure 2: Single-product-two-segment Profit for Cases of Concave and Non-concave Profit

model (Li and Huh 2011, Gallego and Wang 2014, Huh and Li 2015) essentially constrain the degree of asymmetry in the price-sensitivity parameters. Here we note that symmetry of price sensitivity alone does not ensure a unique solution under MMNL. Proposition 1 and the examples in Figure 2 hint that the condition for a unique price solution requires that the difference between segment-specific attractiveness be limited. The next proposition formalizes this idea for the multi-product case.

Proposition 2. *Assume $b_{ik} = b_i$ for all k. There exists $\overline{\boldsymbol{\lambda}} = (\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_n) > \mathbf{1}$ such that, if $\frac{\max_k A_{ik}}{\min_k A_{ik}} < \overline{\lambda}_i$ for all i, then $\hat{\Pi}(\mathbf{q}^1)$ is concave on Ω_1 and the optimal price vector is unique.*

Proposition 2 implies that there is a neighborhood around $\boldsymbol{\lambda} = \mathbf{1}$ where the profit function is concave in $\mathbf{q}^1$. Outside this neighborhood, i.e., when the segments are sufficiently asymmetric, the profit function may not be concave. As we observe in Proposition 2 and its proof, in general, accurate identification of this neighborhood is very difficult when $n > 1$. This complexity inhibits closed-form characterization of $\overline{\boldsymbol{\lambda}}$. Even for a specific problem instance (defined by a set of parameter values), the numerical evaluation of the signs of the diagonal and leading principle minors of the Hessian to test whether it is negative definite over Ω_1 is not promising (i.e., only a finite subset of points in Ω_1 can be evaluated).

Define $\tilde{\Pi}(\mathbf{q}^m)$ as the total profit as a function of $\mathbf{q}^m$, similar to the definition of $\hat{\Pi}(\mathbf{q}^1)$. While the total profit is in general not concave in either $\mathbf{q}^1$ or $\mathbf{q}^m$, the following proposition illustrates that in the case when $b_{ik} = b_i$ (i.e., when the known segment differences are mainly due to taste variations for product features and performance), the profit is bounded from above and below by two concave functions. Let

$$\begin{aligned}
\overline{\Pi}(\mathbf{q}^1) &= \sum_i \frac{q_{i1}}{b_i}\left[\log A_{i1} - \log q_{i1} + \log\left(1 - \sum_j q_{j1}\right) - b_i c_i\right] \sum_k w_k A_{ik}/A_{i1} \\
\underline{\Pi}(\mathbf{q}^m) &= \sum_i \frac{q_{im}}{b_i}\left[\log A_{im} - \log q_{im} + \log\left(1 - \sum_j q_{jm}\right) - b_i c_i\right] \sum_k w_k A_{ik}/A_{im} \,.
\end{aligned}$$

Proposition 3. *Assume $b_{ik} = b_i$ for all k. In addition, assume $A_{i1} \leq A_{ik} \leq A_{im}$ for all k. (i) $\overline{\Pi}(\mathbf{q}^1)$ is concave in $\mathbf{q}^1$ and $\underline{\Pi}(\mathbf{q}^m)$ is concave in $\mathbf{q}^m$. (ii) $\hat{\Pi}(\mathbf{q}^1) \leq \overline{\Pi}(\mathbf{q}^1)$ and $\tilde{\Pi}(\mathbf{q}^m) \geq \underline{\Pi}(\mathbf{q}^m)$.*

Therefore, in this case, the total profit is bounded by functions that are easy to optimize, yielding lower and upper bounds on the optimal total profit. We provide these bounds in Corollary 1 in the online appendix.

To further characterize the profit function, we note that both segment-profit functions $(\hat{R}_1(\mathbf{q}^1), \hat{R}_2(\mathbf{q}^1))$ in Figure 2(b) are quasiconcave, even though $\hat{\Pi}(\mathbf{q}^1)$ is not quasiconcave. In the following proposition, we show that this feature of the segment-profit functions holds in general.

Proposition 4. *For the MMNL model, $\hat{R}_k(\mathbf{q}^1)$ is quasiconcave on Ω_1 for all k, and $\hat{R}_1(\mathbf{q}^1)$ is concave on Ω_1.*

To prove Proposition 4, we decompose the function $f_k(\mathbf{q}^1)$ into more elementary functions, and show that each of these functions preserves convexity. This implies that the set $\{f_k(\mathbf{q}^1)\,|\mathbf{q}^1 \in \Omega_1\}$ is a convex set. Using the fact that the inverse image of a convex set under linear-fractional transformation is convex, and by showing that a certain power transformation also preserves convexity of a set, we prove that the superlevel set of $\hat{R}_k$ is convex, establishing quasiconcavity of $\hat{R}_k$.

Proposition 4 tells us that the MMNL profit function $\hat{\Pi}(\mathbf{q}^1)$ is a weighted sum of quasiconcave segment-profit functions, at least one of which is assured to be concave (i.e., $\hat{R}_1(\mathbf{q}^1)$). While the weighted sum of concave functions is concave (with a unique stationary point), there is no such assurance that the weighted sum of quasiconcave functions is quasiconcave (e.g., Figure 2(b)). Nevertheless, the relatively well-behaved structure of the segment-profit functions in the total profit function $\hat{\Pi}(\mathbf{q}^1) = \sum_{k=1}^{m} w_k \hat{R}_k(\mathbf{q}^1)$ hints that $\hat{\Pi}(\mathbf{q}^1)$ may exhibit a unique stationary point over a range of MMNL parameter values and may help explain the favorable computational results that we observe in Section 5.

3.2 Optimization Algorithms

Proposition 4 characterizes the total profit under the MMNL model as a weighted sum of quasiconcave segment profit functions. Propositions 1 and 2 indicate that when the degree of segment asymmetry is sufficient small, the total profit is concave in $\mathbf{q}^1$ and a unique optimal solution is guaranteed. When the degree of asymmetry becomes sufficiently large, the total profit becomes nonconcave or even nonquasiconcave, as demonstrated in Figure 2. In this section, we propose different algorithms addressing these two scenarios. If the profit function is known to be quasiconcave (e.g., with high degree of symmetry across segments), then the following bisection search algorithm is assured to return an optimal solution.

Algorithm 1 (bisection search)**.** *Step 1. Start with an initial interval* $\left[\hat{\Pi}_L, \hat{\Pi}_H\right]$ *in which the optimum must lie.*

Step 2. Let $t = \left(\hat{\Pi}_L + \hat{\Pi}_H\right)/2$ *and solve the feasibility problem: find* $\mathbf{q}^1 \in \Omega_1$ *s.t.* $\hat{\Pi}(\mathbf{q}^1) \geq t$*. Note that the feasibility problem can be formulated as minimizing a constant over the convex set* $S_t = \left\{\mathbf{q}^1 | \mathbf{q}^1 \in \Omega_1, \hat{\Pi}(\mathbf{q}^1) \geq t\right\}$ *(e.g.,* $\min_{\mathbf{q}^1 \in S_t} k$*) and solved using a convex optimization procedure;* S_t *is convex because* Ω_1 *is convex and the constraint* $\hat{\Pi}(\mathbf{q}^1) \geq t$ *represents a superlevel set of* $\hat{\Pi}(\mathbf{q}^1)$*, which due to quasiconcave* $\hat{\Pi}(\mathbf{q}^1)$*, is convex (see Boyd and Vandenberghe 2004, p. 95.)*

Step 3. If the above problem is feasible, then $\hat{\Pi}^* \geq t$ and set $\hat{\Pi}_L = t$; otherwise $\hat{\Pi}^* < t$ and set $\hat{\Pi}_H = t$.

Step 4. Repeat Steps 2-3 until $\hat{\Pi}_L$ and $\hat{\Pi}_H$ converge.

If the profit function is not quasiconcave, then the feasibility problem in Step 2 is not convex and the algorithm may not find a feasible solution even when one exists. In this case, Algorithm 1 is not assured to return an optimal solution. Instead, we may use a gradient descent procedure to obtain a price vector that is a stationary point of the profit function. Let $h(\mathbf{p}) = -\pi(\mathbf{p})$, which is the function we want to minimize in a gradient descent algorithm. Let $\hat{p}_i = p_i - c_i$ (margin of product i) and recall that $r_k = \sum_{i=1}^{n} \hat{p}_i q_{ik}$. Note that

$$\begin{aligned}
\frac{\partial q_{ik}}{\partial p_i} &= -b_{ik} q_{ik} \left(1 - q_{ik}\right), \qquad \frac{\partial q_{jk}}{\partial p_i} = b_{ik} q_{ik} q_{jk} \quad \text{for } i \neq j, \\
\frac{\partial q_i}{\partial p_i} &= -\sum_{k=1}^{m} w_k b_{ik} q_{ik} \left(1 - q_{ik}\right), \quad \frac{\partial q_j}{\partial p_i} = \sum_{k=1}^{m} w_k b_{ik} q_{ik} q_{jk} \quad \text{for } i \neq j.
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{\partial h\left(\hat{\mathbf{p}}\right)}{\partial \hat{p}_i} &= -\frac{\partial \pi\left(\mathbf{p}\right)}{\partial p_i} = \hat{p}_i \sum_{k=1}^{m} w_k b_{ik} q_{ik} - \sum_{k=1}^{m} w_k b_{ik} q_{ik} \sum_{j=1}^{n} \hat{p}_j q_{jk} - q_i \\
&= \left(\sum_{k=1}^{m} w_k b_{ik} q_{ik}\right) \left[\hat{p}_i - \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum_{l=1}^{m} w_l b_{il} q_{il}}\right) r_k - \frac{1}{\sum_{k=1}^{m} \left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}}\right].
\end{aligned}$$

Algorithm 2 (gradient descent)**.** *Step 1. Select values for an initial margin vector, e.g., $\hat{\mathbf{p}}^1 = (1/b_{11}, ..., 1/b_{n1})$ and let $t = 1$.*

Step 2. At the t^{th} iteration, compute the direction vector $\mathbf{d}^t$ as

$$d_i^t = \frac{1}{\sum_{k=1}^{m} \left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}} + \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum_{l=1}^{m} w_l b_{il} q_{il}} \sum_{j=1}^{n} \hat{p}_j^t q_{jk}\right) - \hat{p}_i^t \textit{ for all } i,$$

where q_{ik}, q_{0k} are functions of $\hat{\mathbf{p}}^t$, and compute the step size $\alpha^t \in [0,1]$ as

$$\alpha^t = \underset{\alpha \in [0,1]}{\arg\min}\, h\left(\hat{\mathbf{p}}^t + \alpha \mathbf{d}^t\right).$$

Step 3. Compute the new margin vector as $\hat{\mathbf{p}}^{t+1} = \hat{\mathbf{p}}^t + \alpha^t \mathbf{d}^t$.

Step 4. Increase t by 1 and repeat steps 2-4 until the markup vector converges.

Proposition 5. *Algorithm 2 converges to a stationary point of the price optimization problem.*

Since a unique optimal solution is not guaranteed when the profit function is not quasi-concave, it is necessary to apply Algorithm 2 with different starting price vectors to avoid a suboptimal stationary point solution. In particular, it can be shown from equation (3) that the optimal price $p_i, i = 1, 2, \ldots, n$ must be bounded in the interval

$$\left[c_i + \frac{1}{\max_k b_{ik}},\ c_i + \frac{1}{\min_k b_{ik}} + \max_k \rho_k\right] \tag{9}$$

where ρ_k is the optimal profit from a segment k customer if prices of all products are set to maximize segment k profit only. Specifically, ρ_k solves the single-variable equation (Li and Huh 2011, Theorem 2)

$$\rho_k = \sum_{j=1}^{n} \frac{e^{a_{jk} - b_{jk}c_{jk} - 1} e^{-b_{jk}\rho_k}}{b_{jk}} .$$

This single-variable equation is easily solved with a bisection search as the left side monotonically increases in ρ_k and the right side monotonically decreases in ρ_k. Therefore, by randomly generating starting price vectors from the above interval and applying Algorithm 2, one can obtain additional stationary points and compare the profits to identify the best among them. In theory, as long as the random distribution used to generate the starting values does not lead to a nonzero probability of repeatedly missing a subset with positive volume, the search will converge to the global optima (Solis and Wets, 1981). For example, a uniform distribution satisfies this requirement. In practice, with reasonably sufficient number of random starting values, a global optima can be achieved with high confidence.

3.3 Efficient Frontier of Profit and Market Share

A practical concern in pricing is balancing profit and market share objectives. At Intel, for example, senior management constantly shifts discussion between profit maximization and market share expansion. On one hand, the profit-maximizing pricing solution may not meet the firm's ambition on market share; on the other hand, the market share-maximizing prices reduce profit margins to nil which is also far from ideal. With this in mind, we describe how our optimization algorithm can be adapted to yield the efficient frontier of optimal profit by market share, allowing a firm to choose the sweet spot that reflects its market strategy.

Let $\bar{q}$ denote the firm's target total share and consider the problem

$$\max_{\mathbf{p}} \left\{ \pi(\mathbf{p}) | \sum_{j=1}^{n} q_j(\mathbf{p}) = \bar{q} \right\},$$

which can be equivalently expressed as maximizing the Lagrangian function, i.e.,

$$\max_{\mathbf{p},\gamma} L(\mathbf{p},\gamma) = \pi(\mathbf{p}) + \gamma \left(\sum_{j=1}^{n} q_j(\mathbf{p}) - \bar{q} \right).$$

Let $\gamma^*(\bar{q})$ denote the Lagrangian multiplier for a given target total share value $\bar{q}$. Note that $\gamma^* : (0,1) \to (-\infty, \infty)$ is a strictly increasing function (e.g., $\gamma^*(q^o) = 0$ where $q^o = \sum_j^n q_j(\mathbf{p}^o)$ and $\mathbf{p}^o$ denotes the optimal unconstrained price vector). Thus, we generate the efficient frontier by solving the following problem

$$L^*(\gamma) \;=\; \max_{\mathbf{p}} \left\{ \pi(\mathbf{p}) + \gamma \left(\sum_{j=1}^{n} q_j(\mathbf{p}) - \bar{q} \right) \right\}$$

for differing values of γ. Letting

$$\mathbf{p}^*(\gamma) \;=\; \operatorname*{argmax}_{\mathbf{p}} \left\{ \pi(\mathbf{p}) + \gamma \sum_{j=1}^{n} q_j(\mathbf{p}) \right\},$$

the efficient frontier is given by the curve $\left(\sum_{j=1}^{n} q_j(\mathbf{p}^*(\gamma)), \pi(\mathbf{p}^*(\gamma)) \right)$.

Next, we illustrate how the gradient search method can be used to obtain the efficient frontier. The gradient descent algorithm described above (Algorithm 2) identifies the gradient function for the special case of $\gamma = 0$. In this section, we identify the generalized gradient function. Let $h^L(\mathbf{p}) = -L(\mathbf{p})$ and define $\hat{p}_i = p_i - c_i$.

$$\begin{aligned}
\frac{\partial h^L(\mathbf{p},\gamma)}{\partial p_i} &= -\frac{\partial L(\mathbf{p},\gamma)}{\partial p_i} = (\hat{p}_i + \gamma) \sum_{k=1}^{m} w_k b_{ik} q_{ik} - \sum_{k=1}^{m} w_k b_{ik} q_{ik} \sum_{j=1}^{n} (\hat{p}_j + \gamma) q_{jk} - q_i \\
&= \left(\sum_{k=1}^{m} w_k b_{ik} q_{ik} \right) \left\{ \hat{p}_i + \gamma - \sum_{k=1}^{m} \left[\frac{w_k b_{ik} q_{ik}}{\sum_{\ell=1}^{m} w_\ell b_{i\ell} q_{i\ell}} \sum_{j=1}^{n} (\hat{p}_j + \gamma) q_{jk} \right] - \frac{1}{\sum_{\ell=1}^{m} \frac{w_\ell q_{i\ell}}{q_i} b_{i\ell}} \right\} \\
&= \left(\sum_{k=1}^{m} w_k b_{ik} q_{ik} \right) \left[\hat{p}_i - \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum_{\ell=1}^{m} w_\ell b_{i\ell} q_{i\ell}} \right) \left(\sum_{j=1}^{n} \hat{p}_j q_{jk} - \gamma q_{0k} \right) - \frac{1}{\sum_{\ell=1}^{m} \frac{w_\ell q_{i\ell}}{q_i} b_{i\ell}} \right].
\end{aligned}$$

The following algorithm solves for a stationary point for the optimization of $L(\mathbf{p}|\gamma)$ for a given γ.

Algorithm 3 (gradient descent for efficient frontier)**.** *Step 1. Select values for an initial margin vector, e.g.,* $\hat{\mathbf{p}}^1 = (1/b_{11}, ..., 1/b_{n1})$ *and let* $t = 1$*.*

Step 2. At the t^{th} *iteration, compute the direction vector* $\mathbf{d}^t$ *as*

$$d_i^t = \frac{1}{\sum_{k=1}^{m} \left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}} + \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum_{l=1}^{m} w_l b_{il} q_{il}} \right) \left(\sum_{j=1}^{n} \hat{p}_j^t q_{jk} - \gamma q_{0k} \right) - \hat{p}_i^t \text{ for all } i$$

where q_{ik}, q_{0k} *are functions of* $\hat{\mathbf{p}}^t$*, and compute the step size* $\alpha^t \in [0, 1]$ *as*

$$\alpha^t = \arg\min_{\alpha \in [0,1]} h\left(\hat{\mathbf{p}}^t + \alpha \mathbf{d}^t\right).$$

Step 3. Compute the new margin vector as $\hat{\mathbf{p}}^{t+1} = \hat{\mathbf{p}}^t + \alpha^t \mathbf{d}^t$*.*

Step 4. Increase t *by 1 and repeat steps 2-4 until the markup vector converges.*

Proposition 6. *Given* γ*, Algorithm 3 converges to a stationary point of* $L(\mathbf{p}|\gamma)$*.*

4 Application

In the following, we describe applications of the pricing solution under MMNL at Intel. Currently at Intel, product prices, along with performance expectations must be announced to customers well in advance of product release for these customers to make product design and purchasing decisions. As a direct consequence, Intel's pricing decisions are often performed with incomplete/uncertain information and are often made in light of the prices of the previous generation of Intel products qualitatively accounting for additional features in the new generation. Internally, there is a strong desire to quantify the impact of pricing decisions with available data.

We apply the model to Intel's microprocessor stock keeping units (SKUs) used in computer servers. Sales data of 16 SKUs spanning four generations of products were used as the initial study of the tools developed. In particular, the first three generations of products (13 SKUs) were used to parameterize the demand model and the fourth generation of products (3 SKUs) were used to test the demand model; we optimize prices for the three SKUs of generation 4 products. Products that are sold concurrently directly compete and form a choice set. Quarterly sales are scaled down by a constant factor and converted to one or multiple choice occasions based on the sales volume. For example, every 200 units is treated as one choice occasion. Then quarterly sales with a volume of 800 units are converted to four choice occasions with the chosen product as the choice decision among the corresponding choice set.

Customers are categorized into seven segments and the weights $w_k, k = 1, 2, \ldots, 7$ are computed based on historical purchasing volumes. The seven customer categories correspond to groupings used by Intel's sales division.

The list of independent variables considered includes processor cores, processor base frequency, TDP (a power consumption index), turbo frequency, performance (a commonly-adopted benchmark score from the Standard Performance Evaluation Corporation), cache, price and price/performance. The regressors for each customer segment are chosen based on the training and test data prediction accuracy. We use the SAS multinomial discrete choice (MDC) procedure to obtain segment-specific coefficients. Data fitting and parameterization details are provided in the online appendix which includes the coefficient table, as well as the training and test errors.

We then use the regression coefficients and the independent variable values to compute values for a_{ik} and b_{ik}. In particular, the non-price attributes of the products and the corresponding coefficients are used to obtain a_{ik}, i.e., $a_{ik} = \boldsymbol{\beta}_k \boldsymbol{x}_i$ where $\boldsymbol{x}_i$ is the vector of attribute values and $\boldsymbol{\beta}_k$ is the vector of coefficients. The price terms (price and price/performance) are used to compute b_{ik} values, i.e., $b_{ik} = -\left(\beta_k^p + \frac{\beta_k^I}{\text{performance}_i}\right)$ where β_k^p and β_k^I are the coefficients for price and price/performance respectively, and performance_i is the performance of product i. Since the no-purchase incidences are not observable, the MDC procedure is run with only Intel product alternatives.

For price optimization, however, we need to account for the no-purchase utilities. In the market of server processors, Intel products have dominant performances and thus a segment-specific no-purchase option is determined by computing the segment-specific utilities of recently retired Intel products comparable to the current choice set (i.e., retired products which would have belonged to the current choice set if still available), using the coefficients obtained from the regression. We then normalize the a_{ik} values of Intel products so that the corresponding no-purchase utility is zero (i.e., subtract the segment-specific no-purchase utility for each segment from the original a_{ik} values). Lastly, the normalized a_{ik} values and b_{ik} values are fed to the price optimization algorithm to compute the optimal price vector. The normalized price-independent utilities, i.e, a_{ik} values, and the price sensitivities of generation 4 products are given in Table 2 and Table 3 respectively. The segment weights (w_k's) are given in Table 4. In this application, the marginal costs c_i's are set to zero due

Table 2: a_{ik} Values.

i \ k	1	2	3	4	5	6	7
1	-1.0334	3.2480	-0.9336	1.7094	0.4187	-0.8904	-0.9804
2	0.7840	4.7161	-0.3438	1.8777	2.1771	-0.4310	-0.4907
3	6.0054	3.8771	1.3506	2.3611	1.1723	0.8889	0.9163

Table 3: b_{ik} Values.

i \ k	1	2	3	4	5	6	7
1	0.00416	0.01840	0.00525	0.01165	0.01015	0.00325	0.00331
2	0.00312	0.01354	0.00394	0.00874	0.00639	0.00244	0.00248
3	0.00181	0.00744	0.00229	0.00508	0.00167	0.00142	0.00144

Table 4: w_k Values

k	1	2	3	4	5	6	7
w_k	0.0753	0.1126	0.1285	0.1180	0.0859	0.2842	0.1953

to Intel's high volume production in which processors are produced in large lots and fixed cost dominates.

Algorithm 3 is implemented for Intel application to obtain the profit-market share efficient frontier, and the corresponding optimal prices for any given market share target. For each share target, we use 30 randomly selected starting values and convergence to a stationary point is achieved for all instances. Convergence takes between 18 and 41 iterations.

Figure 3 illustrates the efficient frontier of profit versus total market share. Recall that γ is the Lagrangian multiplier associated with a given total share value. The case of $\gamma = 0$ corresponds to the unconstrained optimal solution. The profit and market share under Intel's current prices are noted with $*$ in Figure 3. As the desired total market share increases (equivalently, as γ increases), the optimal achievable profit decreases and the profit decline is steeper at higher market shares. Interestingly, Intel's current prices sit quite close to the efficient frontier. That is, if Intel is to maintain the current total market share, the current prices perform well. The room for improvement without compromising on market share is 5.3% (i.e., moving from current practice vertically up to the efficient frontier). Alternatively, Intel may improve the total market share by 2.1% without compromising profit by moving

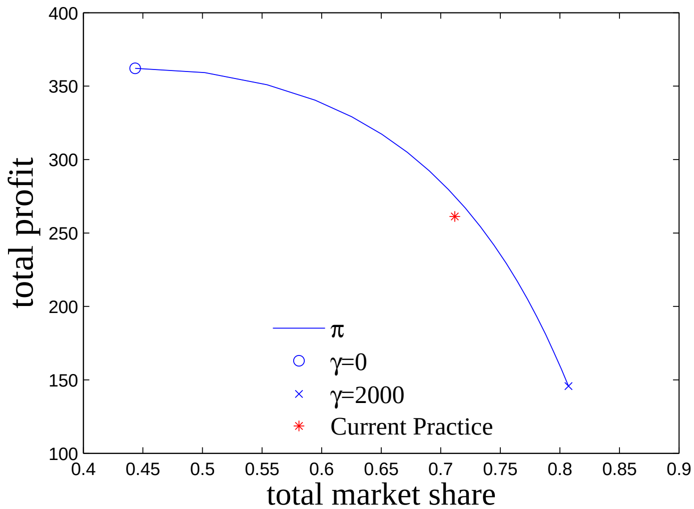


Figure 3: Profit – Market Share Efficient Frontier for Intel Application

from the current practice horizontally to the efficient frontier. Table 5 presents these options.

Table 5: Current versus Efficient Frontier Solutions (Profit-improving maintains same total market share as the current practice whereas Share-improving maintains same profit as the current practice).

Price	Current	Profit-improving	Share-improving
p_1	$207	$120	$107
p_2	$299	$182	$169
p_3	$410	$649	$601
total profit	$261.2	$275.1	$261.2
total market share	0.7117	0.7117	0.7265

Figure 4 illustrates the distribution of sales among the three products under the current and the profit-improving prices respectively. Figure 5 illustrates how the total profit distributes among different customer segments, which reveals an important underlying mechanism that drives profit improvement. Note that, compared to the distribution under the current prices, profit shares for segments 1, 3, 6 and 7 increase while those for segments 2,4, and 5 decrease under the profit-improving prices. That is, by adjusting the product prices to tailor to the preferences of certain customers, Intel can generate a larger total profit. Therefore, profit improvement is in part a consequence of exploiting segment differences and using price as a means for redistributing sales and profits among customer segments. Segment-specific sales distribution provides additional supporting evidence for this and is given in the online appendix.

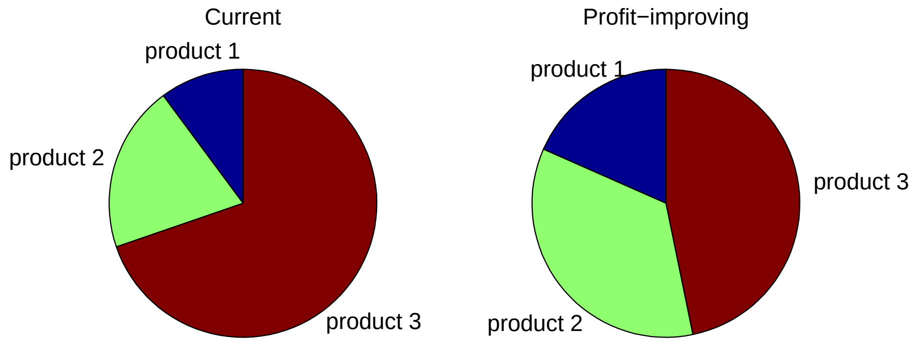


Figure 4: Sales Distribution among Products (Comparison of the Current Practice and the Profit-improving Solution)

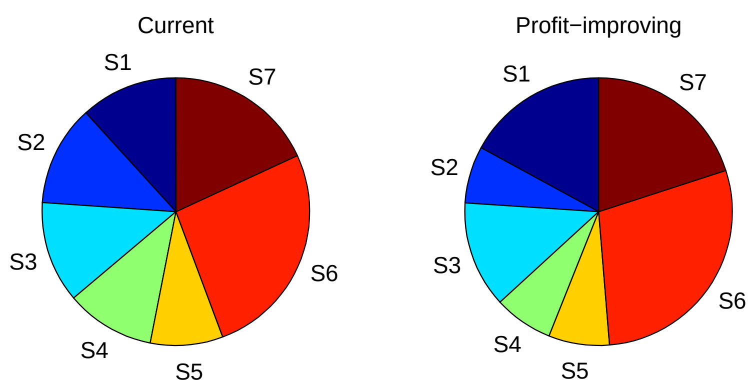
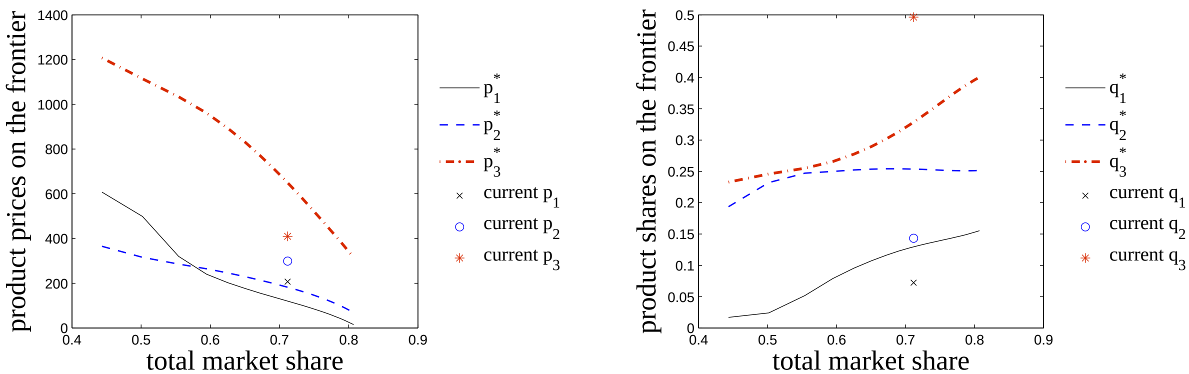


Figure 5: Profit Distribution Among Customer Segments (Comparison of the Current Practice and the Profit-improving Solution)

Intel management is pleased to observe that the current prices perform well (i.e., close to the efficient frontier). They are also enthusiastic about the ability to visualize where Intel is positioned along the entire spectrum of the the efficient frontier and understand how movement along the efficient frontier would help or hurt Intel's objective. For example, while adopting an unconstrained price solution (noted by "$\circ$" in Figure 3) can lead to significant one-time profit increase, losing a substantial amount of market share is not necessarily the strategy that Intel wishes to pursue. The efficient frontier allows Intel to choose a pricing strategy that effectively balances its goals of profit versus market share.

Figures 6(a) and 6(b) present the corresponding prices and product shares along the efficient frontier. The current prices and product shares are also noted in the figures.



(a) Optimal Prices on the Efficient Frontier

(b) Product Shares on the Efficient Frontier

Figure 6: Efficient Frontier Solution (Optimal Prices and Product Shares)

Two observations are salient. (1) As the target total market share increases, the optimal prices of all products decrease and the resulting product shares increase, which is expected. (2) At levels close to the current total market share, the sequence of the optimal prices are consistent with the order of the preference values of the products, i.e., products with higher a values are priced higher; however, when the target market share is low, the price sequence of products 1 and 2 flips.

Observation (2) is counterintuitive. Lemma 1 provides a plausible theoretical explanation for the price sequence reversal. In the case of Intel products, product 1's preference values are low and are often significantly lower than the no-purchase option (see Table 2). Therefore, it is unlikely to be a high-volume product. However, when market share is less of a concern, the firm can take advantage of the segment differences and essentially turn it from a "cheap" product into a "niche" product by charging a higher price and only focusing on customers who value it more than others.

The methods presented in this paper serve multiple purposes at Intel. First, they provide a new alternative for market share prediction among Intel products for different customer segments, adding to the suite of independent demand forecasting tools. Second, they optimize product prices based on segment-specific customer preferences revealed through sales data, a capability that Intel's current pricing tools include only heuristically. Third, they enable the company to quantitatively balance the tradeoff between profit and market share. Furthermore, the ability to locate the current pricing strategy relative to the efficient frontier allows Intel to identify practical improvement opportunities. Thus far, we have applied the model to one product family which includes 16 SKUS and spans four generations. This product family belongs to the server processor market which has total sales of $16 billion. The next step is to apply this to other product families and markets. Ultimately we expect the efficient frontier to become a useful aid for strategic analysis and decision support.

5 Conclusion

We consider optimal pricing under a discrete specification of the MMNL demand model, which aligns with a context in which the customer population consists of a finite number of segments, each of which has specific choice preferences.

We show that the well-known equal markup property identified for the MNL model with symmetric sensitivities does not hold under MMNL even for entirely symmetric price sensitivities across *all* products and *all* segments. This suggests that customer heterogeneity in preferences towards price-independent product attractiveness alone justifies differentiated markups. In addition, the concavity property with respect to the choice probability vector shown in prior research for the MNL and NL model breaks down under MMNL and we demonstrate with examples how the profit function might shift from concave to nonconcave functions as the model primitives change. The analysis that leads to a unique solution under MNL or NL does not carry through to MMNL. In this paper, we characterize the profit function under MMNL as the sum of a set of quasiconcave functions and present efficient optimization algorithms for the pricing problem. In addition, we present a solution method for computing the efficient frontier of profit against total market share, which broadens the applicability of the model in a practical setting. Moreover, we apply our methods using data from Intel Corporation and show that, by adjusting the prices of the products to tailor to the preferences of different customer segments, Intel can redistribute sales and profits among the segments to generate a larger total profit. The efficient frontier of profit against market share enables Intel to examine its pricing decision in light of the firm's desired balance between the two goals.

The MMNL model can approximate any discrete choice model consistent with random utility maximization (RUM) to any degree of precision (McFadden and Train, 2000). Thus, the theoretical results and solution approaches derived in this paper suggest a path to solving the pricing problem under a general discrete choice model that is consistent with RUM.

References

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

Aksoy-Pierson, M., G. Allon, and A. Federgruen. 2013. Price competition under mixed multinomial logit demand functions. *Management Science* 59 (8): 1817–1835.

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

Aydin, G., and E. L. Porteus. 2008. Joint Inventory and Pricing Decisions for an Assortment. *Operations Research* 56 (5): 1247–1255.

Bertsekas, D. 2003. *Nonlinear programming.* Belmont, MA: Athena Scientific.

Blanchet, J., G. Gallego, and V. Goyal. 2013. A markov chain approximation to choice modeling. In *EC*, 103–104. Citeseer.

Boyd, J., and R. Mellman. 1980. The effect of fuel economy standards on the U.S. automotive market: an hedonic demand analysis. *Transportation Research* 14A (5-6): 367–378.

Boyd, S., and L. Vandenberghe. 2004. *Convex optimization.* Cambridge, UK: Cambridge University Press.

Cardell, N., and F. Dunbar. 1980. Measuring the societal impacts of automobile downsizing. *Transportation Research* 14A (5-6): 423–434.

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

Eek, M., and K. Berggren. 2007. The Emerging Pricing Capability–A Multi Case Study of Business-to-Business Companies.

Feldman, J., and H. Topaloglu. 2015. Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management* 24 (10): 1598–1620.

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

Greene, W. H., and D. A. Hensher. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37 (8): 681–698.

Guadagni, P. M., and J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2 (3): 203–238.

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

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

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

Kamakura, W. A., and G. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26:379–390.

Kök, A. G., and Y. Xu. 2011. Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science* 57 (9): 1546–1563.

Kouvelis, P., Y. Xiao, and N. Yang. 2015. PBM competition in pharmaceutical supply chain: Formulary design and drug pricing. *Manufacturing and Service Operations Management* 17 (4): 511–526.

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

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

Maddah, B., and E. K. Bish. 2007. Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics* 54 (3): 315–330.

McFadden, D. 1974. The measurement of urban travel demand. *Journal of public economics* 3 (4): 303–328.

McFadden, D. 1978. Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models*, ed. A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, 75–96. North-Holland, Amsterdam.

McFadden, D. 1980. Econometric models for probabilistic choice among products. *Journal of Business* 53 (3): S13–S29.

McFadden, D., and K. Train. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15 (5): 447–470.

Mookherjee, R., J. Mukherjee, J. Martineau, L. Xu, M. Gullo, K. Zhou, A. Hazlewood, X. Zhang, F. Griarte, and N. Li. 2016. End-to-End Predictive Analytics and Optimization in Ingram Micros Two-Tier Distribution Business. *Interfaces* 46 (1): 49–73.

Noble, P. M., and T. S. Gruca. 1999. Industrial pricing: Theory and managerial practice. *Marketing science* 18 (3): 435–454.

Phillips, R. L. 2005. *Pricing and revenue optimization*. Stanford University Press.

Rayfield, W., P. Rusmevichientong, and H. Topaloglu. 2015. Approximation methods for pricing problems under the nested logit model with price bounds. *INFORMS Journal on Computing* 27 (2): 335–357.

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

Silva-Risso, J., and I. Ionova. 2008. Practice Prize Winner-A Nested Logit Model of Product and Transaction-Type Choice for Planning Automakers' Pricing and Promotions. *Marketing Science* 27 (4): 545–566.

Silva-Risso, J., W. V. Shearin, I. Ionova, A. Khavaev, and D. Borrego. 2008. Chrysler and JD Power: Pioneering scientific price customization in the automobile industry. *Interfaces* 38 (1): 26–39.

Solis, F. J., and R. J.-B. Wets. 1981. Minimization by random search techniques. *Mathematics of Operations Research* 6 (1): 19–30.

Song, J.-S., and Z. Xue. 2007. Demand management and inventory control for substitutable products. *Working Paper, Duke University*.

Train, K. 2003. *Discrete choice methods with simulation.* Cambridge University Press.

Train, K. E., D. L. McFadden, and M. Ben-Akiva. 1987. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *RAND Journal of Economics* 18 (1): 109–123.

Wang, R. 2012. Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters* 40 (6): 492–497.

Product-line Pricing under Discrete Mixed Multinomial Logit Demand

Hongmin Li • Scott Webster • Nicholas Mason • Karl Kempf

A Online Appendix

A.1 Proof of Non-equal Markup

Lemma 2. *Assume $b_{ik} = b$ for all i, k. Let p_i^* be the optimal price for product i. Then in general, $p_i^* - c_i \neq p_j^* - c_j$ for $i \neq j$.*

Proof. Since $b_{ik} = b$, the first-order optimality condition becomes

$$p_i - c_i = \frac{1}{b} + \sum_k \frac{w_k q_{ik}}{q_i} r_k \ . \tag{10}$$

Note that

$$\sum_k \frac{w_k q_{ik}}{q_i} r_k = \frac{\sum_k w_k q_{0k} A_{ik} e^{-bp_i} r_k}{\sum_{k'} w_{k'} q_{0k'} A_{ik'} e^{-bp_i}} = \sum_k \frac{w_k q_{0k} A_{ik}}{\sum_{k'} w_{k'} q_{0k'} A_{ik'}} r_k$$

is a weighted average of r_k with the weights given by $\frac{w_k q_{0k} A_{ik}}{\sum_{k'} w_{k'} q_{0k'} A_{ik'}}$. Since $A_{ik} \neq A_{ik'}$ for $k \neq k'$ (otherwise, segments k and k' become degenerate and are considered the same segment), the weights depend on the product index i.

Assume for contradiction that $p_i^* - c_i = \theta$ for all i. Then

$$\begin{aligned} r_k(\mathbf{p}^*) &= \sum_{i'} (p_{i'}^* - c_{i'}) q_{i'k}(\mathbf{p}^*) = \theta \sum_{i'} q_{i'k}(\mathbf{p}^*) = \theta(1 - q_{0k}(\mathbf{p}^*)) \\ &= \theta \left(1 - \frac{1}{1 + \sum_{j=1}^n A_{jk} e^{-bp_j^*}} \right) = \theta \left(1 - \frac{1}{1 + e^{-b\theta} \sum_{j=1}^n A_{jk} e^{-bc_j}} \right) \end{aligned}$$

whose value depends on the segment index k, thus in general r_k's are not equal across segments. Since the right side of equation (10) is a weighted average of the vector $(r_1, r_2, \ldots, r_m)$ with nonequal values and the weights depend on the product index i, this weighted average is a value that depends on i. This contradicts the assumption that $p_i^* - c_i = \theta$ for all i. ☐

A.2 Proof of Lemma 1

Proof. Since $b_{ik} = b$, the first-order optimality condition becomes

$$\begin{aligned} p_i - c_i &= \frac{1}{b} + \frac{w_A q_{iA}}{q_i} r_A + \frac{w_B q_{iB}}{q_i} r_B \\ &= \frac{1}{b} + r_B + \frac{w_A q_{iA}}{q_i} (r_A - r_B) \ . \end{aligned}$$

Thus $p_i^* - c_i \geq p_j^* - c_j$ if and only if

$$\left(\frac{w_A q_{iA}}{q_i} - \frac{w_A q_{jA}}{q_j} \right) (r_A - r_B) \geq 0 \ .$$

It is easy to verify that $\left(\frac{w_A q_{iA}}{q_i} - \frac{w_A q_{jA}}{q_j} \right)$ has the same sign as $[(a_{iA} - a_{iB}) - (a_{jA} - a_{jB})]$.

$$\Big(\text{Note that } \frac{w_A q_{iA}}{q_i} \geq \frac{w_A q_{jA}}{q_j} \Leftrightarrow \frac{q_i}{w_A q_{iA}} \leq \frac{q_j}{w_A q_{jA}} \Leftrightarrow \frac{w_A q_{iA} + w_B q_{iB}}{w_A q_{iA}} \leq \frac{w_A q_{jA} + w_B q_{jB}}{w_A q_{jA}}$$

$$\Leftrightarrow \frac{q_{iB}}{q_{iA}} \leq \frac{q_{jB}}{q_{jA}} \Leftrightarrow \frac{q_{iB}}{q_{jB}} \leq \frac{q_{iA}}{q_{jA}} \Leftrightarrow \frac{e^{a_{iB} - b_i p_i}}{e^{a_{jB} - b_j p_j}} \leq \frac{e^{a_{iA} - b_i p_i}}{e^{a_{jA} - b_j p_j}} \Leftrightarrow a_{iA} - a_{iB} \geq a_{jA} - a_{jB} \ . \Big)$$

Therefore, $p_i^* - c_i \geq p_j^* - c_j$ if and only if

$$[(a_{iA} - a_{iB}) - (a_{jA} - a_{jB})] (r_A - r_B) \geq 0 \quad \text{for } i \neq j.$$

□

A.3 An Example with Preference-value-inconsistent Optimal Prices

Consider a two-product (products 1 and 2) two-segment (segments A and B) example with $w_A = w_B = 0.5, b_{ik} = 1, c_i = 0, a_{1A} = 6, a_{2A} = 5, a_{1B} = 3, a_{2B} = 1$. Product 1 has higher price-independent utility values than product 2 for both segments of customers. The optimal prices for product 1 and product 2 are $p_1 = 4.011, p_2 = 4.387$, which is a sequence that is the opposite of the preference value sequence.

A.4 Proof of Proposition 1

Proof. To simplify presentation, we suppress the product subscript in our notation (e.g., q_k in place of q_{1k}). Note that the purchase probability of the product by a segment k customer is $\mathbf{q}^k = q_k$. Accordingly,

$$f_k(q_1) = \frac{A_k \left(\frac{q_1}{A_1(1-q_1)} \right)}{1 + A_k \left(\frac{q_1}{A_1(1-q_1)} \right)} \ ,$$

$$g_k(q_k) = \frac{1}{b}\log\left(\frac{A_k(1-q_k)}{q_k}\right) ,$$

and the profit contribution from a segment k customer as a function of q_1 simplifies to

$$\hat{R}_k(q_1) = (g_k(f_k(q_1)) - c)\, f_k(q_1) = \left(\frac{\log A_k - \log(\lambda_k q_1) + \log(1-q_1) - bc}{b}\right)\frac{\lambda_k q_1}{1 - q_1 + \lambda_k q_1}$$

where $\lambda_k = A_k/A_1$. We can derive that

$$-bz\frac{\partial^2 \hat{R}_k}{\partial q_1^2} = \frac{\lambda_k^2}{x} + \frac{2\lambda_k}{y} + \frac{x}{y^2} + 2L(\lambda_k - 1)^2\left(\frac{1}{z} - \frac{x}{z^2}\right) + \frac{2}{z}(\lambda_k - 1)(L - \lambda_k) - \frac{2x}{yz}(\lambda_k - 1) ,$$

where $x := \lambda_k q_1$, $y := q_{01}$, $z := \lambda_k q_1 + q_{01}$, and $L := \log A_k - \log(\lambda_k q_1) + \log q_{01} - bc$. Assume without loss of generality that $\lambda_k \geq 1$. From $\frac{\max_k A_k}{\min_k A_k} \leq 2$, we have $\lambda_k \leq 2$, equivalently, $\lambda_k \geq 2(\lambda_k - 1)$. Therefore,

$$\begin{aligned}
-bz\frac{\partial^2 R_k}{\partial {q_1}^2} &\geq \frac{\lambda_k^2}{x} + \frac{2\lambda_k}{y} + \frac{x}{y^2} + \frac{2}{z}L(\lambda_k - 1)^2 - \frac{2x}{z^2}L(\lambda_k - 1)^2 - \frac{2}{z}\lambda_k(\lambda_k - 1) - \frac{2x}{yz}(\lambda_k - 1) \\
&\geq \frac{\lambda_k^2}{x} + \frac{2\lambda_k}{y} + \frac{x}{y^2} - \frac{2}{z}\lambda_k(\lambda_k - 1) - \frac{2x}{yz}(\lambda_k - 1) \\
&= \lambda_k\left[\frac{\lambda_k}{x} - \frac{2}{z}(\lambda_k - 1)\right] + \frac{2}{y}\left[\lambda_k - \frac{x}{z}(\lambda_k - 1)\right] + \frac{x}{y^2} \\
&\geq \frac{x}{y^2} \geq 0
\end{aligned}$$

where the first inequality holds due to $L > 0$ (i.e., $L = b(p - c) > 0$), the second inequality holds because $\frac{x}{z} \leq 1$, and the third inequality holds due to $\lambda_k \geq 2(\lambda_k - 1)$ and $x \leq z$. Therefore, $\hat{R}_k$ is concave on Ω_1. The weighted sum of concave functions is concave, and thus $\hat{\Pi}$ is concave on Ω_1. □

A.5 Proof of Proposition 2

Proof. From (7), to establish the concavity of $\hat{\Pi}(\mathbf{q}^1)$, it suffices to show concavity of $\hat{R}_k(\mathbf{q}^1)$. Recall that $f_k(\mathbf{q}^1)$ is the vector of product purchase probabilities for segment k as a function of vector $\mathbf{q}^1$. Let $f_{ik}(\mathbf{q}^1)$ denote the ith element in $f_k(\mathbf{q}^1)$. Then we can write profit contribution of product i in segment k as $\hat{R}_{ik}(\mathbf{q}^1) = (g_{ik}(f_k(\mathbf{q}^1)) - c_i)f_{ik}(\mathbf{q}^1)$ and the segment-k profit as $\hat{R}_k(\mathbf{q}^1) = \sum_{i=1}^{n} \hat{R}_{ik}(\mathbf{q}^1)$.

From (4) and (6),

$$
\begin{aligned}
b_i \hat{R}_{ik}\left(\mathbf{q}^1\right) &= \frac{A_{ik}\left(\frac{q_{i1}}{A_{i1}q_{01}}\right)}{1+\sum_{j=1}^{n} A_{jk}\left(\frac{q_{j1}}{A_{j1}q_{01}}\right)} \left[\log\left(\frac{A_{i1}q_{01}}{q_{i1}}\right) - b_i c_i\right] \\
&= \frac{\frac{A_{ik}q_{i1}}{A_{i1}q_{01}}}{1+\sum_{j=1}^{n} \left(\frac{A_{jk}q_{j1}}{A_{j1}q_{01}}\right)} \left[\log A_{ik} - \log\left(\frac{A_{ik}q_{i1}}{A_{i1}q_{01}}\right) - b_i c_i\right] \\
&= \frac{\lambda_{ik}q_{i1}}{q_{01}+\sum_{j=1}^{n} (\lambda_{jk}q_{j1})} \left[\log A_{ik} - \log(\lambda_{ik}q_{i1}) + \log q_{01} - b_i c_i\right]
\end{aligned}
$$

where $\lambda_{ik} = A_{ik}/A_{i1}$ and $q_{01} = 1 - \sum_{i'=1}^{n} q_{i'1}$. For a given segment k, define $Z_i\left(\mathbf{q}^1\right) := b_i \hat{R}_{ik}\left(\mathbf{q}^1\right)$. We can derive that

$$
\begin{aligned}
\frac{\partial^2 Z_i}{\partial q_{\ell 1}^2} &= -\frac{x_i}{y^2 z} + \frac{2x_i}{z^3}(\lambda_{\ell k}-1)^2 L_i + \frac{2x_i}{yz^2}(\lambda_{\ell k}-1) \quad \ell \neq i \\
\frac{\partial^2 Z_i}{\partial q_{i1}^2} &= -\frac{x_i}{y^2 z} + \frac{2x_i}{z^3}(\lambda_{ik}-1)^2 L_i + \frac{2x_i}{yz^2}(\lambda_{ik}-1) - \frac{\lambda_{ik}^2}{x_i z} - \frac{2\lambda_{ik}}{yz} - \frac{2\lambda_{ik}(\lambda_{ik}-1)}{z^2}(L_i - 1) \\
\frac{\partial^2 Z_i}{\partial q_{\ell 1}\partial q_{j1}} &= -\frac{x_i}{y^2 z} + \frac{2x_i}{z^3}(\lambda_{\ell k}-1)(\lambda_{jk}-1) L_i + \frac{x_i}{yz^2}(\lambda_{\ell k}-1+\lambda_{jk}-1) \quad \ell, j \neq i \\
\frac{\partial^2 Z_i}{\partial q_{i1}\partial q_{\ell 1}} &= -\frac{x_i}{y^2 z} + \frac{2x_i}{z^3}(\lambda_{ik}-1)(\lambda_{\ell k}-1) L_i + \frac{x_i}{yz^2}(\lambda_{ik}-1+\lambda_{\ell k}-1) \\
&\quad -\frac{\lambda_{ik}}{yz} - \frac{1}{z^2}\lambda_{ik}(\lambda_{\ell k}-1)(L_i-1) \quad \ell \neq i\ ,
\end{aligned}
$$

where $x_i := \lambda_{ik}q_{i1}$, $y := q_{01}$, $z := q_{01} + \sum_j \lambda_{jk}q_{j1}$, and $L_i := \log A_{ik} - \log x_i + \log y - b_i c_i = b_i(p_i - c_i)$. The Hessian of $\hat{R}_k\left(\mathbf{q}^1\right) = \sum_{i=1}^{n} \hat{R}_{ik}(\mathbf{q}^1) = \sum_{i=1}^{n} Z_i(\mathbf{q}^1)/b_i$ is

$$
H\left(\mathbf{q}^1\right) = \begin{bmatrix}
\frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{11}^2}, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{11}\partial q_{21}}, & \ldots, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{11}\partial q_{n1}} \\
\frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{21}\partial q_{11}}, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{21}^2}, & \ldots, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{21}\partial q_{n1}} \\
\cdots & & & \\
\frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{n1}\partial q_{11}}, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{n1}\partial q_{21}}, & \ldots, & \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{n1}\partial q_{n1}}
\end{bmatrix}.
$$

The function $\hat{R}_k(\mathbf{q}^1)$ is concave on Ω_1 if and only if $\boldsymbol{\theta}^{\mathrm{T}} H \boldsymbol{\theta} < 0$ for any nonzero vector $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \theta_2, \ldots, \theta_n)$ and any $\mathbf{q}^1 \in \Omega_1$. Note that

$$\begin{aligned}
\boldsymbol{\theta}^{\mathrm{T}} H \boldsymbol{\theta} &= \sum_{\ell=1}^{n} \theta_\ell^2 \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{\ell 1}^2} + \sum_{\ell=1}^{n} \sum_{j \neq \ell}^{n} \theta_\ell \theta_j \frac{\partial^2 \sum_{i=1}^n Z_i/b_i}{\partial q_{\ell 1} \partial q_{j1}} \\
&= -\sum_i \frac{1}{b_i x_i z} \left(\theta_i \lambda_{ik} + \frac{x_i}{y} \sum_\ell \theta_\ell \right)^2 - \sum_i \frac{L_i - 1}{b_i z^2} 2\theta_i \lambda_{ik} \left[\sum_\ell \theta_\ell (\lambda_{\ell k} - 1) \right] \\
&\quad + \sum_i \frac{2 x_i L_i}{b_i z^3} \left[\sum_\ell \theta_\ell (\lambda_{\ell k} - 1) \right]^2 \\
&\quad + \sum_i \frac{2 x_i}{b_i y z^2} \left[\sum_\ell \theta_\ell^2 (\lambda_{\ell k} - 1) + \sum_\ell \sum_{j \neq \ell} \theta_\ell \theta_j (\lambda_{\ell k} - 1 + \lambda_{jk} - 1) \right] .
\end{aligned}$$

Define $G(\boldsymbol{\lambda}) := \boldsymbol{\theta}^{\mathrm{T}} H \boldsymbol{\theta}$ where $\boldsymbol{\lambda} = [\lambda_{ik}]$ for $i = 1, ..., n$ and $k = 1, ..., m$ (i.e., $\boldsymbol{\lambda}$ is an $n \times m$ matrix). The function G has a strict negative value at $\boldsymbol{\lambda} = \mathbf{1}$ (because it is not possible to find a nonzero vector $\boldsymbol{\theta}$ such that $\theta_i \lambda_{ik} + \frac{x_i}{y} \sum_\ell \theta_\ell = 0$ for all i). G is continuous in $\boldsymbol{\lambda}$. So there must exist a rectangle region near $\boldsymbol{\lambda} = \mathbf{1}$ in which the values of G stay negative. □

A.6 Proof of Proposition 3

Proof. The concavity of $\overline{\Pi}(\mathbf{q}^1)$ and $\underline{\Pi}(\mathbf{q}^m)$ follows from Lemma 2 in Li and Huh (2011). From (2) and $A_{i1} \leq A_{ik} \leq A_{im}$, $q_{01} \geq q_{0k} \geq q_{0m}$. From (4),

$$p_i = \frac{1}{b_i} \left[\log A_{i1} - \log q_{i1} + \log \left(1 - \sum_j q_{j1} \right) \right] . \tag{11}$$

Since $q_{ik} = A_{ik} q_{0k} \frac{q_{i1}}{A_{i1} q_{01}}$,

$$\begin{aligned}
\hat{\Pi}(\mathbf{q}^1) &= \sum_i (p_i(\mathbf{q}^1) - c_i) \sum_k w_k q_{ik} = \sum_i (g_{i1}(\mathbf{q}^1) - c_i) \sum_k w_k A_{ik} q_{0k} \left(\frac{q_{i1}}{A_{i1} q_{01}} \right) \\
&= \sum_i \frac{q_{i1}}{b_i} \left[\log A_{i1} - \log q_{i1} + \log \left(1 - \sum_j q_{j1} \right) - b_i c_i \right] \sum_k w_k \left(\frac{A_{ik} q_{0k}}{A_{i1} q_{01}} \right) \\
&\leq \sum_i \frac{q_{i1}}{b_i} \left[\log A_{i1} - \log q_{i1} + \log \left(1 - \sum_j q_{j1} \right) - b_i c_i \right] \sum_k w_k \left(\frac{A_{ik}}{A_{i1}} \right) \\
&= \overline{\Pi}(\mathbf{q}^1) .
\end{aligned}$$

We can similarly show that $\tilde{\Pi}(\mathbf{q}^m) \geq \underline{\Pi}(\mathbf{q}^m)$. □

A.7 Corollary 1 (Upper Bounds of $\overline{\Pi}(\mathbf{q}^1)$ and $\underline{\Pi}(\mathbf{q}^m)$)

Corollary 1. *Let $\overline{\mathbf{q}^1} = argmax_{\mathbf{q}^1}\,\overline{\Pi}(\mathbf{q}^1)$ and $\overline{\mathbf{p}^1}$ be the corresponding price vector. In addition, let $\overline{\mathbf{q}^m} = argmax_{\mathbf{q}^m}\,\underline{\Pi}(\mathbf{q}^m)$ and $\overline{\mathbf{p}^m}$ be the corresponding price vector.*
(i) The maximum of $\overline{\Pi}(\mathbf{q}^1)$ is given by $\overline{\theta}$ where $\overline{\theta}$ is the unique solution to the single-variable equation

$$\overline{\theta} = \sum_i \left(\frac{e^{a_{i1} - b_i c_i - 1 - \frac{b_i \overline{\theta}}{\sum_k w_k A_{ik}/A_{i1}}}}{b_i} \sum_k w_k A_{ik}/A_{i1} \right) .$$

(ii) The maximum of $\underline{\Pi}(\mathbf{q}^m)$ is given by $\underline{\theta}$ where $\underline{\theta}$ is the unique solution to the single-variable equation

$$\underline{\theta} = \sum_i \left(\frac{e^{a_{im} - b_i c_i - 1 - \frac{b_i \underline{\theta}}{\sum_k w_k A_{ik}/A_{im}}}}{b_i} \sum_k w_k A_{ik}/A_{im} \right) .$$

Proof. Since $\overline{\Pi}(\mathbf{q}^1)$ is concave in $\mathbf{q}^1$, we take the first order derivative with respect to q_{j1}, and set it to zero to obtain the first-order condition

$$p_j - c_j = \frac{1}{b_j} + \frac{\overline{\theta}}{\sum_k w_k A_{jk}/A_{j1}} , \tag{12}$$

where

$$\overline{\theta} = \sum_i \frac{q_{i1}/q_{01}}{b_i} \sum_k w_k A_{ik}/A_{i1} . \tag{13}$$

Thus

$$q_{i1}/q_{01} = e^{a_{i1} - b_i p_i} = e^{a_{i1} - b_i c_i - 1 - b_i \frac{\overline{\theta}}{\sum_k w_k A_{ik}/A_{i1}}} . \tag{14}$$

From (13) and (14), we have

$$\overline{\theta} = \sum_i \frac{e^{a_{i1} - b_i c_i - 1 - \frac{b_i \overline{\theta}}{\sum_k w_k A_{ik}/A_{i1}}}}{b_i} \sum_k w_k A_{ik}/A_{i1} .$$

Therefore, the profit $\overline{\Pi}$ can be rewritten as

$$\begin{aligned}
\overline{\Pi} &= \sum_i (p_i - c_i) q_{i1} \sum_k w_k A_{ik}/A_{i1} \\
&= \sum_i \left(\frac{1}{b_i} + \frac{\overline{\theta}}{\sum_k w_k A_{ik}/A_{i1}} \right) q_{i1} \sum_k w_k A_{ik}/A_{i1} \\
&= \left(\sum_i \frac{q_{i1}/q_{01}}{b_i} \sum_k w_k A_{ik}/A_{i1} \right) q_{01} + \overline{\theta} \sum_i q_{i1} \\
&= \overline{\theta} q_{01} + \overline{\theta} \sum_i q_{i1} = \overline{\theta} ,
\end{aligned}$$

where the second equality follows from (12) and the fourth equality follows from (13). This proves (i). The proof of (ii) follows a similar argument. □

A.8 Proof of Proposition 4

Proof. Note that $f_1(\mathbf{q}^1) = \mathbf{q}^1$. Thus $\hat{R}_1(\mathbf{q}^1) = R_1(f_1(\mathbf{q}^1)) = R_1(\mathbf{q}^1)$ is a profit function based on MNL demand which, as noted above, is concave.

What remains is to show that $\hat{R}_k(\mathbf{q}^1)$ is quasiconcave for $k \geq 2$. Without loss of generality, we set $k = 2$. Let us outline the main steps of our proof. We first show that $\Omega_4 := \{f_2(\mathbf{q}^1) | \mathbf{q}^1 \in \Omega_1\}$ is a convex set by decomposing function f_2 into a composition of more elementary functions, and show that each of these functions preserves convexity. We then explain why the convexity of Ω_4 implies that superlevel set $S_\alpha(R_2, \Omega_4) = \{\mathbf{q}^2 \in \Omega_4 | R_2(\mathbf{q}^2) \geq \alpha\}$ is convex. Finally, we show that the inverse image $S_\alpha(R_2, \Omega_4)$ under function f_2 is convex (using a similar decomposition approach), which implies that superlevel set $S_\alpha\left(\hat{R}_2, \Omega_1\right) = \left\{\mathbf{q}^1 \in \Omega_1 | \hat{R}_2(\mathbf{q}^1) \geq \alpha\right\}$ is convex, thereby proving that $\hat{R}_2$ is quasiconcave. Our proof will rely on the following definitions, remark, and property.

Definition 1. Function $f : R^n \to R$ is *quasiconcave* if its domain is convex and its superlevel sets $S_\alpha(f, \mathbf{dom}\, f) = \{\mathbf{x} \in \mathbf{dom}\, f | f(\mathbf{x}) \geq \alpha\}$ are convex for all $\alpha \in R$ (Boyd and Vandenberghe 2004, p. 95).

Definition 2. Let $A \in R^{n \times m}, \mathbf{b} \in R^n, \mathbf{c} \in R^m, d \in R$. Function $f : R^m \to R^n$ with $f(\mathbf{x}) = (A\mathbf{x} + \mathbf{b})/(\mathbf{c}^{\mathrm{T}}\mathbf{x} + d)$ defined on $\mathbf{dom}\, f = \left\{\mathbf{x} | \mathbf{c}^{\mathrm{T}}\mathbf{x} + d > 0\right\}$ is a *linear-fractional function* (Boyd and Vandenberghe 2004, p. 41).

Let $C \in \mathbf{dom}\, f$ be a convex set. Note that $S_\alpha(f, C) = \{\mathbf{x} \in C | f(\mathbf{x}) \geq \alpha\} = C \cap S_\alpha(f, \mathbf{dom}\, f)$. The following remark follows from the fact that the intersection of two convex sets is a convex set.

Remark 1. Let $C \in \mathbf{dom}\, f$ be a convex set. If f is quasiconcave on $\mathbf{dom}\, f$, then f is quasiconcave on C.

Property 1. Let f be a linear-fractional function and let $C \in \mathbf{dom}\, f$ be a convex set. Then image $D = \{f(\mathbf{x}) | \mathbf{x} \in C\}$ is a convex set. Furthermore, the inverse image of a convex set under a linear-fractional function is also convex, i.e., $\{f^{-1}(\mathbf{y}) | \mathbf{y} \in D\}$ is convex if D is convex (Boyd and Vandenberghe 2004, p. 42).

Note that $\Omega_1 = \left\{ \mathbf{q}^1 | \sum_{i=1}^{n} q_{i1} \le 1,\ q_{i1} \ge 0,\ g_{i1}(\mathbf{q}^1) \ge c_i\ \forall i \right\} = \mathbf{dom}\ \hat{R}_k$ is a convex set. Consider the following function F_1 that maps $\mathbf{q}^1 \in \Omega_1$ to $\mathbf{x} \in R^n$:

$$\mathbf{x} = F_1\left(\mathbf{q}^1\right) = \left(\frac{q_{11}/A_{11}}{1 - \sum_{l=1}^{n} q_{l1}}, ..., \frac{q_{n1}/A_{n1}}{1 - \sum_{l=1}^{n} q_{l1}} \right).$$

Function F_1 is a linear-fractional function (see Definition 2), and thus it follows from Property 1 that the image of Ω_1 under F_1, $\Omega_2 = \left\{ F_1\left(\mathbf{q}^1\right) | \mathbf{q}^1 \in \Omega_1 \right\}$, is a convex set.

Next consider the following function F_2 that maps $\mathbf{x} \in \Omega_2$ to $\mathbf{y} \in R^n$:

$$\mathbf{y} = F_2\left(\mathbf{x}\right) = \left(A_{12} x_1^{b_{12}/b_{11}}, ..., A_{n2} x_n^{b_{n2}/b_{n1}} \right).$$

The image of Ω_2 under F_2 is $\Omega_3 = \left\{ F_2\left(\mathbf{x}\right) | \mathbf{x} \in \Omega_2 \right\} = \left\{ A_{12} x_1^{b_{12}/b_{11}}, ..., A_{n2} x_n^{b_{n2}/b_{n1}} | \mathbf{x} \in \Omega_2 \right\}$. We next show that Ω_3 is a convex set. Let $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ denote two distinct points in Ω_2, so that $F_2\left(\mathbf{x}^{(1)}\right)$ and $F_2\left(\mathbf{x}^{(2)}\right)$ are two points in Ω_3. Note that Ω_3 is convex if and only if $\alpha F_2\left(\mathbf{x}^{(1)}\right) + (1-\alpha) F_2\left(\mathbf{x}^{(2)}\right) \in \Omega_3$ for all $\alpha \in [0,1]$ and all $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ in Ω_2, i.e., for any $\alpha \in [0,1]$ and any $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ in Ω_2, we require $\alpha F_2\left(\mathbf{x}^{(1)}\right) + (1-\alpha) F_2\left(\mathbf{x}^{(2)}\right) = F_2\left(\mathbf{x}^{(3)}\right)$ for some $\mathbf{x}^{(3)} \in \Omega_2$. We use a subscript on function F_2 to denote the functional element in vector $F_2(\mathbf{x})$, i.e., $F_{2i}\left(x_i\right) = A_{i2} x_i^{b_{i2}/b_{i1}}$. Thus,

$$\alpha F_{2i}\left(x_i^{(1)}\right) + (1-\alpha) F_{2i}\left(x_i^{(2)}\right) = \alpha A_{i2}\left(x_i^{(1)}\right)^{b_{i2}/b_{i1}} + (1-\alpha) A_{i2}\left(x_i^{(2)}\right)^{b_{i2}/b_{i1}}.$$

Assume without loss of generality that $x_i^{(1)} \le x_i^{(2)}$. Then, because $F_{2i}(x_i)$ is strictly increasing in x_i,

$$A_{i2}\left(x_i^{(1)}\right)^{b_{i2}/b_{i1}} \le \alpha A_{i2}\left(x_i^{(1)}\right)^{b_{i2}/b_{i1}} + (1-\alpha) A_{i2}\left(x_i^{(2)}\right)^{b_{i2}/b_{i1}} \le A_{i2}\left(x_i^{(2)}\right)^{b_{i2}/b_{i1}}$$

and there exists $x_i^{(3)} \in \left[x_i^{(1)}, x_i^{(2)}\right]$ such that $\alpha A_{i2}\left(x_i^{(1)}\right)^{b_{i2}/b_{i1}} + (1-\alpha) A_{i2}\left(x_i^{(2)}\right)^{b_{i2}/b_{i1}} = A_{i2}\left(x_i^{(3)}\right)^{b_{i2}/b_{i1}}$, and equivalently, there exists $\theta_i \in [0,1]$ that satisfies

$$\alpha A_{i2}\left(x_i^{(1)}\right)^{b_{i2}/b_{i1}} + (1-\alpha) A_{i2}\left(x_i^{(2)}\right)^{b_{i2}/b_{i1}} = A_{i2}\left(\theta_i x_i^{(1)} + (1-\theta_i) x_i^{(2)}\right)^{b_{i2}/b_{i1}} = A_{i2}\left(x_i^{(3)}\right)^{b_{i2}/b_{i1}}.$$

Of course, if $b_{i2}/b_{i1} = 1$, then $\theta_i = \alpha$. Combining the above, we have the following identity:

$$\alpha F_{2i}\left(x_i^{(1)}\right) + (1-\alpha) F_{2i}\left(x_i^{(2)}\right) = F_{2i}\left(\theta_i x_i^{(1)} + (1-\theta_i) x_i^{(2)}\right)$$

for some $\theta_i \in [0,1]$ and all i. Therefore, $\alpha F_2\left(\mathbf{x}^{(1)}\right) + (1-\alpha) F_2\left(\mathbf{x}^{(2)}\right) \in \Omega_3$ if and only if

$$\mathbf{x}^{(3)} := \left(\theta_1 x_1^{(1)} + (1-\theta_1) x_1^{(2)}, ..., \theta_n x_n^{(1)} + (1-\theta_n) x_n^{(2)}\right) \in \Omega_2. \tag{15}$$

To determine whether (15) holds, we need to characterize set Ω_2. Note that $\mathbf{x} = F_1\left(\mathbf{q}^1\right) = \left(\frac{q_{11}}{A_{11}q_{01}}, ..., \frac{q_{n1}}{A_{n1}q_{01}}\right)$. For pair $i,j \in \{1,...,n\}$ with $i \neq j$, let $\Delta = \sum_{l=1}^{n} q_{l1} - q_{i1} - q_{j1}$. We keep q_{01} and Δ fixed, and examine the (x_i, x_j) curve as q_{i1} varies over its feasible range $[0, 1-q_{01}-\Delta]$. Note that $q_{j1} = 1 - q_{01} - \Delta - q_{i1}$, and thus

$$x_j = \frac{1-q_{01}-\Delta-q_{i1}}{A_{j1}q_{01}} = \frac{1-q_{01}-\Delta}{A_{j1}q_{01}} - \frac{A_{i1}}{A_{j1}}\left(\frac{q_{i1}}{A_{i1}q_{01}}\right) = \frac{1-q_{01}-\Delta}{A_{j1}q_{01}} - \frac{A_{i1}}{A_{j1}}x_i$$

for $x_i \in \left[0, \frac{1-q_{01}-\Delta}{A_{i1}q_{01}}\right]$. It is apparent that the function $x_j(x_i)$ is a line with slope $-A_{i1}/A_{j1}$ connecting points $\left(0, \frac{1-q_{01}-\Delta}{A_{j1}q_{01}}\right)$ and $\left(\frac{1-q_{01}-\Delta}{A_{i1}q_{01}}, 0\right)$. By letting $\delta := q_{01} + \Delta$ vary over interval $[0,1]$ and q_{01} vary over interval $[0, \delta - \Delta]$, we see that our $x_j(x_i)$ curves cover the entire positive orthant in two dimensions. This holds for all $i,j \in \{1,...,n\}$ with $i \neq j$, and thus Ω_2 is the positive orthant in n dimensions. Therefore, (15) holds if $\mathbf{x}^{(3)}$ is in the positive orthant. This is clearly the case because $\theta_i \in [0,1]$ for all i and both $\mathbf{x}^{(1)} \in \Omega_2$ and $\mathbf{x}^{(2)} \in \Omega_2$ are in the positive orthant. By the above arguments, we have shown that $\alpha F_2\left(\mathbf{x}^{(1)}\right) + (1-\alpha) F_2\left(\mathbf{x}^{(2)}\right) \in \Omega_3$, and thus Ω_3 is a convex set.

Finally, consider the following function F_3 that maps $\mathbf{y} \in \Omega_3$ to $\mathbf{z} \in R_n$:

$$\mathbf{z} = F_3(\mathbf{y}) = \left(\frac{y_1}{1 + \sum_{j=1}^{n} y_j}, ..., \frac{y_n}{1 + \sum_{j=1}^{n} y_j}\right).$$

F_3 is a linear-fractional function (see Definition 2), and thus it follows from Property 1 that the image of Ω_3 under F_3, $\Omega_4 = \{F_3(\mathbf{y}) | \mathbf{y} \in \Omega_3\}$, is a convex set.

Now, to conclude that $\hat{R}_2$ is quasiconcave, we need to show that all of its superlevel sets are convex. Note that $\hat{R}_2\left(\mathbf{q}^1\right) = R_2\left(F_3\left(F_2\left(F_1\left(\mathbf{q}^1\right)\right)\right)\right) = R_2\left(f_2\left(\mathbf{q}^1\right)\right)$. We see that $\hat{R}_2\left(\mathbf{q}^1\right)$ is obtained by evaluating R_2 at a point in convex set Ω_4. Because the segment profit function $R_2\left(\mathbf{q}^2\right)$ is concave (and quasiconcave) on $\mathbf{dom}\ R_2 = \left\{q_{i2} | \sum_{i=1}^{n} q_{i2} \leq 1,\ q_{i2} \geq 0\ \forall i\right\}$ and $\Omega_4 \subset \mathbf{dom}\ R_2$ is convex set, we know from Remark 1 that R_2 is quasiconcave on Ω_4, and thus $S_\alpha(R_2, \Omega_4) = \{\mathbf{q}^2 \in \Omega_4 | R_2(\mathbf{q}^2) \geq \alpha\}$ is a convex set for any α (follows from Definition 1). To establish that

$$S_\alpha\left(\hat{R}_2, \Omega_1\right) = \left\{\mathbf{q}^1 | \mathbf{q}^1 \in \Omega_1, \hat{R}_2\left(\mathbf{q}^1\right) = R_2\left(F_3\left(F_2\left(F_1\left(\mathbf{q}^1\right)\right)\right)\right) \geq \alpha\right\}$$

is a convex set, we need to show that the inverse image of convex set $S_\alpha(R_2, \Omega_4)$ under $f_2 = F_3 \circ F_2 \circ F_1$ is convex (i.e., the inverse image of convex set $S_\alpha(R_2, \Omega_4)$ under f_2 is $S_\alpha\left(\hat{R}_2, \Omega_1\right)$). Now F_1 and F_3 are linear-fractional functions, and from Property 1, we know that an inverse image of a convex set under F_1 and F_3 is a convex set. What remains is to show that the inverse image of a convex set under F_2 is a convex set.

Let $D \in \Omega_3$ be a convex set. Its inverse image under F_2 is $C = \left\{F_2^{-1}(\mathbf{y})\,|\mathbf{y} \in D\right\}$. Recall that $F_2(\mathbf{x}) = \left(A_{12}x_1^{b_{12}/b_{11}}, ..., A_{n2}x_n^{b_{n2}/b_{n1}}\right)$, and thus

$$F_2^{-1}(\mathbf{y}) = \left(F_{21}^{-1}(y_1), ..., F_{2n}^{-1}(y_n)\right) = \left(\left(\frac{y_1}{A_{12}}\right)^{b_{11}/b_{12}}, ..., \left(\frac{y_n}{A_{n2}}\right)^{b_{n1}/b_{n2}}\right).$$

Suppose that $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are in D. Then $\mathbf{x}^{(1)} = F_2^{-1}\left(\mathbf{y}^{(1)}\right)$ and $\mathbf{x}^{(2)} = F_2^{-1}\left(\mathbf{y}^{(2)}\right)$. Inverse image C is convex if and only if $\alpha\mathbf{x}^{(1)} + (1 - \alpha\mathbf{x}^{(2)}) \in C$ for all $\alpha \in [0, 1]$ and for all $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ in C (i.e., $\mathbf{x}^{(1)} = F_2^{-1}\left(\mathbf{y}^{(1)}\right)$ and $\mathbf{x}^{(2)} = F_2^{-1}\left(\mathbf{y}^{(2)}\right)$ are in C because $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are in D). Because the elements of $\mathbf{y}$ are independent, if the above condition holds for the i^{th} element in $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, then it holds for all elements, i.e., we need to check if $\alpha x_i^{(1)} + (1 - \alpha)\, x_i^{(2)} \in C_i := \{x_i | \mathbf{x} \in C\}$. Note that

$$\alpha x_i^{(1)} + (1 - \alpha)\, x_i^{(2)} = \alpha \left(\frac{y_i^{(1)}}{A_{i2}}\right)^{b_{i1}/b_{i2}} + (1 - \alpha) \left(\frac{y_i^{(2)}}{A_{i2}}\right)^{b_{i1}/b_{i2}}$$

and that $\left(\frac{y_i}{A_{i2}}\right)^{b_{i1}/b_{i2}}$ is a strictly increasing function. Thus, there exists $\theta_i \in [0, 1]$ that satisfies

$$\alpha \left(\frac{y_i^{(1)}}{A_{i2}}\right)^{b_{i1}/b_{i2}} + (1 - \alpha) \left(\frac{y_i^{(2)}}{A_{i2}}\right)^{b_{i1}/b_{i2}} = \left(\frac{\theta_i y_i^{(1)} + (1 - \theta_i)\, y_i^{(2)}}{A_{i2}}\right)^{b_{i1}/b_{i2}} = F_{2i}^{-1}\left(\theta_i y_i^{(1)} + (1 - \theta_i)\, y_i^{(2)}\right).$$

Because D is convex, we know that $\theta_i y_i^{(1)} + (1 - \theta_i)\, y_i^{(2)} \in D_i := \{y_i | \mathbf{y} \in D\}$, which implies $\alpha x_i^{(1)} + (1 - \alpha)\, x_i^{(2)} \in C_i$. Therefore, C is a convex set.

Let us summarize the implications of the above. We now know that inverse image of a convex set under F_1, under F_2, and under F_3 is a convex set. Therefore, beginning with convex set $S_\alpha(R_2, \Omega_4)$, we obtain its convex inverse image under F_3. From this convex set, we obtain its convex inverse image under F_2, then repeat to obtain the convex inverse image under F_1. This process results in convex set $S_\alpha\left(\hat{R}_2, \Omega_1\right)$, which proves that $\hat{R}_2$ is quasiconcave on Ω_1. Therefore $\hat{R}_k$ is quasiconcave for any segment k. $\square$

A.9 Proof of Proposition 5

Proof. We first show that the sequence generated by Algorithm 2 has at least one limit point. From equation (3), the optimal price $p_i, i = 1, 2, \ldots, n$ must be bounded in the interval

$$\left[c_i + \frac{1}{\max_k b_{ik}},\ c_i + \frac{1}{\min_k b_{ik}} + \max_k \rho_k\right] \tag{16}$$

where ρ_k is the optimal profit from a segment k customer if prices of all products are set to maximize segment k profit only. Specifically, ρ_k solves the single-variable equation (Li and Huh 2011, Theorem 2)

$$\rho_k = \sum_{j=1}^{n} \frac{e^{a_{jk} - b_{jk} c_{jk} - 1} e^{-b_{jk} \rho_k}}{b_{jk}}$$

and is finite. Thus the optimal price $p_i, i = 1, 2, \ldots, n$ must be finite. Hence we assume that one always starts with a finite price vector in Algorithm 2.

Note that given any bounded margin vector at the t^{th} iteration,

$$\hat{p}_i^{t+1} = \hat{p}_i^t + \alpha^t d_i^t = \hat{p}_i^t + \alpha^t \left(\frac{1}{\sum\limits_{k=1}^{m} \left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}} + \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum\limits_{l=1}^{m} w_l b_{il} q_{il}} \right) r_k^t - \hat{p}_i^t \right)$$

where $r_k^t = \sum\limits_{i=1}^{n} \hat{p}_i^t q_{ik}\left(\hat{\mathbf{p}}^t\right)$. Since $\alpha^t \in [0, 1]$, $\hat{p}_i^{t+1}$ is bounded in the interval $\left[\min\left(\hat{p}_i^t, M^t\right), \max\left(\hat{p}_i^t, M^t\right)\right]$ where $M^t = \frac{1}{\sum_{k=1}^{m}\left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}} + \sum_{k=1}^{m}\left(\frac{w_k b_{ik} q_{ik}}{\sum_{l=1}^{m} w_l b_{il} q_{il}}\right) r_k^t$. M^t is the sum of the multiplicative inverse of a weighted average of b_{ik} values and a weighted average of the segment profits r_k^t. Since $r_k^t \leq \rho_k$ and ρ_k is finite, M^t is bounded by a finite constant $\frac{1}{\min_{ik} b_{ik}} + \rho_k$. As a result, $\hat{p}_i^{t+1} \leq \max\{\hat{p}_i^t, \frac{1}{\min_{ik} b_{ik}} + \rho_k\}$. Hence, the sequence $\{\hat{\mathbf{p}}^t\}$ is bounded and consequently has at least one limit point (see Bertsekas 2003, Proposition A.5, p. 666). Furthermore,

$$\nabla h\left(\hat{\mathbf{p}}^t\right)^{\mathrm{T}} \mathbf{d}^t = -\sum_{i=1}^{n} \sum_{k=1}^{m} w_k b_{ik} q_{ik} \left(\hat{p}_i^t - \sum_{k=1}^{m} \left(\frac{w_k b_{ik} q_{ik}}{\sum\limits_{l=1}^{m} w_l b_{il} q_{il}} \right) r_k^t - \frac{1}{\sum\limits_{k=1}^{m} \left(\frac{w_k q_{ik}}{q_i}\right) b_{ik}} \right)^2 < 0$$

unless $\hat{\mathbf{p}}^t$ is already a stationary point. Hence, $\{\mathbf{d}^t\}$ is gradient-related to $\{\hat{\mathbf{p}}^t\}$ and every limit point of the sequence $\{\hat{\mathbf{p}}^t\}$ is a stationary point of h (See Bertsekas 2003, Proposition 1.2.1, p. 43). $\square$

A.10 Proof of Proposition 6

Proof. The proof follows the same argument as in the proof of Proposition 5 and is omitted here. □

A.11 Data Fitting Details

In the following, we provide the details of data fitting and testing. Because not all product attributes are relevant for all customer segments, Intel suggested that segment-specific subset of regressors should be used to prevent problems stemming from over-fitting or over-simplifying. To that end, we test a variety of models for each segment where a model refers to a particular subset of the regressors.

We use the first three generations of products (13 SKUs) to parameterize the demand model and the fourth generation of products (3 SKUs) to test the model, mimicking the practical context at Intel. For any given customer segment, the market share prediction is computed for each product; we select the model using the mean absolute error (MAE) for the market share of each product.

Table 6 presents a summary of goodness-of-fit and test measures for the selected model for each segment including the Estrella index (which is a value between 0 and 1, larger number corresponding to a better fit), the training MAE, and the test MAE. The model for each segment is chosen by focusing primarily on the test MAE and secondly on the training MAE, and by balancing model parsimony and the test errors.

Table 6: Fit and Forecast Accuracy.

Segment	Chosen Model	Estrella Index	Training MAE	Test MAE
1	TDP, Performance, Price/performance	89%	10%	17%
2	Frequency, TDP, Price, Price/performance	62%	13%	1%
3	Performance, Price/performance	78%	12%	2%
4	Performance, Price/performance	53%	13%	9%
5	Frequency, Price, Price/performance	71%	10%	6%
6	Performance, Price/performance	50%	14%	10%
7	Performance, Price/performance	54%	13%	13%

The coefficients and the corresponding standard errors (in parenthesis) of the selected regression model for each segment are given in Table 7.

Table 7: Linear Utility Coefficients for Each Customer Segment.

Segment	Frequency	TDP	Performance	Price	Price/performance
1	–	-0.2791 (0.0095)	0.02885 (0.00066)	–	-0.786 (0.170)
2	2.097 (0.162)	-0.0244 (0.0086)	–	0.00105 (0.00051)	-3.677 (0.131)
3	–	–	0.00936 (0.00031)	–	-0.993 (0.106)
4	–	–	0.00267 (0.00027)	–	-2.201 (0.099)
5	2.512 (0.135)	–	–	0.00490 (0.00032)	-2.846 (0.109)
6	–	–	0.00729 (0.00030)	–	-0.615 (0.084)
7	–	–	0.00777 (0.00030)	–	-0.625 (0.087)

A.12 Segment-specific Sales Distribution among Products

Table 8: Sales Distribution under Current Practice (Each Number Represents Segment-specific Choice Probability for the Corresponding Product).

Product \ Segment	S1	S2	S3	S4	S5	S6	S7
1	0.0008	0.0982	0.0464	0.1505	0.0451	0.0726	0.0661
2	0.0044	0.3359	0.0764	0.1457	0.3169	0.1087	0.1018
3	0.9897	0.3938	0.5274	0.4003	0.3954	0.4716	0.4829

Table 9: Sales Distribution under Profit-improving Solution (Each Number Represents Segment-specific Choice Probability for the Corresponding Product).

Product \ Segment	S1	S2	S3	S4	S5	S6	S7
1	0.0017	0.2063	0.0862	0.3339	0.0848	0.1043	0.0962
2	0.0097	0.6924	0.1425	0.3256	0.5198	0.1565	0.1486
3	0.9807	0.0282	0.3590	0.0957	0.2065	0.3634	0.3736