Assortment and Price Optimization under an Endogenous Context-Dependent Multinomial Logit Model

Yicheng Bai[1], Omar El Housni[1], Paat Rusmevichientong[2], Huseyin Topaloglu[1]

[1]School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10044
[2]Marshall School of Business, University of Southern California, Los Angeles, CA 90089
{yb279,oe46}@cornell.edu, rusmevic@marshall.usc.edu, topaloglu@orie.cornell.edu

August 8, 2023

Motivated by empirical evidence that the utility of each product depends on the assortment of products offered along with it, we propose an endogenous context-dependent multinomial logit model (Context-MNL) under which the utility of each product depends on both the product's intrinsic value and the deviation of the intrinsic value from the expected maximum utility among all the products in the offered assortment. Under the Context-MNL model, an assortment provides a context in which customers evaluate the utility of each product. Our model generalizes the standard multinomial logit model and allows the utility of each product to depend on the offered assortment. The model is parsimonious, requires only one parameter more than the standard multinomial logit model, captures the assortment-dependent effect endogenously, and does not require the decision-maker to determine in advance the relevant attributes of the assortment that might affect the product utility. The Context-MNL model also admits tractable maximum likelihood estimation and is operationally tractable, with efficient solution methods for solving assortment and price optimization problems. Our numerical study, which is based on data from Expedia, shows that compared to the standard multinomial logit model, the Context-MNL model substantially improves out-of-sample goodness of fit and prediction accuracy.

1. Introduction

Discrete choice models describe how customers choose a product from an assortment of alternatives presented to them. Choice models have been widely used in revenue management problems, such as finding good assortments of products to offer to customers or determining product prices that maximize total revenue. Most choice models used in the revenue management literature assume that each customer assigns a utility to each product consisting of the product's intrinsic value plus an idiosyncratic noise. When a customer is offered an assortment of products, she chooses the product from the assortment that maximizers her utility. Under this assumption, the probability of selecting a particular product changes with an assortment, but the utility of each product is independent of the assortment offered to each customer.

The marketing literature has provided ample evidence that a customer constructs a product's utility in real time based on the specific choice context, which corresponds to the assortment of alternatives that are shown alongside the target product. Researchers have documented numerous effects of offered assortments on customer preferences and choice behaviors; these effects are consistent with the theory that each product's utility depends partly on the assortment within which the product is offered. The prospect theory in economics also points to the effect of context, and posits that the utility of a product depends on the product's intrinsic value and the deviation of the intrinsic value from some reference point. The deviation from the reference point can correspond to the gain or loss a customer experiences when comparing a product to its alternatives in the assortment.

Despite considerable evidence of assortment-dependent effects on each product's utility, only a few revenue management papers have examined choice models with assortment-dependent utility specification. The existing literature has assumed that the impact of assortment on the utility of each product is modeled as a function of prespecified assortment-specific attributes, such as the cheapest price in the assortment, the assortment size, or the minimum quality among products in the assortment. Relying on such prespecified attributes requires that the utility of a product is formulated as a function of specific attributes, so the assortment and price optimization algorithms developed under these choice models become ineffective when the utility of a product is formulated in a different way. We formulate the utility of each product as a function of the utilities of all other products in the assortment, so our approach is fully independent of how the utilities of the products are formulated as a function of the product features.

1.1 Main Contributions

Our main contributions are the formulation of an endogenous context-dependent multinomial logit model, the validation of the model on data, and the development of efficient methods for parameter estimation, assortment and price optimization under this choice model.

Endogenous Context-Dependent Multinomial Logit Model: We propose an *endogenous* context-dependent multinomial logit model, which we refer to as the Context-MNL model. Our model does not require the decision-maker to determine in advance the relevant features of an assortment. Rather, we define the utility of a product in an offered assortment as the sum of three terms: (i) the product's intrinsic value, (ii) the deviation of the intrinsic value from the expected maximum utility among all the items in the offered assortment, and (iii) an idiosyncratic noise with a Gumbel distribution. Under our Context-MNL model, the utility of each product in an assortment depends partly on the expected maximum utility among all the products in the offered assortment. The expected maximum utility serves as an endogenous reference point. By defining each product's utility as a function of the utility of all the products in the assortment, we endogenously create a system of nonlinear equations whose unique solution defines the mean utility of each product (Equation (2) and Theorem 2.2). Our model is parsimonious. Compared to the standard multinomial logit model, the Context-MNL model has just one extra parameter α, which measures the impact of the deviation of each product's intrinsic value from the expected maximum utility in the offered assortment and which ranges from negative one to infinity. When α is zero, the Context-MNL model reduces to the standard multinomial logit model. Under the Context-MNL model, products are substitutable (Corollary 2.4), and the selection probability can be computed efficiently (Theorems 2.2 and 2.3).

Flexibility in Modeling Demand Spills and Recaptures: In a situation in which a customer's first-choice demand is unavailable, spills occur when the first-choice demand is redirected to the no-purchase option, and recaptures occur when the first-choice demand is redirected to other products. It is well-known that the standard multinomial logit model is overly optimistic about recaptures (Gallego et al. 2015). The extra parameter α in our Context-MNL model provides flexibility in modeling spills and recaptures. According to Theorem 2.5, the recapture-to-spill ratio is strictly increasing in α, so when α is less than zero, the recaptures under the Context-MNL model will be less optimistic than those under the standard multinomial logit model.

Improved Goodness of Fit and Prediction Accuracy: We validate our Context-MNL model using data from Expedia, which provides the results of search queries for hotels on Expedia for each customer and whether the customer chooses to book a hotel displayed in the queries. Our numerical experiments show that by incorporating the context-dependent effect on product utility, the Context-MNL model significantly improves the goodness of fit and prediction accuracy regarding customer choice behavior. We generate multiple datasets based on the Expedia data by varying no-purchase probabilities and the number of queries. In every dataset we generate, our experiment shows that the expected maximum utility among the products in an offered assortment serves as an important endogenous reference point and plays an important role in explaining customer choice behavior in the data. The estimated parameter α is statistically significant, with p-values less than 10^{-5} in all the datasets. The fitted Context-MNL model also improves the out-of-sample log-likelihood over that of the standard multinomial logit, with an average improvement of 2.30%. Compared to the standard multinomial logit model, our model also improves the ability to predict the bookings made by customers, with an average improvement in prediction accuracy of 1.31%.

Efficient Assortment and Price Optimization: Revenue management problems can also be solved efficiently under the Context-MNL model. We consider an assortment optimization problem under a cardinality constraint. When α is nonnegative, letting n denote the number of problems, the assortment optimization problem can be solved by evaluating the revenue of at most n^2 assortments (Corollary 3.2). For α strictly between negative one and zero, the assortment optimization problem is NP-hard, and we develop a fully polynomial-time approximation scheme (Theorem 3.5). In numerical experiments, we validate the performance of our approximation scheme by choosing the parameter to ensure that the resulting assortment has an expected revenue within 50% of the optimal expected revenue. We observe that the approximation scheme yields solutions whose expected revenues far exceed the theoretical performance guarantee, and the resulting expected revenues are on average within 0.33% of the optimal expected revenue, with a maximum gap of 2.30%. Under the Context-MNL model, we can also efficiently compute prices that maximize the total expected profit. As Theorem 4.1 shows, the optimal prices exhibit the constant-markup property, in which every product is charged the same markup above its marginal cost. We also provide an explicit formula for the optimal markup that generalizes the formula in the standard multinomial logit setting.

Tractable Estimation: Under the Context-MNL model, for each value of α, the negative log-likelihood function is convex in the remaining parameters (Theorem 5.1). Because α is a real number, we can create a grid of possible values of α, and for each value, we estimate the remaining parameters by minimizing a convex function. We pick the value of α that gives the highest log-likelihood value. Because α is a one-dimensional parameter, searching over α can be done relatively efficiently, and we can increase the quality of the solution by using a finer grid of values. Thus, maximum likelihood estimation is tractable for our choice model.

1.2 Literature Review

Our work is related to the literature in marketing, economics, and revenue management. We describe the connection to each stream of literature below. The marketing literature has provided ample evidence that the choice behavior of customers is driven partly by the offered assortment, which provides a context on the basis of which customers construct the product utilities (Payne 1982, Chakravarti and Lynch 1983, Ratneshwar et al. 1987, Prelec et al. 1997). These assortment-dependent effects support the view that customers construct the utility of each product in real time based on the particular assortment of items offered along with the target product (Payne et al. 1992, Bettman et al. 1998, Simonson 1999). Researchers have proposed a number of models that incorporate the assortment-dependent utility. Kamakura and Srivastava (1984) propose a multinomial probit model with distance-dependent covariates to measure the similarity between a product and its alternatives in the assortment. Tversky and Simonson (1993) develop a componential context model that defines the value of a product as the sum of its intrinsic value and the advantage (or disadvantage) that the product has over its alternatives in the assortment; the advantage is computed by comparing each of the prespecified attributes of the target product with the corresponding attributes of each alternative in the assortment. Kivetz et al. (2004)

generalize the multinomial logit model and incorporate the assortment effect by computing the distance between the product's attributes and an anchor that consists of the minimum value of each attribute among all the alternatives in the assortment. Rooderkerk et al. (2011) consider a similar model but allow Gaussian idiosyncratic noise. A common theme across these models is the need to predetermine the assortment-specific attributes that will contribute to the product utility. The number of additional parameters in these models often increases with the number of attributes chosen. Moreover, most of the marketing literature has focused on parameter estimation and does not consider the tractability of the operational decisions. An exception is Orhun (2009), which uses comparative valuation model to examine the problem of designing an optimal product line when consumers exhibit assortment-dependent preferences, but the proposed model does not admit efficient solutions for large-scale assortment and price optimization.

Our study is also related to the prospect theory in economics, pioneered by Kahneman and Tversky (1979). Under the prospect theory, each customer evaluates a product against a reference point, which might correspond to the cheapest product in the assortment, and the product utility will depend partly on the deviation of the target product from the reference point. Kőszegi and Rabin (2006) propose an additive form of the reference-dependent preference, under which the utility of each product is the sum of the product's intrinsic value and the gain–loss utility function that measures the distance between the target product and the reference point. The reference-dependent model of preferences has received increasing attention in the revenue management literature. Researchers have investigated the dynamic pricing problem under reference effects (Popescu and Wu 2007, Hu et al. 2016, Chen et al. 2017). In a recent paper, Cao et al. (2022) propose a spiked multinomial logit model under which a product receives a boost in the assigned utility if the product is the cheapest item in the assortment, and they consider various revenue management problems under this model. Wang (2018) extends the standard multinomial logit model by allowing each product's utility to depend on the deviation of the product's price from the minimum price in the assortment, and he examines assortment and price optimization under this model. These models assume that the reference attributes are prespecified, so they are tied to a particular specification of the product utilities.

Our study is also related to the substantial research on assortment optimization under the standard multinomial logit model. We show that for the unconstrained assortment optimization problem under the Context-MNL model with a nonnegative α parameter, revenue-ordered assortments are optimal. This result generalizes the classical result, which establishes that, under the standard multinomial logit model, it is optimal to offer a revenue-ordered assortment, which consists of products with the highest revenues (Gallego et al. 2004, Talluri and van Ryzin 2004).

Rusmevichientong et al. (2010), Wang (2012), Jagabathula (2016), and Sumida et al. (2021) examine the problem under various constraints on the offered assortment under the standard multinomial logit model. Bront et al. (2009), Mendez-Diaz et al. (2014), Rusmevichientong et al. (2014), and El Housni and Topaloglu (2023) consider the problem under a mixture of multinomial logit models.

Regarding pricing under the standard multinomial logit model, Song and Xue (2007) show that the expected revenue is concave in the product market shares. Hopp and Xu (2005) and Li and Huh (2011) show that the optimal prices possess the constant mark-up property, according to which the optimal price of each product exceeds its marginal cost by the same constant. Zhang et al. (2018) show that both of these results hold under all generalized extreme value models. Like our study, all of these papers assume that the products have the same price sensitivity. Our review is limited to studies based on the multinomial logit model, but we refer the reader to Farias et al. (2013), Davis et al. (2014), Gallego and Wang (2014), Aouad et al. (2021), Blanchet et al. (2016), Désir et al. (2016), and Li and Webster (2017) for representative work based on other choice models.

The concavity of the log-likelihood function under the standard multinomial logit model is well-known (Train 2003). Wang and Wang (2017) extend the standard multinomial logit model to capture endogenous network effects, under which the market share of each product is implicitly defined as a solution to a system of nonlinear equations. As a result of the implicitly defined market shares, the log-likeliood function of their model is no longer concave, and they use a heuristic to find a stationary point of the log-likelihood function. In contrast, although the utility in our Context-MNL model is also implicitly defined, for each value of α, the log-likelihood function under the Context-MNL model remains concave in all the other parameters.

Organization: In Section 2, we formulate our Context-MNL model, provide motivating examples, establish properties of the model, and show how our model provides flexibility in modeling of demand spills and recaptures. In Section 3, we develop efficient methods for solving the assortment optimization problem under our model. Then, in Section 4, we consider the price optimization problem and give a formula for the optimal markup. In Section 5, we show that for each value of α, the negative log-likelihood function under the Context-MNL model is convex in all the remaining parameters. Section 6 contains numerical experiments and we conclude in Section 7.

2. Model

We have a universe of n products indexed by $\mathcal{N} = \{1, 2, \ldots, n\}$ and we let 0 denote the no-purchase option. Under the standard multinomial logit model, each customer assigns each product in $\mathcal{N}$ a random utility, consisting of a product-specific deterministic term plus a product-specific random noise. As noted in Section 1, studies have documented many settings in which the utility assigned by a customer to each product depends on the assortment of products offered alongside it. The assortment provides a context on the basis of which customers evaluate the utility of each product. For example, if the assortment contains similar products with favorable reviews by past buyers, this may increase the customer's assessment of the utility of the target product. On the other hand, if the population of past buyers have negative opinions of related products, this may decrease the customer's assessment of the target product's value.

2.1 Endogenous Context-Dependent Multinomial Logit

Our *endogenous context-dependent multinomial logit* (Context-MNL) model captures the assortment-dependent effect by assuming that the random utility associated with each product depends on the assortment of products containing it, as follows: for each assortment $S \subseteq \mathcal{N}$, the random utility for each product in S is given by

$$\mathsf{Util}_i(S) \;=\; \mu_i - \alpha \mathbb{E}\Big[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\Big] + \epsilon_i \quad \text{for all} \;\; i \in S \qquad \text{and} \qquad \mathsf{Util}_0 \;=\; \mu_0 + \epsilon_0, \tag{1}$$

where $\{\epsilon_i : i \in \mathcal{N} \cup \{0\}\}$ are idiosyncratic noises and assumed to be independent and identically distributed mean-zero Gumbel random variables with a scaling parameter of one.[1] The parameter $\alpha \in \mathbb{R}$ is the assortment-effect parameter, which captures the impact of the context on the utility of each product.

We assume each customer follows the random utility maximization principle and chooses from an assortment the product that maximizes her utility. Customers are also assumed to be independent. Thus, the term $\mathbb{E}\big[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\big]$ captures the population average of the maximum utility among all the products shown in the assortment. By the random utility maximization principle, given an assortment $S \subseteq \mathcal{N}$, the probability that a customer chooses product $i \in S$, denoted by $\phi_i(S)$, is the probability that product i has the highest utility among the offered products; that is, $\phi_i(S) = \Pr\{\mathsf{Util}_i(S) > \mathsf{Util}_\ell(S) \;\; \forall\, \ell \in S \cup \{0\}, \ell \neq i\}$. We refer to $\phi_i(S)$ as the choice probability or the selection probability, and we use these two terms interchangeably.

The model in Equation (1) captures the idea that the random utility of each product i in an assortment S depends on the sum of three terms: (i) the intrinsic value of product i, (ii) the

[1] Recall that a random variable X follows a Gumbel distribution with a location parameter μ and a scaling parameter β if for all $x \in \mathbb{R}$, $\Pr\{X \leq x\} = e^{-e^{-(x-\mu)/\beta}}$. We denote this by $X \sim \mathsf{Gumbel}(\mu, \beta)$. It is well-known that $\mathbb{E}[X] = \mu + \beta\gamma_{\mathsf{EM}}$ where $\gamma_{\mathsf{EM}} = 0.57721...$ is the Euler–Mascheroni constant. Because ϵ_i has a mean of zero, it follows that $\epsilon_i \sim \mathsf{Gumbel}(-\gamma_{\mathsf{EM}}, 1)$ for all $i \in \mathcal{N} \cup \{0\}$. The distribution of ϵ_i is the same as the distribution of the noise in the utility specification in the standard multinomial logit model.

deviation of the intrinsic value from the expected maximum utility among all the products in the assortment, and (iii) idiosyncratic noise. To see this, letting η_i denote the intrinsic value of product i, then the sum of the three terms correspond to $\eta_i + \beta\big(\mathbb{E}[\max_{\ell \in S\cup\{0\}} \mathsf{Util}_\ell(S)] - \eta_i\big) + \epsilon_i$, which is equivalent to $(1-\beta)\eta_i + \beta\mathbb{E}[\max_{\ell \in S\cup\{0\}} \mathsf{Util}_\ell(S)] + \epsilon_i$, and this is an instance of the Context-MNL model with $\mu_i = (1-\beta)\eta_i$ and $\alpha = -\beta$.

Under Equation (1), the utility of each product now depends on the accompanying assortment. As noted in the introduction, other context-dependent utility models assume that the utility of each product depends on *exogenous* assortment-specific characteristics, such as the minimum price of the products in the assortment, the assortment size, or other prespecified exogenous characteristics. In many applications, however, it is unclear what these exogenous characteristics should be. In contrast, our Context-MNL model incorporates the impact of the accompanying assortment *endogenously*. Through Equation (1), the utility of each product depends on the utility of all the products in the assortment. As we will demonstrate, this definition endogenously specifies the mean utility of each product as the unique solution to a system of nonlinear equations.

The standard multinomial logit model is a special case of the Context-MNL where $\alpha = 0$. The Context-MNL model thus generalizes the standard multinomial logit model by introducing an additional parameter α, which captures the effect of context on product utility. As the remainder of the paper demonstrates, the Context-MNL model enjoys many advantages of the standard multinomial logit model, such as ease of estimation, operational tractability, and structural insights. More importantly, as our numerical experiments show, with the addition of a single parameter α, the Context-MNL model provides a significantly better fit with real-world data and better prediction accuracy than the standard multinomial logit model.

To ensure that the utility random variable is well-defined, we need to restrict the range of α. The following example shows that if $\alpha \le -1$, then the utility random variable can have an infinite mean. We will later prove that if $\alpha > -1$, then the utility has a finite mean and is well-defined.

Example 2.1 (Infinite Mean) Suppose $\alpha \le -1$. For ease of notation, let $\beta = -\alpha$; note that $\beta \ge 1$. Letting $S = \{1\}$, $\mu_0 = 0$, $\mu_1 = 1$, and $X = \mathsf{Util}_1(S)$, it follows from Equation (1) that $X = 1 + \beta\mathbb{E}[\max\{\epsilon_0, X\}] + \epsilon_1$, and because $\mathbb{E}[\epsilon_1] = 0$, we have $\mathbb{E}[X] = 1 + \beta\mathbb{E}[\max\{\epsilon_0, X\}]$. Because of Jensen's inequality, $\mathbb{E}[\epsilon_0] = 0$, and $\beta \ge 1$, we have $\mathbb{E}[X] \ge 1 + \beta\max\{0, \mathbb{E}[X]\}\}$. This implies that $\mathbb{E}[X] \ge 1$, so we have $\mathbb{E}[X] \ge 1 + \beta\mathbb{E}[X]$. Repeated application of this equality shows that for all $k \in \mathbb{Z}_+$, $\mathbb{E}[X] \ge \sum_{\ell=0}^{k-1} \beta^\ell + \beta^k\mathbb{E}[X] \ge \sum_{\ell=0}^{k} \beta^\ell$, where the last inequality follows because $\mathbb{E}[X] \ge 1$. Because $\beta \ge 1$, this shows that $\mathbb{E}[X]$ is infinite.

Motivating Examples for Different Ranges of α: Example 2.1 motivates us to restrict the range of α to be strictly higher than -1. Building on the intuition given in the beginning of this section, we now describe two motivating examples for our $\mathsf{Context\text{-}MNL}$ model. The first example focuses on the case in which $\alpha \geq 0$ and the second example on the case in which $\alpha \in (-1, 0]$.

An Example for $\alpha \geq 0$: Consider a platform that sells a single product from multiple sellers. When a customer arrives to purchase the product, the platform's goal is to decide on the assortment of sellers to show to the customers. Although all the sellers offer the same product, the quality of the products may differ slightly from seller to seller; some sellers may offer a brand new product, whereas others may offer a slightly used one. The service provided may also differ from one seller to the next; for example, some sellers may offer one-day shipping via UPS, whereas others may require at least one week to deliver the product.

In this setting, $\mathcal{N} = \{1, 2, \ldots, n\}$ corresponds to the set of sellers. We assume that customers are homogenous but independent. Each arriving customer assigns a value $\tilde{\mu}_i$ to seller i, which can depend on the various characteristics of the product and the service offered by seller i.

According to the prospect theory (Kahneman and Tversky 1979), it is well-known that consumers exhibit an anchoring effect; they would thus compare the reputations of the sellers. For example, if seller i's reputation is worse than the reputation of other sellers shown in the assortment, this information would negatively affect the utility the customer assigns to seller i. In other words, the customer compares the utility of seller i to a reference point. If $\tilde{\mu}_i$ is lower than this reference point, the overall utility the customer assigns to seller i decreases, but if $\tilde{\mu}_i$ is higher than the reference point, the overall utility the customer assigns to seller i increases.

Because each customer independently chooses the seller that gives her the highest utility, the population-level expected maximum utility that the population of customers derives from the sellers in assortment S is $\mathbb{E}\left[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\right]$. Our $\mathsf{Context\text{-}MNL}$ model uses the population-level expected maximum utility as a reference point, so the overall utility the customer assigns to seller $i \in S$ is given by: for some $\beta \geq 0$,

$$\mathsf{Util}_i(S) \equiv \tilde{\mu}_i - \beta\Big(\mathbb{E}\Big[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\Big] - \tilde{\mu}_i\Big) + \epsilon_i = (1+\beta)\tilde{\mu}_i - \beta\mathbb{E}\Big[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\Big] + \epsilon_i ,$$

which is an instance of the endogenous context-dependent utility model where $\alpha = \beta$ and $\mu_i = (1+\beta)\tilde{\mu}_i$.

An Example for $\alpha \in (-1, 0]$: We now describe a setting in which $\alpha \in (-1, 0]$ may be applicable. Consider a firm selling n substitutable products indexed by $\mathcal{N} = \{1, 2, \ldots, n\}$. We assume the customers are independent and homogenous. Each customer initially assigns a value $\hat{\mu}_i$ to product i that depends on various characteristics of product i, which represent the product's intrinsic value. However, the utility the customer ultimately assigns to product i may also depend on the population average of the utility that other customers obtain from purchasing products in the assortment S. This is because products are substitutable and are offered by a single firm, so if the other customers are unhappy with the products offered in the assortment, then this may indicate inferior product quality or subpar support services offered by the firm. Customers choose products that maximize their utility, and we capture the maximum expected utility with the term $\mathbb{E}\big[\max_{\ell \in S\cup\{0\}} \mathsf{Util}_\ell(S)\big]$, which reflects the population-level expected utility other customers obtain from purchasing the products in S. We model the customer's final utility as a convex combination of $\hat{\mu}_i$ and $\mathbb{E}\left[\max_{\ell \in S\cup\{0\}} \mathsf{Util}_\ell(S)\right]$. So, if the population-level expected utility is lower than the customer's initial assessment, the customer's final utility will be lower than her initial value. Thus, for some $\gamma \in (0,1)$, the overall utility a customer assigns to product $i \in S$ is given by $\mathsf{Util}_i(S) \equiv (1-\gamma)\hat{\mu}_i + \gamma\mathbb{E}\big[\max_{\ell \in S\cup\{0\}} \mathsf{Util}_\ell(S)\big] + \epsilon_i$, which is an instance of the endogenous context-dependent utility model where $\alpha = -\gamma \in (-1,0)$ and $\mu_i = (1-\gamma)\hat{\mu}_i$.

Preference Weights and Choice Probabilities: As in the standard multinomial logit model, we can assume without loss of generality that the parameter μ_0 for the no-purchase option is set to zero; Appendix A provides a detailed account of how to transform a $\mathsf{Context\text{-}MNL}$ model with nonzero μ_0 to an equivalent model with $\mu_0 = 0$. The parameters of the $\mathsf{Context\text{-}MNL}$ model thus consist of $\alpha > -1$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$.

Determining the utility random variables directly from Equation (1) may not be straightforward. Theorem 2.2 below provides an equivalent characterization of the utility random variables under the $\mathsf{Context\text{-}MNL}$ model and gives an explicit formula for the selection probability. The theorem also shows that we can express the selection probability in terms of the assortment-dependent preference weight. The key difference in our problem is that the preference weight of each product now depends on the assortment of other products offered alongside it. Given the parameters $\alpha > -1$ and $\boldsymbol{\mu} \in \mathbb{R}^n$ of the $\mathsf{Context\text{-}MNL}$ model, we introduce the system of equations that defines the "preference weight" associated with each product within each assortment. For each $S \subseteq \mathcal{N}$, letting $(v_\ell(S) : \ell \in S)$ be the set of variables, consider the following system of nonlinear equations that define the *preference weights for products in the assortment* S:

$$v_i(S) \;=\; e^{\mu_i - \alpha \log\left[1 \;+\; \sum_{\ell \in S} v_\ell(S)\right]} \qquad \forall\; i \in S\;. \tag{2}$$

The following theorem shows that the above system of equations has a unique solution and that the solution can be used to describe the choice probability. As Theorem 2.2 shows, the formula for the selection probability under the $\mathsf{Context\text{-}MNL}$ model is the same as that under the standard multinomial logit model; the only difference is that the preference weight $(v_\ell(S) : \ell \in S)$ of each product now depends on the assortment containing it. Noting that $\phi_i(S) = 0$ for all $i \notin S$, and $\phi_0(S) \equiv 1 - \sum_{i \in S} \phi_i(S)$ is the probability that a customer does not select any product in S.

Theorem 2.2 (Equivalent Utility Specification via Preference Weights) *Fix arbitrary $\alpha > -1$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$. For each $S \subseteq \mathcal{N}$, the system of equations* (2) *has a unique solution. Moreover, for all $S \subseteq \mathcal{N}$, the utility random variable $\mathsf{Util}_i(S)$ in Equation* (1) *can be equivalently expressed as follows:*

$$\mathsf{Util}_i(S) \;=\; \log v_i(S) \;+\; \epsilon_i \quad \forall\, i \in S \qquad \textit{and} \qquad \mathsf{Util}_0(S) \;=\; \epsilon_0 \;,$$

and for all $i \in S$, the probability $\phi_i(S)$ that product i is selected when assortment S is offered, is given by $\phi_i(S) = v_i(S) / \left(1 + \sum_{\ell \in S} v_\ell(S)\right)$.

Proof: Fix an arbitrary $\boldsymbol{\mu} \in \mathbb{R}^n$, $\alpha > -1$, and $S \subseteq \mathcal{N}$. Letting $w = \sum_{i \in S} v_i(S)$ and adding Equation (2) over all $i \in S$, it follows that $w = \frac{\sum_{i \in S} e^{\mu_i}}{(1\,+\,w)^\alpha}$ or, equivalently, $w\,(1 \,+\, w)^\alpha = \sum_{i \in S} e^{\mu_i}$. Because $\alpha > -1$, by taking the derivative, it is easy to verify that the mapping $x \in \mathbb{R}_+ \to x\,(1+x)^\alpha$ is strictly increasing, unbounded, and has a value of zero when $x = 0$. Thus, there exists a unique w^* such that $w^*\,(1 \,+\, w^*)^\alpha = \sum_{i \in S} e^{\mu_i}$. It follows that $v_i(S) = e^{\mu_i - \alpha \log(1+w^*)}$ for all $i \in S$, so $(v_i(S) : i \in S)$ is uniquely defined.

Because the noises are i.i.d. mean-zero Gumbel random variables with a scaling parameter of one, it follows from Equation (1) that $\{\mathsf{Util}_i(S) : i \in S \cup \{0\}\}$ are also independent Gumbel random variables with a scaling parameter of one. Letting $x_i(S) = \mathbb{E}[\mathsf{Util}_i(S)]$ for all $i \in S \cup \{0\}$ and noting that $x_0(S) = 0$, it follows that $\mathsf{Util}_i(S) \sim \mathsf{Gumbel}(x_i(S) - \gamma_{\mathsf{EM}}, 1)$, where $\gamma_{\mathsf{EM}} = 0.57721...$ is the Euler–Mascheroni constant. Because $\{\mathsf{Util}_i(S) : i \in S \cup \{0\}\}$ are independent Gumbel random variables with a scaling parameter of one, $\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)$ is still a Gumbel random variable with a scaling parameter of one and a location parameter $e^{-\gamma_{\mathsf{EM}}} + \sum_{\ell \in S} e^{x_\ell(S) - \gamma_{\mathsf{EM}}}$ (Gumbel 2004), which implies that $\mathbb{E}\big[\max_{\ell \in S \cup \{0\}} \mathsf{Util}_\ell(S)\big] = \gamma_{\mathsf{EM}} + \log\left(e^{-\gamma_{\mathsf{EM}}} + \sum_{\ell \in S} e^{x_\ell(S) - \gamma_{\mathsf{EM}}}\right)$. Because $\mathbb{E}[\epsilon_i] = 0$ for all i, it follows from Equation (1) that

$$x_i(S) = \mu_i - \alpha\Big(\gamma_{\mathsf{EM}} + \log\Big(e^{-\gamma_{\mathsf{EM}}} + \sum_{\ell \in S} e^{x_\ell(S) - \gamma_{\mathsf{EM}}}\Big)\Big) = \mu_i - \alpha \log\Big(1 + \sum_{\ell \in S} e^{x_\ell(S)}\Big) \;,$$

and thus $(e^{x_i(S)} : i \in S)$ satisfies the system of equations (2), which has a unique solution by the proof we present in the previous paragraph. Therefore, $x_i(S) = \log v_i(S)$, which implies that $\mathsf{Util}_i(S) = \log v_i(S) + \epsilon_i$ for all $i \in S$.

To derive the choice probability, we use the fact that $\mathsf{Util}_i(S) \sim \mathsf{Gumbel}(x_i(S) - \gamma_{\mathsf{EM}}, 1)$ for all $i \in S \cup \{0\}$, and the utility random variables are independent. It follows from the property of the Gumbel random variables that for $i \in S$, $\Pr\left\{\mathsf{Util}_i(S) > \max_{\ell \in S\cup\{0\}, \ell \neq i} \mathsf{Util}_\ell(S)\right\} = \frac{e^{x_i(S)}}{\sum_{\ell \in S \cup \{0\}} e^{x_\ell(S)}} = \frac{v_i(S)}{1+\sum_{\ell \in S} v_\ell(S)}$, where the last equality follows because $x_i(S) = \log v_i(S)$ and $x_0(S) = 0$. ■

2.2 Properties of the Context-MNL Model

The expression for the choice probability in Theorem 2.2 requires solving a system of nonlinear equations. We can give an equivalent expression in terms of the primitive parameters α and $\boldsymbol{\mu}$, and this expression requires only the evaluation of a one-dimensional function. For each $\alpha > -1$, let $f : [1, \infty) \to \mathbb{R}_+$ and $g : \mathbb{R}_+ \to [1, \infty)$ be defined by

$$f(x) = x - x^{\alpha/(1+\alpha)} \quad \forall\, x \in [1, \infty) \qquad \text{and} \qquad g(y) = f^{-1}(y) \quad \forall\, y \in \mathbb{R}_+ \,. \tag{3}$$

Because $\alpha > -1$, by taking the derivative, it is easy to verify that f is strictly increasing over the interval $[1, \infty)$ and that its range is the nonnegative real line (Corollary B.2); thus, its inverse g is well-defined. Note that g is also strictly increasing and unbounded. The following theorem gives an equivalent expression for the choice probability in terms of g.

Theorem 2.3 (Choice Probability for Context-MNL) *For the Context-MNL model with parameters $\alpha > -1$ and $\boldsymbol{\mu} \in \mathbb{R}^n$, for each $S \subseteq \mathcal{N}$, we have $\phi_i(S) = e^{\mu_i} / g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)$ for all $i \in S$ and $\phi_0(S) = 1/\left[g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)\right]^{1/(1+\alpha)}$.*

Proof: Fix $\alpha > -1$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $S \subseteq \mathcal{N}$. Let $w = \sum_{\ell \in S} v_\ell(S \,|\, \alpha, \boldsymbol{\mu})$. It follows from Equation (2) that $w(1+w)^\alpha = \sum_{\ell \in S} e^{\mu_\ell}$, which is equivalent to

$$\sum_{\ell \in S} e^{\mu_\ell} \;=\; (1+w)^{1+\alpha} - (1+w)^\alpha \;\overset{(a)}{=}\; f\left((1+w)^{1+\alpha}\right) \quad \Leftrightarrow \quad (1+w)^{1+\alpha} \;\overset{(b)}{=}\; g\left(\sum_{\ell \in S} e^{\mu_\ell}\right) \,,$$

where (a) and (b) follow from the definition of f and g in Equation (3). For all $i \in S$, it follows from Theorem 2.2 that $\phi_i(S) = v_i(S)/(1+w) = e^{\mu_i}/(1+w)^{1+\alpha} = e^{\mu_i}/g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)$, which is the desired result. Similarly, $\phi_0(S) = 1/(1+w) = 1/\left[g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)\right]^{1/(1+\alpha)}$. ■

We will use the choice probability expression given in Theorem 2.3 extensively in our analysis of assortment and price optimization problems in later sections. We note that it follows from Theorem 2.3 that under the Context-MNL model, the products are substitutable; that is, the removal of a product from an assortment increases the choice probability of the remaining items in the assortment.

Corollary 2.4 (Substitutability) *For the* Context-MNL *model with parameters* $\alpha > -1$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n$*, each* $S \subseteq \mathcal{N}$ *and* $i \in S \cup \{0\}$*,* $j \in S$ *with* $j \neq i$*, we have* $\phi_i(S \setminus \{j\}) \geq \phi_i(S)$*.*

Proof: By Theorem 2.3, $\phi_0(S) = 1/\left[g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)\right]^{1/(1+\alpha)}$ and $\phi_i(S) = e^{\mu_i}/g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)$ for all $i \in S$. Because g is a strictly increasing function and $1/(1+\alpha) > 0$, we have $g\left(\sum_{\ell \in S \setminus \{j\}} e^{\mu_\ell}\right) \leq g\left(\sum_{\ell \in S} e^{\mu_\ell}\right)$, and the desired result follows. ■

Flexibility in the Modeling of Demand Spill and Recapture: When a customer's preferred choice is unavailable, her first-choice demand is redirected to other alternatives. As defined in Gallego et al. (2015), spill refers to redirected demand that is lost to the no-purchase option, and recapture refers to redirected demand that results in sales of a different product. It is well-known that the standard multinomial logit model is overly optimistic about the recaptured demand. Moreover, under the standard multinomial logit model, when a particular product becomes unavailable, the relative increase in the demand for other available products is constrained to be the same as the relative increase in the spill. To see this, note that the standard multinomial logit model corresponds to a special case of the Context-MNL model with $\alpha = 0$, and it is easy to verify that

$$\underbrace{\frac{\phi_i(S \setminus \{k\} \,|\, \alpha = 0, \boldsymbol{\mu})}{\phi_i(S \,|\, \alpha = 0, \boldsymbol{\mu})}}_{\substack{\text{Increase in Demand for} \\ \text{Product } i \text{ from Recapture}}} \quad = \quad \underbrace{\frac{\phi_0(S \setminus \{k\} \,|\, \alpha = 0, \boldsymbol{\mu})}{\phi_0(S \,|\, \alpha = 0, \boldsymbol{\mu})}}_{\substack{\text{Increase in Demand for the} \\ \text{No-Purchase Option from Spill}}} \qquad \forall\ i \in S \setminus \{k\}\ ,$$

where $\phi_i(S \,|\, \alpha, \boldsymbol{\mu})$ is the choice probability; we explicitly denote the dependence on the parameters α and $\boldsymbol{\mu}$.

By allowing for nonzero α, our proposed Context-MNL model allows for more flexibility in the modeling of demand spill and recapture. Fix $\boldsymbol{\mu} \in \mathbb{R}^n$. Consider an assortment $S \subseteq \mathcal{N}$ and product $k \in S$. If we remove k from S, some of the demand for product k will be recaptured by product i, and the increase in the demand for product $i \in S \setminus \{k\}$ from recapture is $\frac{\phi_i(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_i(S \,|\, \alpha, \boldsymbol{\mu})}$. The increase demand for the no-purchase option from spill is $\frac{\phi_0(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_0(S \,|\, \alpha, \boldsymbol{\mu})}$. The **recapture-to-spill ratio** is defined as $\frac{\phi_i(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_i(S \,|\, \alpha, \boldsymbol{\mu})} \Big/ \frac{\phi_0(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_0(S \,|\, \alpha, \boldsymbol{\mu})}$. Under the standard multinomial logit model, this ratio is always one. The following theorem shows that the recapture-to-spill ratio is strictly increasing in α. As we vary α, we can have recapture-to-spill strictly less than one and strictly larger than one. The proof is given in Appendix B.

Theorem 2.5 (Increasing Recapture-to-Spill Ratio) *For each* $\boldsymbol{\mu} \in \mathbb{R}^n$*,* $S \subseteq \mathcal{N}$*,* $k \in S$*, and* $i \in S \setminus \{k\}$*, the function* $\alpha \mapsto \frac{\phi_i(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_i(S \,|\, \alpha, \boldsymbol{\mu})} \Big/ \frac{\phi_0(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_0(S \,|\, \alpha, \boldsymbol{\mu})}$ *is strictly increasing for all* $\alpha > -1$*.*

The standard multinomial logit model, corresponding to $\alpha = 0$, has a recapture-to-spill ratio of exactly one. Choosing $\alpha \in (-1, 0)$ makes the recapture-to-spill ratio less than one, allowing us to model settings under which the increase in recapture demand is less than the increase in the demand for the no-purchase option from spill. This property of the Context-MNL model provides additional flexibility in the modeling of more realistic spills and recaptures.

3. Assortment Optimization

In this section, we fix the parameters $\alpha > -1$ and $\boldsymbol{\mu} \in \mathbb{R}^n$ and consider the cardinality-constrained assortment optimization problem under the Context-MNL model. Letting $r_i \in \mathbb{R}$ be the marginal profit of product i and K be the maximum assortment size, we are interested in the following optimization problem:

$$Z^* \equiv \max_{S \subseteq \mathcal{N} : |S| \leq K} \sum_{i \in S} r_i \phi_i(S) = \max_{S \subseteq \mathcal{N} : |S| \leq K} \frac{\sum_{i \in S} r_i \, e^{\mu_i}}{g(\sum_{i \in S} e^{\mu_i})} , \tag{4}$$

where the equality follows because $\phi_i(S) = e^{\mu_i}/g\big(\sum_{\ell \in S} e^{\mu_\ell}\big)$ for all $i \in S$ by Theorem 2.3 and g is defined in Equation (3). We can assume without loss of generality that $r_i > 0$ for all $i \in \mathcal{N}$ because according to the substitutability property (Corollary 2.4), it is not optimal to include a product with a nonpositive marginal profit.

The tractability of the assortment optimization problem depends on the value of α. In Section 3.1, we show that for $\alpha \geq 0$, the assortment optimization problem can be solved by evaluating the revenue of at most n^2 assortments. For $-1 < \alpha < 0$, however, the problem is NP-hard, and in Section 3.2 we construct a fully polynomial-time approximation scheme for the problem.

3.1 Computing an Optimal Assortment when $\alpha \geq 0$

The following theorem establishes an important relationship between Z^* and the assortment optimization problem under the standard multinomial logit model. We use this relationship to design an algorithm for computing an optimal assortment.

Theorem 3.1 (Connection to the Standard Multinomial Logit Model) *For each $\alpha \geq 0$ and $\boldsymbol{\mu} \in \mathbb{R}^n$, there exist constants $c_0 \geq 1$ and $c_1 > 0$ such that an assortment S^* is an optimal solution to the optimization problem in Equation (4) if and only if S^* is an optimal solution to the problem $\max_{S \subseteq \mathcal{N} : |S| \leq K} \big(\sum_{i \in S} r_i e^{\mu_i} c_1\big) / \big(c_0 + \sum_{i \in S} e^{\mu_i} c_1\big)$.*

Proof: Fix an arbitrary $\alpha \geq 0$ and $\boldsymbol{\mu} \in \mathbb{R}^n$. Let S^* be an optimal assortment for the problem associated with Z^* in Equation (4), and note that $|S^*| \leq K$. Define $\eta^* = \sum_{i \in S^*} e^{\mu_i}$. Because g :

$\mathbb{R}_+ \to [1,\infty)$ is concave by Corollary B.3, it follows that $1 \le g(x) \le g(\eta^*) + g'(\eta^*)\,(x-\eta^*)$ for all $x \in \mathbb{R}_+$, where the last inequality follows from the subgradient inequality for concave functions. Replacing g by its upper bound, it follows that

$$Z^* \;\ge\; \max_{S\subseteq\mathcal{N}\,:\,|S|\le K} \;\frac{\sum_{i\in S} r_i\, e^{\mu_i}}{g(\eta^*) + g'(\eta^*)\left[\sum_{i\in S} e^{\mu_i} - \eta^*\right]} \;\ge\; \frac{\sum_{i\in S^*} r_i\, e^{\mu_i}}{g(\eta^*)} \;=\; Z^*\;, \tag{5}$$

where the second inequality follows from the fact that $|S^*| \le K$, and from evaluating the function $\left(\sum_{i\in S} r_i\, e^{\mu_i}\right)/\left(g(\eta^*) + g'(\eta^*)\left[\sum_{i\in S} e^{\mu_i} - \eta^*\right]\right)$ at S^*. The final equality follows from the definition of η^*. Thus, all the inequalities above must hold as equalities.

Letting $c_0 = g(\eta^*) - g'(\eta^*)\eta^*$ and $c_1 = g'(\eta^*)$, we note that $1 = g(0) \le g(\eta^*) + g'(\eta^*)\,(0-\eta^*)$, so $c_0 \ge 1$. Because g is strictly increasing by Corollary B.2, we have $c_1 > 0$. It follows from (5) that

$$Z^* \;=\; \max_{S\subseteq\mathcal{N}\,:\,|S|\le K} \;\frac{\sum_{i\in S} r_i\, e^{\mu_i}}{c_0 \;+\; \sum_{i\in S} e^{\mu_i} c_1} \;=\; \frac{1}{c_1} \times \max_{S\subseteq\mathcal{N}\,:\,|S|\le K} \;\frac{\sum_{i\in S} r_i\, e^{\mu_i} c_1}{c_0 \;+\; \sum_{i\in S} e^{\mu_i} c_1}\;,$$

and this completes the proof. ∎

Computing an Optimal Assortment without Knowing c_0 and c_1: We emphasize that the values of the constants $c_0 \ge 1$ and $c_1 > 0$ in Theorem 3.1 are not known in advance because they depend on S^*. However, because of the structure of the optimization problem, we do **not** need to know their values to determine an optimal assortment. To see this, letting $W^* = \max_{S\subseteq\mathcal{N}\,:\,|S|\le K} \left(\sum_{i\in S} r_i\, e^{\mu_i} c_1\right) \big/ \left(c_0 + \sum_{i\in S} e^{\mu_i} c_1\right)$ denote the optimization problem in Theorem 3.1, then we can equivalently express W^* as follows

$$\begin{aligned} W^* \;&=\; \min\left\{ t \,\middle|\, t \ge \frac{\sum_{i\in S} r_i\, e^{\mu_i} c_1}{c_0 + \sum_{i\in S} e^{\mu_i} c_1}\; \forall\; S \subseteq \mathcal{N},\, |S| \le K \right\} \\ &=\; \min\left\{ t \,\middle|\, t c_0 \ge c_1 \sum_{i\in S} (r_i - t)\, e^{\mu_i}\; \forall\; S \subseteq \mathcal{N},\, |S| \le K \right\} = \min\left\{ t \,\middle|\, \frac{c_0 t}{c_1} \;\ge\; \max_{S:|S|\le K} \sum_{i\in S} (r_i - t)\, e^{\mu_i} \right\}. \end{aligned}$$

We construct a collection of assortments $\mathcal{A}$ with the following three properties: (i) $\mathcal{A}$ has at most n^2 assortments, (ii) each assortment in $\mathcal{A}$ has a cardinality of at most K, and (iii) for all $t \in \mathbb{R}_+$, the optimization problem $\max_{S\subseteq\mathcal{N}\,:\,|S|\le K} \sum_{i\in S} (r_i - t)\, e^{\mu_i}$ has an optimal solution in $\mathcal{A}$. It then follows from the above equations that $W^* \;=\; \min\left\{ t \,\middle|\, \frac{c_0 t}{c_1} \;\ge\; \max_{S\in\mathcal{A}} \sum_{i\in S} (r_i - t)\, e^{\mu_i} \right\} = \max_{S\in\mathcal{A}} \frac{\sum_{i\in S} r_i\, e^{\mu_i} c_1}{c_0 + \sum_{i\in S} e^{\mu_i} c_1}$. Thus, to find an optimal assortment, it suffices to evaluate the revenue of at most n^2 assortments in $\mathcal{A}$.

The construction of the collection $\mathcal{A}$ has been discussed in the literature in other contexts (Rusmevichientong et al. 2010, Gallego and Topaloglu 2014). For the sake of completeness, we give a brief overview of the idea here. For each $i \in \mathcal{N}$, the function $h_i(t) = (r_i - t) e^{\mu_i}$ represents a line in the two-dimensional plane. These n lines intersect at no more than $\binom{n}{2} = \frac{n(n-1)}{2}$ intersection points. We use $0 = t^0 \leq t^1 \leq \ldots \leq t^{K-1} \leq t^K = \infty$ with $K = O(n^2)$ to denote the x-coordinates of the intersection points. These points divide the nonnegative real line into $O(n^2)$ intervals. If t takes values in an interval $[t^{k-1}, t^k]$, then the ordering between the values of the functions $\{h_i(t) : i \in \mathcal{N}\}$ remains constant; thus, within each interval, an optimal solution to the problem $\max_{S \subseteq \mathcal{N}\,:\,|S| \leq K} \sum_{i \in S} (r_i - t)\, e^{\mu_i}$ corresponds to the top K lines whose values are nonnegative. Let $\mathcal{A}$ denote the collection of optimal assortments across $O(n^2)$ intervals. It follows that $|\mathcal{A}| \leq n^2$ and each set in $\mathcal{A}$ has a cardinality of at most K. Moreover, the collection contains an optimal solution to the problem $\max_{S \subseteq \mathcal{N}\,:\,|S| \leq K} \sum_{i \in S} (r_i - t)\, e^{\mu_i}$ for all $t \in \mathbb{R}_+$, which is the desired result.

Moreover, when there is no cardinality constraint ($K = n$), the assortment optimization problem associated with Z^* is the unconstrained assortment optimization under the multinomial logit model. It is well-known that the optimal assortment is revenue-ordered (Talluri and van Ryzin 2004); that is, if $r_1 \geq r_2 \geq \cdots \geq r_n$, then S^* is of the form $\{1, 2, \ldots, k\}$ for some product k. The discussion above is summarized in the following corollary.

Corollary 3.2 (Finding an Optimal Assortment) *For each $\alpha \geq 0$, we can compute an optimal assortment for the optimization problem associated with Z^* by evaluating the revenue of at most n^2 assortments. Moreover, when $K = n$, the optimal assortment is revenue-ordered.*

3.2 Computing an Optimal Assortment when $-1 < \alpha < 0$

We study the assortment optimization problem when $-1 < \alpha < 0$. The following theorem shows that when $\alpha = -1/2$ and there is no constraint on the assortment size ($K = n$), the problem is NP-hard. The proof of this result uses a reduction from the partition problem, which is a well-known NP-complete problem (Garey and Johnson 1979), and the proof is presented in Appendix C.

Theorem 3.3 (Complexity) *For $\alpha = -1/2$ and $K = n$, the assortment optimization problem is NP-hard.*

Although the assortment optimization problem is NP-hard, it turns out that we can develop a fully polynomial time approximation scheme (FPTAS) for it. The key idea behind our FPTAS is the observation that we can find in polynomial time a feasible assortment whose revenue is approximately within a certain target. This observation is stated in the following lemma; see also Désir et al. (2022). We provide a proof in Appendix D. For each $\epsilon > 0$ and $(\nu, \xi) \in \mathbb{R}_+ \times \mathbb{R}_+$, let $\mathcal{F}_\epsilon(\nu, \xi) = \{S \subseteq \mathcal{N} : |S| \leq K,\ \sum_{i \in S} r_i e^{\mu_i} \geq \nu,\ \sum_{i \in S} e^{\mu_i} \leq \xi\}$.

Lemma 3.4 (Targeted Assortment) *For each $\epsilon > 0$ and $(\nu,\xi) \in \mathbb{R}_+ \times \mathbb{R}_+$, there exists an algorithm based on dynamic programming with a running time of $O(n^3/\epsilon^2)$ that is independent of ν and ξ, such that whenever $\mathcal{F}_\epsilon(\nu,\xi) \neq \varnothing$, the algorithm outputs a subset $S_{\epsilon,\nu,\xi}$ such that $|S_{\epsilon,\nu,\xi}| \le K$, $\sum_{i\in S_{\epsilon,\nu,\xi}} r_i e^{\mu_i} \ge \nu(1-2\epsilon)$, and $\sum_{i\in S_{\epsilon,\nu,\xi}} e^{\mu_i} \le \xi(1+2\epsilon)$.*

To describe our FPTAS, we let $\mathsf{r}_{\max} = \max\{r_i : i \in \mathcal{N}\}$ and $\mathsf{r}_{\min} = \min\{r_i : i \in \mathcal{N}\}$ denote the maximum and minimum revenues of the products, respectively. Also, let $\mathsf{u}_{\max} = \max\{e^{\mu_i} : i \in \mathcal{N}\}$ and $\mathsf{u}_{\min} = \min\{e^{\mu_i} : i \in \mathcal{N}\}$, and set $\bar{R} = \mathsf{r}_{\max}/\mathsf{r}_{\min}$ and $\bar{U} = \mathsf{u}_{\max}/\mathsf{u}_{\min}$. Given an approximation accuracy $\epsilon > 0$, define the following set of grid points:

$$\Gamma_\epsilon = \left\{ \mathsf{r}_{\min}\mathsf{u}_{\min}(1+\epsilon)^\ell \,:\, \ell = 0,\ldots,\left\lceil \frac{\log(n\bar{R}\bar{U})}{\log(1+\epsilon)} \right\rceil \right\} \text{ and } \Delta_\epsilon = \left\{ \mathsf{u}_{\min}(1+\epsilon)^\ell \,:\, \ell = 0,\ldots,\left\lceil \frac{\log(n\bar{U})}{\log(1+\epsilon)} \right\rceil \right\}.$$

We consider $(\nu,\xi) \in \Gamma_\epsilon \times \Delta_\epsilon$. Here is the description of our algorithm.

FPTAS for the Assortment Optimization Problem when $-1 < \alpha < 0$

INITIALIZATION: We choose an approximation parameter $\epsilon > 0$.
DESCRIPTION: For each $(\nu,\xi) \in \Gamma_\epsilon \times \Delta_\epsilon$, let $\tilde{S}_{\epsilon,\nu,\xi}$ be the output of the algorithm from Lemma 3.4.
OUTPUT: Among the assortments $\{\tilde{S}_{\epsilon,\nu,\xi} : (\nu,\xi) \in \Gamma_\epsilon \times \Delta_\epsilon\}$, return the assortment with the largest expected revenue.

The main result of this section is stated in the following theorem.

Theorem 3.5 (FPTAS) *Our FPTAS returns an assortment whose revenue is at least $(1-6\epsilon)Z^*$ and has a running time of $\mathcal{O}\big((n^3/\epsilon^4) \log(n\bar{R}\bar{U}) \log(n\bar{U}) \big)$.*

Proof: Fix an arbitrary $-1 < \alpha < 0$. Let S^* be the optimal solution to the assortment optimization problem and (q_1,q_2) be such that $\mathsf{r}_{\min}\mathsf{u}_{\min}(1+\epsilon)^{q_1} \le \sum_{j\in S^*} r_j e^{\mu_j} \le \mathsf{r}_{\min}\mathsf{u}_{\min}(1+\epsilon)^{q_1+1}$ and $\mathsf{u}_{\min}(1+\epsilon)^{q_2-1} \le \sum_{j\in S^*} e^{\mu_j} \le \mathsf{u}_{\min}(1+\epsilon)^{q_2}$. For $(\nu,\xi) = (\mathsf{r}_{\min}\mathsf{u}_{\min}(1+\epsilon)^{q_1}, \mathsf{u}_{\min}(1+\epsilon)^{q_2})$, the dynamic programming algorithm from Lemma 3.4 returns an assortment $\tilde{S}$ such that $\sum_{j\in\tilde{S}} r_j e^{\mu_j} \ge \mathsf{r}_{\min}\mathsf{u}_{\min}(1+\epsilon)^{q_1}(1-2\epsilon) \ge \frac{1-2\epsilon}{1+\epsilon}\sum_{j\in S^*} r_j e^{\mu_j}$, where the last inequality follows from the definition of q_1. A similar argument shows that $\sum_{j\in\tilde{S}} e^{\mu_j} \le \mathsf{u}_{\min}(1+\epsilon)^{q_2}(1+2\epsilon) \le (1+2\epsilon)(1+\epsilon)\sum_{j\in S^*} e^{\mu_j}$.

There are two cases: $\sum_{j\in\tilde{S}} e^{\mu_j} \le \sum_{j\in S^*} e^{\mu_j}$ and $\sum_{j\in\tilde{S}} e^{\mu_j} > \sum_{j\in S^*} e^{\mu_j}$. In the first case, if $\sum_{j\in\tilde{S}} e^{\mu_j} \le \sum_{j\in S^*} e^{\mu_j}$, then because g is increasing (Corollary B.2) and $r_j \ge 0$ for all j, we have

$$\frac{\sum_{j\in\tilde{S}} r_j e^{\mu_j}}{g\big(\sum_{j\in\tilde{S}} e^{\mu_j}\big)} \;\ge\; \frac{\sum_{j\in\tilde{S}} r_j e^{\mu_j}}{g\big(\sum_{j\in S^*} e^{\mu_j}\big)} \;\ge\; \frac{1-2\epsilon}{1+\epsilon}\frac{\sum_{j\in S^*} r_j e^{\mu_j}}{g\big(\sum_{j\in S^*} e^{\mu_j}\big)} \ge (1-3\epsilon)Z^* \;,$$

where the second inequality follows from the property of $\tilde{S}$ that $\sum_{j\in\tilde{S}} r_j e^{\mu_j} \ge \frac{1-2\epsilon}{1+\epsilon}\sum_{j\in S^*} r_j e^{\mu_j}$.

In the second case, $\sum_{j\in\tilde{S}} e^{\mu_j} > \sum_{j\in S^*} e^{\mu_j}$. Because $-1<\alpha<0$, it follows from Lemma B.1 that $0<g'(y)\le 1$ for all $y\ge 0$, so $0<g\big(\sum_{j\in\tilde{S}} e^{\mu_j}\big)-g\big(\sum_{j\in S^*} e^{\mu_j}\big)\le \sum_{j\in\tilde{S}} e^{\mu_j} - \sum_{j\in S^*} e^{\mu_j} \le \big((1+2\epsilon)(1+\epsilon)-1\big)\sum_{j\in S^*} e^{\mu_j} = (3\epsilon+2\epsilon^2)\sum_{j\in S^*} e^{\mu_j}$, where the second inequality follows from the property of $\tilde{S}$ that $\sum_{j\in\tilde{S}} e^{\mu_j} \le (1+2\epsilon)(1+\epsilon)\sum_{j\in S^*} e^{\mu_j}$. Letting $W=\sum_{j\in S^*} e^{\mu_j}$ and noting that if $W\le 1$, we have $W\le g(W)$ because the range of g is $[1,\infty)$; on the other hand, if $W>1$, then $f(W)<W$ by the definition of f in Equation (3), and thus $W=g(f(W))<g(W)$. In all cases, $W\le g(W)$, which implies that $0<g\big(\sum_{j\in\tilde{S}} e^{\mu_j}\big)-g\big(\sum_{j\in S^*} e^{\mu_j}\big)\le(3\epsilon+2\epsilon^2)g\big(\sum_{j\in S^*} e^{\mu_j}\big)$, and thus $g\big(\sum_{j\in\tilde{S}} e^{\mu_j}\big) \le (1+3\epsilon+2\epsilon^2)g\big(\sum_{j\in S^*} e^{\mu_j}\big) = (1+2\epsilon)(1+\epsilon)g\big(\sum_{j\in S^*} e^{\mu_j}\big)$. Therefore,

$$\frac{\sum_{j\in\tilde{S}} r_j e^{\mu_j}}{g\big(\sum_{j\in\tilde{S}} e^{\mu_j}\big)} \ \ge\ \frac{((1-2\epsilon)/(1+\epsilon))\sum_{j\in S^*} r_j e^{\mu_j}}{(1+2\epsilon)(1+\epsilon)g\big(\sum_{j\in S^*} e^{\mu_j}\big)} \ \ge\ (1-6\epsilon)Z^*\ ,$$

where the first inequality also uses the property that $\sum_{j\in\tilde{S}} r_j e^{\mu_j} \ge \frac{1-2\epsilon}{1+\epsilon}\sum_{j\in S^*} r_j e^{\mu_j}$.

The above argument establishes a performance guarantee for $\tilde{S}$. Our proposed FPTAS returns a set whose revenue is at least as large as $\tilde{S}$, so we have established that it returns an assortment that is a $(1-6\epsilon)$-approximation to the assortment optimization problem.

For the running time, there are $\left(\left\lceil\frac{\log(n\bar{R}\bar{U})}{\log(1+\epsilon)}\right\rceil+1\right)\left(\left\lceil\frac{\log(n\bar{U})}{\log(1+\epsilon)}\right\rceil+1\right)=O\big((1/\epsilon^2)\log(n\bar{R}\bar{U})\log(n\bar{U})\big)$ guesses for (ν,ξ). For each guess, we run the dynamic program algorithm from Lemma 3.4 with a running time of $O(n^3/\epsilon^2)$, and the total running time of the algorithm is thus $\mathcal{O}\big((n^3/\epsilon^4)\log(n\bar{R}\bar{U})\log(n\bar{U})\big)$. ■

4. Price Optimization

We consider the price optimization problem under the Context-MNL model, which assumes that the parameter $\mu_i(p_i)$ for each product $i\in\mathcal{N}$ is an affine function of its price, with $\mu_i(p_i)=\gamma_i-\beta p_i$ where $\gamma_i\in\mathbb{R}$ is the product-specific parameter and $\beta>0$ is the price sensitivity parameter. Like in earlier studies (Hopp and Xu 2005, Song and Xue 2007, Li and Huh 2011, Zhang et al. 2018), we assume that all products have the same price sensitivity parameter. Given a price vector $\boldsymbol{p}\in\mathbb{R}^n$, the probability $\phi_i(\boldsymbol{p})$ that product i is chosen is given by

$$\phi_i(\boldsymbol{p}) \ = \ \begin{cases} v_i(\boldsymbol{p})/\big(1+\sum_{\ell\in\mathcal{N}} v_\ell(\boldsymbol{p})\big) & \text{if} \quad i\in\mathcal{N} \\ 1/\big(1+\sum_{\ell\in\mathcal{N}} v_\ell(\boldsymbol{p})\big) & \text{if} \quad i=0 \end{cases},$$

where $(v_\ell(\boldsymbol{p}):\ell\in\mathcal{N})$ is the unique solution to the following system of nonlinear equations: $v_i(\boldsymbol{p}) = e^{\gamma_i-\beta p_i-\alpha\log\left(1+\sum_{\ell\in\mathcal{N}} v_\ell(\boldsymbol{p})\right)}$ for all $i\in\mathcal{N}$.

Letting $c_i\ge 0$ be the marginal cost of product i, the price optimization problem is given by:

$$Y^* \ \equiv\ \max_{\boldsymbol{p}\in\mathbb{R}^n} \sum_{i\in\mathcal{N}} (p_i-c_i)\,\phi_i(\boldsymbol{p})\ , \tag{6}$$

and let $\boldsymbol{p}^* = (p_1^*, \ldots, p_n^*)$ denote an optimal price. The following theorem shows that for all $\alpha > -1$, the optimal price $\boldsymbol{p}^*$ has the constant-markup property according to which the optimal price of each product exceeds its marginal cost by the same constant. We also give the formula for the optimal markup. To state our result, for each $\alpha > -1$, we let the function $Q : [1, \infty) \to \mathbb{R}$ be defined as follows: for each $x \geq 1$, $Q(x) = -\dfrac{f(x)}{x} \log\left(\dfrac{f(x)}{\sum_{i \in \mathcal{N}} e^{\gamma_i - \beta c_i}}\right)$, where f is defined in Equation (3).

Theorem 4.1 (Constant Markup is Optimal) *For each $\alpha > -1$, we have that*

(a) *The function Q is quasi-concave and has a unique maximizer h^* such that $h^* < g\big(\sum_{i \in \mathcal{N}} e^{\gamma_i - \beta c_i}\big)$, where g is defined in Equation (3).*

(b) *The optimal price is given by $p_i^* = c_i + \frac{1}{\beta} \log\left(\frac{\sum_{\ell \in \mathcal{N}} e^{\gamma_\ell - \beta c_\ell}}{f(h^*)}\right)$ for all $i \in \mathcal{N}$, and $Y^* = \frac{1}{\beta} Q(h^*)$.*

Before proceeding to the proof, we discuss the implication of the theorem. Let $A = \sum_{i \in \mathcal{N}} e^{\gamma_i - \beta c_i}$. The *markup* of product i under the optimal price is $p_i^* - c_i$, which is the optimal amount to charge for product i beyond its marginal cost. By Theorem 4.1(b), under the optimal price $\boldsymbol{p}^*$, the markup is the **same** for all the products and is equal to $\frac{1}{\beta} \log\left(\frac{A}{f(h^*)}\right)$. Moreover, it follows from Theorem 4.1(a) that $h^* < g(A)$, which implies that $f(h^*) < f(g(A)) = A$ because f is strictly increasing and g is the inverse of f, so $\frac{1}{\beta} \log\left(\frac{A}{f(h^*)}\right) > 0$ and the optimal markup is positive. Moreover, the theorem allows us to reduce the n-dimensional optimization problem to a maximization of a one-dimensional quasi-concave function Q over a compact interval $[0, g(A)]$, which can be done efficiently using a method like the golden-section search (Press et al. 2007).

We note that in the special case of $\alpha = 0$, the formula in Theorem 4.1 reduces to the optimal pricing formula for the standard multinomial logit model. In particular, when $\alpha = 0$, we have $f(x) = x - 1$, $Q(x) = -\frac{x-1}{x} \log(\frac{x-1}{A})$, and $Q'(x) = \frac{1}{x^2} \left(\log(A) - \log(x-1) - x\right)$. This implies that

$$Q'(h^*) = 0 \Leftrightarrow \frac{A}{h^* - 1} = e^{h^*} \Leftrightarrow (h^* - 1) e^{h^* - 1} = A e^{-1} \Leftrightarrow f(h^*) = h^* - 1 = \mathsf{LW}(A e^{-1}) ,$$

where $\mathsf{LW}(\cdot)$ is the Lambert-W function. Then, the optimal markup is given by

$$\frac{1}{\beta} \log\left(\frac{A}{f(h^*)}\right) = \frac{\log A - \log \mathsf{LW}(A e^{-1})}{\beta} = \frac{\log A - \log(A e^{-1}) + \mathsf{LW}(A e^{-1})}{\beta} = \frac{1 + \mathsf{LW}(A e^{-1})}{\beta} ,$$

and we have $Z^* = \frac{1}{\beta} Q^*(h^*) = \frac{1}{\beta} \mathsf{LW}(A e^{-1})$. We thus recover the optimal pricing formula for the standard multinomial logit model (Aydin and Ryan 2000, Li and Graves 2012).

Here is the proof of Theorem 4.1.

Proof: Fix $\alpha > -1$. Let $A = \sum_{i \in \mathcal{N}} e^{\gamma_i - \beta c_i}$. The proof of part (a) is a straightforward application of calculus and is given in Appendix E. We focus on part (b). Using the same argument as in Theorem 2.3, we can show that $\phi_i(\boldsymbol{p}) = e^{\gamma_i - \beta p_i} / g\big(\sum_{\ell \in S} e^{\gamma_\ell - \beta p_\ell}\big)$ for all $i \in \mathcal{N}$. Thus,

$$Y^* = \max_{\boldsymbol{p} \in \mathbb{R}^n} \frac{\sum_{i \in \mathcal{N}} (p_i - c_i) e^{\gamma_i - \beta p_i}}{g\big(\sum_{i \in \mathcal{N}} e^{\gamma_i - \beta p_i}\big)} = \max_{h \geq 1} \max_{\boldsymbol{p} \in \mathbb{R}^n} \left\{ \frac{1}{h} \sum_{i \in \mathcal{N}} (p_i - c_i) e^{\gamma_i - \beta p_i} \ : \ g\Big(\sum_{i \in \mathcal{N}} e^{\gamma_i - \beta p_i}\Big) = h \right\}$$

$$= \max_{h\geq 1}\max_{\boldsymbol{p}\in\mathbb{R}^n} \left\{\frac{1}{h}\sum_{i\in\mathcal{N}}(p_i-c_i)\,e^{\gamma_i-\beta p_i} \;:\; \sum_{i\in\mathcal{N}} e^{\gamma_i-\beta p_i}=f(h)\right\}\;,$$

where the last equality follows because f is the inverse function of g. Using the change of variables $q_i=e^{\gamma_i-\beta p_i}$, it follows that $p_i = \frac{1}{\beta}(\gamma_i-\log q_i)$, which implies that $Y^*=\max_{h\geq 1}\max_{\boldsymbol{q}\in\mathbb{R}^n_+}\left\{\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\beta c_i-\log q_i)q_i \;:\; \sum_{i\in\mathcal{N}}q_i=f(h)\right\}$. Note that the mapping $x\mapsto(a-\log x)x$ is concave on the nonnegative real line, so for each $h\geq 1$, the inner maximization $\max_{\boldsymbol{q}\in\mathbb{R}^n_+}\left\{\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\beta c_i-\log q_i)q_i \;:\; \sum_{i\in\mathcal{N}}q_i=f(h)\right\}$ is a concave maximization subject to a linear constraint, and thus strong duality holds (Boyd and Vandenberghe 2004). Thus, for each $h\geq 1$,

$$\begin{aligned}
\max_{\boldsymbol{q}\in\mathbb{R}^n_+}\left\{\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\log q_i)q_i : \sum_{i\in\mathcal{N}}q_i=f(h)\right\} &= \min_{\lambda\in\mathbb{R}}\max_{\boldsymbol{q}\in\mathbb{R}^n_+}\left\{\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\log q_i)q_i + \lambda\Big(f(h)-\sum_{i\in\mathcal{N}}q_i\Big)\right\}\\
&= \min_{\lambda\in\mathbb{R}}\left\{\lambda f(h)+\max_{\boldsymbol{q}\in\mathbb{R}^n_+}\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\beta c_i-\beta h\lambda-\log q_i)q_i\right\}\\
&= \min_{\lambda\in\mathbb{R}}\left\{\lambda f(h)+\frac{1}{h\beta}\sum_{i\in\mathcal{N}}\max_{q_i\geq 0}\;(\gamma_i-\beta c_i-\beta h\lambda-\log q_i)q_i\right\}\\
&= \min_{\lambda\in\mathbb{R}}\left\{\lambda f(h)+\frac{e^{-1}A}{h\beta}e^{-\beta h\lambda}\right\}\;,
\end{aligned}$$

where the second equality follows from algebra and the third equality follows because the optimization over $\boldsymbol{q}$ is separable. The final equality follows because $\max_{x\geq 0}(a-\log x)x=e^{a-1}$ and the unique maximizer is at $x^*=e^{a-1}$. Thus, for each $i\in\mathcal{N}$, $\max_{q_i\geq 0}\;(\gamma_i-\beta c_i-\beta h\lambda-\log q_i)q_i = e^{\gamma_i-\beta c_i-\beta h\lambda-1}$, and the unique maximizer $q_i^*(h,\lambda)$ is given by

$$q_i^*(h,\lambda) \;=\; e^{\gamma_i-\beta c_i-\beta h\lambda-1} \qquad\Leftrightarrow\qquad p_i^*(h,\lambda)-c_i \;=\; \frac{1}{\beta}+h\lambda\;. \tag{7}$$

Moreover, summing over all $i\in\mathcal{N}$ yields $\frac{1}{h\beta}\sum_{i\in\mathcal{N}}(\gamma_i-\beta c_i-\beta h\lambda-\log q_i)q_i=\frac{e^{-1}A}{h\beta}e^{-\beta h\lambda}$. Putting everything together, we have $Y^*=\max_{h\geq 1}\min_{\lambda\in\mathbb{R}}\;\lambda f(h)\;+\;\frac{e^{-1}A}{h\beta}e^{-\beta h\lambda}$.

For each $h\geq 1$, the mapping $\lambda\mapsto\lambda f(h)\;+\;\frac{e^{-1}A}{\beta h}e^{-\beta h\lambda}$ is strictly convex in λ, and its unique minimizer $\lambda^*(h)$ is given by $\lambda^*(h)=\dfrac{1}{\beta h}\log\left(\dfrac{e^{-1}A}{f(h)}\right)$ and

$$\min_{\lambda\in\mathbb{R}}\;\lambda f(h)\;+\;\frac{e^{-1}A}{\beta h}e^{-\beta h\lambda} \;=\; \frac{f(h)}{\beta h}\log\left(\frac{e^{-1}A}{f(h)}\right)\;+\;\frac{f(h)}{\beta h} \;=\; \frac{f(h)}{\beta h}\log\left(\frac{A}{f(h)}\right) \;=\; \frac{1}{\beta}Q(h)\;.$$

Thus, $Y^* \;=\; \max_{h\geq 1}\;\frac{1}{\beta}Q(h) \;=\; \frac{1}{\beta}Q(h^*)$, where the last equality follows from part (a). Then, we have $\lambda^*(h^*)=\frac{1}{\beta h^*}\log\left(\frac{e^{-1}A}{f(h^*)}\right)$, and by Equation (7), we have established that for all $i\in\mathcal{N}$,

$$p_i^*(h^*,\lambda^*)-c_i \;=\; \frac{1}{\beta}+h^*\lambda^* \;=\; \frac{1}{\beta}\;+\;\frac{1}{\beta}\log\left(\frac{e^{-1}A}{f(h^*)}\right) \;=\; \frac{1}{\beta}\log\left(\frac{A}{f(h^*)}\right)\;,$$

which completes the proof. ■

5. Estimation

We want to show that the parameters of the $\mathsf{Context\text{-}MNL}$ model can be estimated efficiently by showing that the maximum likelihood estimate can be computed efficiently. Here is the setup for our estimation problem. The dataset is given by $\{(S_t, c_t) : t = 1, \ldots, T\}$, which consists of the assortment S_t shown to each customer t and the selection c_t of customer t, including the no-purchase option. We assume that the customers' selections are independent and identically distributed. By Theorem 2.3, for each customer t, the probability of observing the selection c_t is given by $e^{\mu_{c_t}}/g_\alpha(\sum_{\ell \in S_t} e^{\mu_\ell})$ if $c_t \neq 0$, and is $1/\left[g_\alpha\left(\sum_{\ell \in S_t} e^{\mu_\ell}\right)\right]^{1/(1+\alpha)}$ if $c_t = 0$. Instead of g, we write g_α with an explicit dependence on α to emphasize that the function depends on the parameter α. The *negative log-likelihood* function associated with the dataset is then given by

$$\mathsf{NLL}(\alpha, \boldsymbol{\mu}) = \sum_{t=1}^{T} \left\{ \mathbb{1}_{\{c_t \neq 0\}} \log\left(g_\alpha\Big(\sum_{\ell \in S_t} e^{\mu_\ell}\Big)\right) + \mathbb{1}_{\{c_t = 0\}} \frac{\log\left(g_\alpha\left(\sum_{\ell \in S_t} e^{\mu_\ell}\right)\right)}{1+\alpha} \right\} - \sum_{t=1}^{T} \mathbb{1}_{\{c_t \neq 0\}} \mu_{c_t} \,. \quad (8)$$

The maximum likelihood estimate is obtained by solving the optimization problem $\min_{\alpha > -1,\ \boldsymbol{\mu} \in \mathbb{R}^n} \mathsf{NLL}(\alpha, \boldsymbol{\mu})$, where we are minimizing because we are working with the negative log-likelihood. The main result is stated in the following theorem.

Theorem 5.1 (Convexity) *For each $\alpha > -1$, the function $\boldsymbol{\mu} \mapsto \mathsf{NLL}(\alpha, \boldsymbol{\mu})$ is convex in $\boldsymbol{\mu}$.*

Theorem 5.1 gives us the following method for computing the maximum likelihood estimate. We first discretize α by creating a one-dimensional grid on the interval $(-1, -\infty)$. For each α on the grid, we can solve the problem $\min_{\boldsymbol{\mu} \in \mathbb{R}^n} \mathsf{NLL}(\alpha, \boldsymbol{\mu})$ efficiently because it is an unconstrained convex minimization problem. Then, we simply choose the value of the α in the grid that gives the minimum objective value. Because α is a one-dimensional parameter, the grid search can be done relatively efficiently, and we can improve the quality of our solution by refining the grid.

The above result also implies that if each μ_i is parameterized as a linear combination of product features, then for each α, the negative log-likelihood function remains convex in the product-feature coefficients, and we can estimate these coefficients efficiently by minimizing the convex negative log-likelihood function.

We emphasize that the convexity result of Theorem 5.1 is the first such result for a choice model with endogenous effects whose preference weights are defined implicitly through a system of nonlinear equations. Wang and Wang (2017) was the first paper to consider a variant of the standard multinomial logit model with an endogenous network effect. The choice probabilities in their model are also defined as a solution to a system of nonlinear equations, but in their model, the negative log-likelihood function is not convex in the problem parameters, and they are thus unable to determine the maximum likelihood estimate. They offer a heuristic that yields a stationary point under certain technical conditions, with no guarantee on how close their estimate is to the true maximum likelihood estimate.

The fact that our model admits an efficient method for computing the maximum likelihood estimate gives further support to our formulation. With an efficient estimation procedure, we can calibrate the model's parameters from real data to determine whether there are statistically significant assortment-dependent effects on the utility of each product.

The proof of Theorem 5.1 makes use of a series of lemmas. The first lemma establishes a lower bound on the derivative of the composition of the natural logarithm and g_α. The proof is in Appendix F. Recall that g_α is the inverse of the function f_α defined in Equation (3). Because α is a parameter that we need to estimate, we write the functions f and g as f_α and g_α, respectively, to denote the explicit dependence of these functions on α. Note that for each $\alpha > -1$, $g_\alpha : \mathbb{R}_+ \to [1, \infty)$; it is clear that g_α is a strictly increasing function because f_α is.

Lemma 5.2 *For each $\alpha > -1$, if $h_\alpha = \log \circ g_\alpha$, then $h''_\alpha(y) + \frac{h'_\alpha(y)}{y} \geq 0$ for all $y > 0$.*

The next lemma uses the property of the derivative of g_α to establish the convexity of the mapping $\boldsymbol{\mu} \mapsto \log(g_\alpha(\sum_{\ell=1}^n e^{\mu_\ell}))$. This result can be viewed as a generalization of the convexity of the log-sum-exp function because when $\alpha = 0$, we have $g_\alpha(y) = y$, and this function reduces to the classical log-sum-exp function, which is known to be convex. This result is key to establishing the convexity of our negative log-likelihood function.

Lemma 5.3 (Convexity of Generalized Log-Sum-Exp) *For each $\alpha > -1$, the mapping $\boldsymbol{\mu} \mapsto \log(g_\alpha(\sum_{\ell=1}^n e^{\mu_\ell}))$ is convex in $\boldsymbol{\mu}$.*

Proof: Fix an arbitrary $\alpha > -1$. Let $h = \log \circ g_\alpha$ denote the composition of the natural log and g_α. Let $W : \mathbb{R}^n \to \mathbb{R}$ be defined by: for each $\boldsymbol{\mu} \in \mathbb{R}^n$, $W(\boldsymbol{\mu}) = h(\sum_{\ell=1}^n e^{\mu_\ell})$. Then, for all i and j,

$$\frac{\partial W}{\partial \mu_i}(\boldsymbol{\mu}) = e^{\mu_i}\, h'\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) \text{ and } \frac{\partial^2 W}{\partial \mu_i \partial \mu_j}(\boldsymbol{\mu}) = \begin{cases} e^{\mu_i} e^{\mu_i}\, h''\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) + e^{\mu_i}\, h'\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) & \text{if } i = j \\ e^{\mu_i} e^{\mu_j}\, h''\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) & \text{if } i \neq j \end{cases},$$

which implies that the n-by-n Hessian matrix $\nabla^2 W(\boldsymbol{\mu})$ of W is given by

$$\nabla^2 W(\boldsymbol{\mu}) \;=\; h''\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) \begin{bmatrix} e^{\mu_1} \\ e^{\mu_2} \\ \vdots \\ e^{\mu_n} \end{bmatrix} [e^{\mu_1} e^{\mu_2} \cdots e^{\mu_n}] \;+\; h'\Big(\sum_{\ell=1}^n e^{\mu_\ell}\Big) \mathsf{diag}\,(e^{\mu_1}, e^{\mu_2}, \ldots, e^{\mu_n}) \;,$$

where $\mathsf{diag}\,(a_1,\ldots,a_n)$ is a diagonal matrix with entries $a_1,\ldots,a_n$.

To complete the proof, it suffices to show that $\nabla^2 W(\boldsymbol{\mu})$ is positive semidefinite for all $\boldsymbol{\mu}$; that is, $\boldsymbol{x}^\top \nabla^2 W(\boldsymbol{\mu})\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Letting $s = \sum_{\ell=1}^n e^{\mu_\ell}$ and noting that $s > 0$, it follows from the above equality that

$$\boldsymbol{x}^\top \nabla^2 W(\boldsymbol{\mu})\boldsymbol{x} \;=\; h''(s)\left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 + h'(s)\sum_{i=1}^n e^{\mu_i} x_i^2 \;.$$

If $\left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 = 0$, then we have $\boldsymbol{x}^\top \nabla^2 W(\boldsymbol{\mu})\boldsymbol{x} \geq 0$ because $h'(s) \geq 0$ since h is a composition of logarithm and g, both of which are increasing. So, suppose that $\left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 > 0$. It follows that

$$\begin{aligned}
\boldsymbol{x}^\top \nabla^2 W(\boldsymbol{\mu})\boldsymbol{x} \;&=\; \left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 \left(h''(s) + \frac{h'(s)}{s}\cdot \frac{\left(\sum_{i=1}^n e^{\mu_i} x_i^2\right)\left(\sum_{\ell=1}^n e^{\mu_\ell}\right)}{\left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2}\right)\\
&\overset{(a)}{\geq}\; \left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 \left(h''(s) + \frac{h'(s)}{s}\right) \;\overset{(b)}{\geq}\; 0\;,
\end{aligned}$$

where (a) follows from $h'(s) > 0$ and the Cauchy–Schwarz inequality, which shows that

$$\left(\sum_{i=1}^n e^{\mu_i} x_i\right)^2 \;=\; \left(\sum_{i=1}^n e^{\mu_i/2} e^{\mu_i/2} x_i\right)^2 \;\leq\; \left(\sum_{i=1}^n e^{\mu_i} x_i^2\right)\left(\sum_{i=1}^n e^{\mu_i}\right)\;.$$

The last inequality (b) follows from Lemma 5.2. This completes the proof. ■

We are now ready to give the proof of the main theorem.

Proof of Theorem 5.1: The second term in the expression for the negative log likelihood in Equation (8) is $-\sum_{t=1}^T \mathbb{1}_{\{c_t \neq 0\}} \mu_{c_t}$, which is linear in $\boldsymbol{\mu}$. The first term is equal to $\sum_{t=1}^T \left\{ \mathbb{1}_{\{c_t \neq 0\}} \log\left(g_\alpha\left(\sum_{\ell \in S_t} e^{\mu_\ell}\right)\right) \;+\; \mathbb{1}_{\{c_t = 0\}} \frac{\log\left(g_\alpha\left(\sum_{\ell \in S_t} e^{\mu_\ell}\right)\right)}{1+\alpha}\right\}$. For each $\alpha > -1$, it follows from Lemma 5.3 that the function inside the braces is convex in $\boldsymbol{\mu}$, so the first term is also convex in $\boldsymbol{\mu}$. This gives the desired result. ■

6. Numerical Experiments

We provide two sets of computational experiments. First, we use a dataset from Expedia to compare our Context-MNL model with the standard multinomial logit model in terms of their abilities to predict customer purchases. Second, we test the performance of the FPTAS we construct in Section 3.2, where we develop upper bounds on the optimal expected revenue and use the upper bounds to check optimality gaps attained by our FPTAS.

6.1 Prediction Accuracy of the Context-MNL Model on the Expedia Dataset

Using the publicly available dataset provided by Expedia as part of a Kaggle competition (Kaggle 2013), we conduct numerical experiments to test the accuracy with which the Context-MNL model predicts customer purchases.

Description of Dataset: The dataset provides the results of search queries for hotels on Expedia. In the dataset, the rows correspond to different hotels displayed in different search queries performed by different customers, and the columns present information on the characteristics of the displayed hotels in a search query, the characteristics of the requested stay, and the booking decision of the customer. The dataset has fifteen columns. The first column in the dataset is the unique code for each query, and this information allows us to identify all the hotels displayed in each search query, which represent the assortment of products from which a customer chooses. The second column indicates whether the customer booked the hotel in the search query, allowing us to identify the customer's purchase. Each customer books at most one hotel in a search query, but she may choose not to book a hotel. The remaining columns show the display position of the hotel in the search query, the star rating and the average review score for the hotel, an indicator for whether the hotel is part of a chain, two location desirability scores, the average price of the hotel over the last trading period, the displayed price, an indicator for whether the hotel is on promotion, the number of days until the day of stay, the number of adults and children in the search query, and an indicator for whether the requested stay is over the weekend.

We preprocessed the dataset to remove values that were either missing or uninterpretable; this procedure resulted in a final dataset with 15 columns and $595,965$ rows, representing $34,561$ queries. Among all the queries, $5,848$ queries (about 17%) resulted in a booking, and we refer to these as booking queries; the remaining $28,713$ queries (about 83%) did not result in a booking, and we refer to these as nonbooking queries. The average number of hotels displayed in a search query is 17.24, and the the maximum number of hotels displayed is 37. In Appendix G, we explain our approach to preprocessing the dataset.

Parameter Estimation: To enrich our experimental setup, we use bootstrapping to generate multiple datasets based on the Expedia dataset. We ensure that each dataset we generate has n queries and that a fraction p_0 of the n queries do not result in a booking, where n and p_0 are values chosen such that np_0 is an integer. To obtain such dataset, we randomly sample $n(1-p_0)$ queries from the booking queries with replacement and randomly sample np_0 queries from the nonbooking queries with replacement. Putting the two samples together, we obtain a dataset with n queries, and a fraction p_0 of these queries do not result in any booking. The value of p_0 controls the balance in the dataset between the number of customers who book a hotel and those who do not. For each value of n and p_0, we repeat the bootstrapping procedure 10 times to obtain 10 datasets. We vary $n = \{10000, 20000, 30000\}$ and $p_0 = \{0.3, 0.5, 0.7\}$, which results in 90 datasets in our experiment.

Under our Context-MNL model, we assume that the parameter μ_i of hotel i is given by $\mu_i = \beta_0 + \sum_{\ell=1}^{12} \beta_\ell x_{i,\ell}$, where $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,12})$ are the values in the dataset of the last 12 columns giving the characteristics of hotel i and the corresponding requested stay, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{12}) \in \mathbb{R}^{13}$ are the coefficients that capture the impact of each characteristic. Therefore, the parameters of our Context-MNL model are the coefficients $\boldsymbol{\beta}$ and parameter $\alpha > -1$. In our estimation, we represent $\alpha = e^{\gamma} - 1$ where $\gamma \in \mathbb{R}$. As noted in the discussion that immediately follows Theorem 5.1, for each α, the negative log-likelihood function is convex in $\boldsymbol{\beta}$. For each dataset, we randomly split 80% of the data as the training dataset and use the remaining 20% as the test dataset. We obtain the maximum likelihood estimate by using the fminunc function in MATLAB to minimize the negative log-likelihood function. By setting $\alpha = 0$, we also estimate the parameters of the standard multinomial logit model. Throughout this section, we refer to our Context-MNL model as "CML" and the standard multinomial logit model as "SML".

Results on Prediction Accuracy: We use two performance measures to compare CML and SML. The first performance measure is the out-of-sample log-likelihood evaluated on the test dataset, which is then normalized by dividing by the number of queries n in the test dataset. The normalized log-likelihood allows for apple-to-apple comparison across different-sized datasets. The second performance measure is the k-hit score on the test dataset. To compute the k-hit score of the fitted CML model, for each query that results in a booking in the test dataset, we compute the selection probability of each offered hotel based on the estimated parameters computed from the training dataset. If the actual hotel booked by a customer is among the the hotels with the k-highest probabilities, then we say that the query is a k-hit under the CML model. The k-hit score for the CML model is the fraction of the k-hit queries among all the queries that result in a booking. The k-hit score of the fitted SML model is similar. We use $k \in \{1, 2, 3\}$. For the k-hit score, we focus only on the search queries resulting in a booking, because a large fraction of the customers do not book. If we included the nonbooking search queries in the k-hit score, then the k-hit scores would be driven mainly by the customers who do not book, which would undermine our objective of testing the accuracy with which we can predict the specific hotel booked.

Table 1 presents computational results. Each row in the table corresponds to a different value of (n, p_0). Recall that we generate 10 datasets for each value of (n, p_0). In the top portion, we compare the out-of-sample log-likelihoods of CML and SML, normalized by the number of queries n. The

(n, p_0)	Out-of-Sample log-likelihood (Normalized by n)				
	CML $\succ$ SML	CML LogLik.	SML LogLik.	Avg. % Gap	Std. Err.
(10000, 0.3)	10	-2.3470	-2.3945	1.99%	0.13%
(10000, 0.5)	10	-1.9135	-1.9663	2.69%	0.15%
(10000, 0.7)	10	-1.3617	-1.3922	2.20%	0.30%
(20000, 0.3)	10	-2.3428	-2.3916	2.04%	0.07%
(20000, 0.5)	10	-1.9320	-1.9807	2.46%	0.08%
(20000, 0.7)	10	-1.3609	-1.3930	2.31%	0.20%
(30000, 0.3)	10	-2.3490	-2.4005	2.14%	0.08%
(30000, 0.5)	10	-1.9206	-1.9694	2.48%	0.08%
(30000, 0.7)	10	-1.3528	-1.3859	2.39%	0.15%

(n, p_0)	1-Hit Score				2-Hit Score				3-Hit Score			
	CML $\succ$ SML	CML 1-hit	Avg. % Gap.	Std. Err.	CML $\succ$ SML	CML 2-hit	Avg. % Gap	Std. Err.	CML $\succ$ SML	CML 3-hit	Avg. % Gap.	Std. Err.
(10000, 0.3)	8	0.23	1.09%	0.41%	9	0.36	0.96%	0.25%	9	0.46	1.31%	0.25%
(10000, 0.5)	10	0.23	1.29%	0.29%	9	0.36	1.54%	0.45%	9	0.46	1.61%	0.37%
(10000, 0.7)	7	0.23	0.78%	0.42%	8	0.35	0.96%	0.36%	8	0.45	1.38%	0.71%
(20000, 0.3)	8	0.23	0.79%	0.22%	9	0.35	1.02%	0.17%	10	0.46	1.03%	0.11%
(20000, 0.5)	8	0.22	0.85%	0.28%	10	0.36	1.42%	0.17%	10	0.46	1.28%	0.20%
(20000, 0.7)	10	0.23	1.80%	0.25%	10	0.36	1.91%	0.28%	9	0.46	1.85%	0.32%
(30000, 0.3)	10	0.23	0.72%	0.14%	10	0.35	0.98%	0.12%	10	0.46	1.07%	0.11%
(30000, 0.5)	10	0.23	1.20%	0.17%	8	0.35	0.93%	0.30%	10	0.46	1.00%	0.18%
(30000, 0.7)	10	0.23	1.46%	0.24%	10	0.36	2.27%	0.26%	10	0.46	2.76%	0.25%

Table 1 Performance comparison of the fitted CML and SML models using the Expedia dataset.

second column shows the number of datasets out of 10 in which the normalized out-of-sample log-likelihood of the fitted CML model is larger than that of SML. The third and fourth columns, respectively, show the average normalized out-of-sample log-likelihood of the fitted CML and SML models, under which the average is over the 10 datasets. The fifth and sixth columns, respectively, show the average and standard error of the percentage gaps between the normalized out-of-sample log-likelihoods of the two fitted choice models, under which the standard error is the standard deviation of the percentage gaps over the 10 datasets divided by $\sqrt{10}$.

In the bottom portion of Table 1, we compare the k-hit scores. The second column shows the number of datasets out of 10 in which the 1-hit score of the fitted CML model is larger than that of SML. The third column shows the average 1-hit score of the fitted CML model over the 10 datasets. The fourth and fifth columns, respectively, show the average and standard error of the percentage gaps between the 1-hit scores of the two fitted choice models. Positive values favor CML. We compare the 2-hit and 3-hit scores similarly.

The fitted CML model improves the out-of-sample log-likelihoods of the fitted SML model in 90 out of 90 datasets, with an average improvement of 2.30%. To further quantify the improvements in the prediction accuracies, we turn to k-hit scores. The fitted CML model improves the 1-hit score of the fitted SML model in 81 out of 90 datasets. The fitted CML model improves the 1-hit score by 1.11% on average. The gaps between the k-hit scores are maintained for $k \in \{2, 3\}$, with our CML providing an average improvement of 1.33% and 1.48% for the 2-hit and 3-hit scores, respectively.

The average improvement across 1-hit, 2-hit, and 3-hit scores is 1.31%. Noting the 3-hit scores, one of the three alternatives with the largest purchase probabilities ends up being the hotel booked by the customer about 46% of the time. Our bootstrapped datasets are independent samples. In testing the hypothesis that the CML yields higher out-of-sample log-likelihoods and k-hit scores, all the average gaps in the out-of-sample log-likelihoods and k-hit scores are statistically significant in a one-sided paired t-test at the 95% level (see Goulden (1939), Chapter 4.6).

We also conduct tests of the statistical significance of the estimated coefficients $\boldsymbol{\beta}$ and α. All the estimated coefficients are statistically significant, with p-values less than 10^{-5} when we test the null hypothesis that a coefficient is zero. In particular, using both likelihood-ratio and Quasi-t tests, we can reject the null hypothesis that $\alpha = 0$ in all 90 of the datasets used in our experiments, with both methods yielding p-values less than 10^{-5}. We present details of the procedure for testing the statistical significance of α in Appendix H.

6.2 Assortment Optimization when $\alpha \in (-1, 0)$

We test the practical performance of the FPTAS given in Section 3.2 for the unconstrained assortment optimization problem when $\alpha \in (-1, 0)$, which is NP-hard by Theorem 3.3.

Upper Bound: To assess the optimality gap of the solutions obtained by our FPTAS, we give an efficiently computable upper bound on the optimal objective value Z^* of the unconstrained assortment optimization problem, where $Z^* = \max_{\boldsymbol{x} \in \{0,1\}^n} \big(\sum_{i \in \mathcal{N}} r_i e^{\mu_i} x_i\big) / g\big(\sum_{i \in \mathcal{N}} e^{\mu_i} x_i\big)$ by Equation (4). For each product i, the binary variable x_i determines whether product i is offered. We upper bound Z^* by solving the continuous relaxation $W^* = \max_{\boldsymbol{x} \in [0,1]^n} \big(\sum_{i \in \mathcal{N}} r_i e^{\mu_i} x_i\big) / g\big(\sum_{i \in \mathcal{N}} e^{\mu_i} x_i\big)$, allowing each x_i to be any real number between zero and one. We can compute W^* using bisection search because for each $t \ge 0$,

$$W^* \ge t \quad \Leftrightarrow \quad \max_{\boldsymbol{x} \in [0,1]^n} \sum_{i \in \mathcal{N}} r_i e^{\mu_i} x_i - t\, g\Big(\sum_{i \in \mathcal{N}} e^{\mu_i} x_i\Big) \;\ge\; 0 \;.$$

Because g is convex for $\alpha \in (-1, 0)$ by Corollary B.3, the optimization problem $\max_{\boldsymbol{x} \in [0,1]^n} \sum_{i \in \mathcal{N}} r_i e^{\mu_i} x_i - t\, g\big(\sum_{i \in \mathcal{N}} e^{\mu_i} x_i\big)$ is a concave maximization problem, which can be solved efficiently for each $t \ge 0$. We can then use bisection search to determine W^* by iteratively guessing different values of t. Starting with an initial interval $[0, \max_{i \in \mathcal{N}} r_i]$, each iteration of the bisection search maintains an interval that is guaranteed to contain W^*. By checking whether or not W^* is greater than or equal to the midpoint of the interval in the current iteration, the interval is reduced by half in the next iteration. We stop the bisection search when the length of the interval is less than 10^{-8}.

To solve the optimization problem $\max_{\boldsymbol{x}\in[0,1]^n} \sum_{i\in\mathcal{N}} r_i e^{\mu_i} x_i - t\, g\big(\sum_{i\in\mathcal{N}} e^{\mu_i} x_i\big)$, we exploit the convexity of g and replace g by its lower bound, which is a piecewise-linear function on a grid of increment 0.01. The slope of the lower bounding function corresponds to $g'(y)$ for different values of y on a grid. When replacing g by its piecewise-linear lower bound, it is easy to verify that the resulting optimization problem can be written as a linear program. Moreover, because we work with a lower bound on g, the value of t that we find through this process remains an upper bound on W^*.

Experimental Setup: We randomly generate a large number of test problems and compare the expected revenue from the solution obtained by our FPTAS with the upper bound on the optimal expected revenue. In all of our test problems, the number of products is $n=32$. For each product i, we generate r_i from the uniform distribution over the interval $[1,10]$. We then reindex $\{r_1,\ldots,r_n\}$ so that $r_1 \geq r_2 \geq \cdots \geq r_n$. We generate the parameter α from a uniform distribution over the interval $[\gamma_1,\gamma_2]$, where γ_1,γ_2 are parameters we vary. To come up with the parameter $\boldsymbol{\mu}\in\mathbb{R}^n$, we generate ξ_i from the uniform distribution over the interval $[1,10]$ for each $i\in\mathcal{N}$. For all $i\in\mathcal{N}$, we set $\mu_i = \log(\delta\xi_i)$ or, equivalently, $e^{\mu_i}=\delta\xi_i$, where δ is the scaling parameter chosen so that the probability of choosing the no-purchase option under the full assortment is p_0, and p_0 is another parameter we vary. It follows from Theorem 2.3 that δ satisfies the equation $1/\left[g\big(\sum_{\ell\in\mathcal{N}}\delta\xi_\ell\big)\right]^{1/(1+\alpha)} = p_0$. After generating $\boldsymbol{\mu}$, we process them to come up with two problem classes. In the first problem class, we leave $\{\mu_1,\ldots,\mu_n\}$ untouched. In the second problem class, we reindex $\{\mu_1,\ldots,\mu_n\}$ so that $\mu_1\leq\mu_2\leq\cdots\leq\mu_n$. Recall that $r_1\geq r_2\geq\cdots\geq r_n$; in the second problem class, the products with larger revenue have smaller preference weights, so expensive products are less attractive. We refer to the first and second problem class, respectively, as “U” and “O”, where “U” stands for unordered and “O” stands for ordered.

Letting T denote the ordering of the preference, we vary $p_0\in\{0.1,0.3,0.5\}$, $[\gamma_1,\gamma_2]\in\{[-0.9,0],[-0.9,-0.5],[-0.5,0]\}$, and $T\in\{U,O\}$, giving a total of 18 parameter configurations. For each parameter configuration, we generate 10 test problems.

Computational Results: We execute our FPTAS with $\epsilon=0.5/6$ to obtain a solution to the unconstrained assortment optimization problem with a theoretical performance guarantee of $1-6\epsilon=1/2$ (Theorem 3.5). Table 2 shows our computational results. The first column shows the parameter configuration $(p_0,[\gamma_1,\gamma_2])$. The rest of the table consists of two blocks of three columns. The first and second blocks correspond, respectively, to problems with unordered preference weights ($T=\mathrm{U}$) and problems with ordered preference weights ($T=\mathrm{O}$). In each block, the three columns show the average, maximum, and standard deviation of the percent gap between the upper bound on the optimal expected revenue and the expected revenue from the solution produced by our

Param. Config. $(p_0, [\gamma_1, \gamma_2])$	Unordered ($T = \mathrm{U}$) Avg. Gap	Max. Gap	Stdev. Gap	Ordered ($T = \mathrm{O}$) Avg. Gap	Max. Gap	Stdev. Gap
(0.1, [-0.5, 0.0])	0.22%	0.75%	0.23%	0.28%	1.61%	0.49%
(0.1, [-0.9, -0.5])	0.34%	1.24%	0.40%	0.53%	2.30%	0.73%
(0.1, [-0.9, 0.0])	0.22%	0.88%	0.28%	0.49%	1.79%	0.66%
(0.3, [-0.5, 0.0])	0.28%	0.69%	0.19%	0.38%	1.50%	0.52%
(0.3, [-0.9, -0.5])	0.34%	0.92%	0.35%	0.28%	0.90%	0.35%
(0.3, [-0.9, 0.0])	0.43%	1.14%	0.37%	0.58%	2.17%	0.86%
(0.5, [-0.5, 0.0])	0.43%	0.79%	0.24%	0.39%	1.85%	0.57%
(0.5, [-0.9, -0.5])	0.18%	0.68%	0.24%	0.02%	0.16%	0.05%
(0.5, [-0.9, 0.0])	0.31%	1.04%	0.41%	0.21%	0.74%	0.27%

Table 2 Performance of our FPTAS for unconstrained assortment optimization when $\alpha \in (-1, 0)$.

FPTAS. The average, maximum, and standard deviation are computed over 10 problem instances in each parameter configuration. Over all of our test problems, the average gap is 0.33% and the maximum gap is 2.30%. Problems with ordered preference weights have only slightly larger average and maximum gaps. The performance of our FPTAS is substantially stronger than the theoretical guarantee of 1/2. The average running time of our FPTAS is 104.52 seconds per instance using Python 3.9 on a Macbook Pro with 16 GB RAM.

7. Conclusion

We propose an endogenous context-dependent multinomial logit model. Our Context-MNL model parsimoniously captures the impact of assortment on the utility of each product through an extra parameter that measures the impact of the deviation of the product's intrinsic utility from the expected maximum utility among all the alternatives in the assortment. When tested on real data from Expedia, the model offers substantial improvements in terms of the goodness of fit and prediction accuracy. The model can be efficiently calibrated because maximum likelihood estimation is tractable. Assortment and price optimization under the model also admit efficient solution methods. Incorporating the Context-MNL model into network revenue management problems represents an exciting direction for future research. It would also be interesting to explore whether we can generalize other choice models to endogenously capture the assortment effect on the utility of each product and whether such extensions lead to tractable choice models.

References

Aouad, A., V. Farias, R. Levi. 2021. Assortment optimization under consider-then-choose choice models. *Management Science* **67**(6) 3368–3386.

Aydin, G., J. K. Ryan. 2000. Product line selection and pricing under the multinomial logit choice model. *Proceedings of the 2000 MSOM Conference*.

Ben-Akiva, M. E., S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA.

Bettman, J. R., M. F. Luce, J. W. Payne. 1998. Constructive consumer choice processes. *Journal of Consumer Research* **25**(3) 187–217.

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

Boyd, S. P., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Bront, J. J. M., I. Mendez Diaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57**(3) 769–784.

Cao, Y., A. J. Kleywegt, H. Wang. 2022. Network revenue management under a spiked multinomial logit choice model. *Operations Research* **70**(4) 2237–2253.

Chakravarti, D., J. G. Lynch. 1983. A framework for exploring context effects on consumer judgment and choice. R. P. Bagozzi, A. M. Tybout, eds., *Advances in Consumer Research*, vol. 10. Association for Consumer Research, Ann Arbor, MI, 289–297.

Chen, X., P. Hu, Z. Hu. 2017. Efficient algorithms for the dynamic pricing problem with reference price effect. *Management Science* **63**(12) 4389–4408.

Davis, J. M., G. Gallego, H. Topaloglu. 2014. Assortment optimization under variants of the nested logit model. *Operations Research* **62**(2) 250–273.

Désir, A., V. Goyal, S. Jagabathula, D. Segev. 2016. Assortment optimization under the Mallows model. *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS '16, Curran Associates Inc., USA, 4707–4715.

Désir, A., V. Goyal, J. Zhang. 2022. Capacitated assortment optimization: Hardness and approximation. *Operations Research* **70**(2) 893–904.

El Housni, O., H. Topaloglu. 2023. Joint assortment optimization and customization under a mixture of multinomial logit models: On the value of personalized assortments. *Operations Research* **71**(4) 1197–1215.

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

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. CORC Technical Report TR-2004-01.

Gallego, G., R. Ratliff, S. Shebalov. 2015. A general attraction model and sales-based linear programming formulation for network revenue management under customer choice. *Operations Research* **63**(1) 212–232.

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

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

Gao, P., Y. Ma, N. Chen, G. Gallego, A. Li, P. Rusmevichientong, H. Topaloglu. 2021. Assortment optimization and pricing under the multinomial logit model with impatient customers: Sequential recommendation and selection. *Operations Research* **69**(5) 1509–1532.

Garey, M. R., D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, NY.

Goulden, C. H. 1939. *Methods of Statistical Analysis*. John Wiley & Sons, New York, NY.

Gumbel, E. J. 2004. *Statistics of Extremes*. Dover Publications, Meneola, NY.

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

Hu, Z., X. Chen, P. Hu. 2016. Dynamic pricing with gain-seeking reference price effects. *Operations Research* **64**(1) 150–157.

Jagabathula, S. 2016. Assortment optimization under general choice. Tech. rep., NYU, New York, NY.

Kaggle. 2013. Personalize Expedia hotel searches. Last checked: July 2, 2023. URL `https://www.kaggle.com/c/expedia-personalized-sort`.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–291.

Kamakura, W. A., R. K. Srivastava. 1984. Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research* **21**(4) 420–434.

Kivetz, R., O. Netzer, V. Srinivasan. 2004. Alternative models for capturing the compromise effect. *Journal of Marketing Research* **41**(3) 237–257.

Kőszegi, B., M. Rabin. 2006. A model of reference-dependent preferences. *The Quarterly Journal of Economics* **121**(4) 1133–1165.

Li, H., S. C. Graves. 2012. Pricing decisions during inter-generational product transition. *Production and Operations Management* **21**(1) 14–28.

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

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

Mendez-Diaz, I., J. J. M. Bront, G. Vulcano, P. Zabala. 2014. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* **164**(1) 246–263.

Orhun, A. Y. 2009. Optimal product line design when consumers exhibit choice set-dependent preferences. *Marketing Science* **28**(5) 868–886.

Payne, J. W. 1982. Contingent decision behavior. *Psychological Bulletin* **92**(2) 382.

Payne, J. W., J. R. Bettman, E. J. Johnson. 1992. Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology* **43**(1) 87–131.

Popescu, I., Y. Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* **55**(3) 413–429.

Prelec, D., B. Wernerfelt, F. Zettelmeyer. 1997. The role of inference in context effects: Inferring what you want from what is available. *Journal of Consumer Research* **24**(1) 118–125.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. 2007. *Numerical Recipes: The Art of Scientific Computing*. 3rd ed. Cambridge University Press, Cambridge, UK.

Ratneshwar, S., A. D. Shocker, D. W. Stewart. 1987. Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research* **13**(4) 520–533.

Rooderkerk, R. P., H. J. Van Heerde, T. H. A. Bijmolt. 2011. Incorporating context effects into a choice model. *Journal of Marketing Research* **48**(4) 767–780.

Rusmevichientong, P., Z.-J. M. Shen, D. B. Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research* **58**(6) 1666–1680.

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

Severini, T. A. 2000. *Likelihood Methods in Statistics*. Oxford University Press, Cambridge, UK.

Simonson, I. 1999. The effect of product assortment on buyer preferences. *Journal of Retailing* **75**(3) 347–370.

Song, J.-S., Z. Xue. 2007. Demand management and inventory control for substitutable products. Tech. rep., Duke University, Durham, NC.

Sumida, M., G. Gallego, P. Rusmevichientong, H. Topaloglu, J. M. Davis. 2021. Revenue-utility tradeoff in assortment optimization under the multinomial logit model with totally unimodular constraints. *Management Science* **67**(5) 2845–2869.

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

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK.

Tversky, A., I. Simonson. 1993. Context-dependent preferences. *Management Science* **39**(10) 1179–1189.

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

Wang, R. 2018. When prospect theory meets consumer choice models: Assortment and pricing management with reference prices. *Manufacturing & Service Operations Management* **20**(3) 583–600.

Wang, R., Z. Wang. 2017. Consumer choice models with endogenous network effects. *Management Science* **63**(11) 3944–3960.

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

Online Appendix

Assortment and Price Optimization under an Endogenous Context-Dependent Multinomial Logit Model

Appendix A: Normalizing μ_0 to Zero

Consider a Context-MNL model with parameters $\alpha > -1$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\mu_0 \neq 0$. For each assortment S, let $\tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)$ denote the preference weight of product i when it is offered in assortment S, where we explicitly show the dependence on α, $\boldsymbol{\mu}$, and μ_0. Using the same argument as in Theorem 2.2, the utility of product i when it is offered within an assortment S is given by $\mathsf{Util}_i(S) = \log \tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) + \epsilon_i$, where the preference weights $\big(\tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) : i \in S\big)$ satisfy the following system of equations:

$$\tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) = e^{\mu_i - \alpha \log \left[e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)\right]} \qquad \forall\ i \in S\ ,$$

and the same argument also shows that $\phi_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) = \tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) / \big(e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)\big)$ for all $i \in S$, and $\phi_0(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) = 1 - \sum_{i \in S} \phi_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)$. The following proposition shows the relationship between $\tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)$ in the above equation and the "normalized" preference weight in Equation (2) where we set $\mu_0 = 0$.

Proposition A.1 (Normalizing μ_0 to Zero) *For each $\alpha > -1$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$, and $\mu_0 \in \mathbb{R}$, if $\bar{\boldsymbol{\mu}} = (\mu_1 - (1+\alpha)\mu_0, \ldots, \mu_n - (1+\alpha)\mu_0)$, then for all $S \subseteq \mathcal{N}$ and $i \in S$, $\tilde{v}_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) e^{-\mu_0} = v_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}})$ and $\phi_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) = \phi_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}})$, where $(v_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}}) : i \in S)$ is defined in Equation (2) and $\phi_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}})$ is given in Theorem 2.2.*

Proof: Fix $\alpha > -1$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\mu_0 \in \mathbb{R}$. Consider an arbitrary S. For ease of exposition, we will write $\tilde{v}_i = v_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0)$ and $\bar{v}_i = v_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}})$. By definition of $\tilde{v}_i$, we have $\tilde{v}_i = e^{\mu_i - \alpha \log\left(e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell\right)}$ for all $i \in S$, which implies that

$$\tilde{v}_i e^{-\mu_0} = e^{\mu_i - \mu_0 - \alpha \log\left(e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell\right)} = e^{\mu_i - (1+\alpha)\mu_0 - \alpha \log \left[e^{-\mu_0} \times \left(e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell\right)\right]} \ = \ e^{\bar{\mu}_i - \alpha \log \left(1 + \sum_{\ell \in S} \tilde{v}_\ell e^{-\mu_0}\right)}\ ,$$

and by Theorem 2.2, the above system of equation has a unique solution, so $\tilde{v}_i e^{-\mu_0} = \bar{v}_i$. Finally, for all $i \in S$, $\phi_i(S \,|\, \alpha, \boldsymbol{\mu}, \mu_0) = \frac{\tilde{v}_i}{e^{\mu_0} + \sum_{\ell \in S} \tilde{v}_\ell} = \frac{\tilde{v}_i e^{-\mu_0}}{1 + \sum_{\ell \in S} \tilde{v}_\ell e^{-\mu_0}} = \frac{\bar{v}_i}{1 + \sum_{\ell \in S} \bar{v}_\ell} = \phi_i(S \,|\, \alpha, \bar{\boldsymbol{\mu}})$. ■

Appendix B: Properties of f and g and Proof of Theorem 2.5

In this section, we prove properties of the functions f and g defined in Equation (3). The functions f and g were defined for each α, and we will need to understand how these functions change as the parameter α changes. So, we will write f_α and g_α to highlight the dependence on α. Recall from Equation (3) that $f_\alpha : [1,\infty) \to \mathbb{R}_+$ and $g_\alpha : \mathbb{R}_+ \to [1,\infty)$ are defined by:

$$f_\alpha(x) = x - x^{\alpha/(1+\alpha)} \quad \forall\, x \in [1,\infty) \qquad \text{and} \qquad g_\alpha(y) = f_\alpha^{-1}(y) \quad \forall\, y \in \mathbb{R}_+ \;.$$

The first lemma provides expressions for the first derivatives of f_α and g_α.

Lemma B.1 (First Derivatives) *For all $\alpha > -1$, $x \geq 1$ and $y \geq 0$,*

$$\begin{aligned}
\frac{\partial}{\partial x} f_\alpha(x) &= 1 - \frac{\alpha}{1+\alpha} x^{-1/(1+\alpha)} && \text{and} & \frac{\partial}{\partial \alpha} f_\alpha(x) &= -\frac{x^{\alpha/(1+\alpha)} \log x}{(1+\alpha)^2} \\
\frac{\partial}{\partial y} g_\alpha(y) &= \left[\frac{\partial}{\partial x} f_\alpha(x) \Big|_{x=g_\alpha(y)} \right]^{-1} && \text{and} & \frac{\partial}{\partial \alpha} g_\alpha(y) &= \frac{g_\alpha(y) \log g_\alpha(y)}{(1+\alpha)^2 \left[g_\alpha(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha} \right]} \;.
\end{aligned}$$

Proof: The derivatives of f_α with respect to x and α follow from calculus and we omit the details. The derivative of g_α with respect to y follows from the Inverse Function Theorem. We will thus focus on the partial derivative of g_α with respect to α. Because $f_\alpha(g_\alpha(y)) = y$, taking the derivative with respect to α on both sides, it follows from the Chain Rule that

$$\left(\frac{\partial}{\partial x} f_\alpha(x) \Big|_{x=g_\alpha(y)} \times \frac{\partial}{\partial \alpha} g_\alpha(y) \right) + \frac{\partial}{\partial \alpha} f_\alpha(g_\alpha(y)) = 0 \Leftrightarrow \frac{\partial}{\partial \alpha} g_\alpha(y) = -\frac{\partial}{\partial \alpha} f_\alpha(g_\alpha(y)) \Big/ \frac{\partial}{\partial x} f_\alpha(x) \Big|_{x=g_\alpha(y)} \;.$$

It then follows from the partial derivative of f_α that

$$\frac{\partial}{\partial \alpha} g_\alpha(y) = -\frac{g_\alpha(y)^{\alpha/(1+\alpha)} \log g_\alpha(y)}{(1+\alpha)^2 \left[1 - \frac{\alpha}{1+\alpha} g_\alpha(y)^{-1/(1+\alpha)} \right]} = -\frac{g_\alpha(y) \log g_\alpha(y)}{(1+\alpha)^2 \left[g_\alpha(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha} \right]} \,,$$

which is the desired result. ■

The following corollaries follow immediately from the above lemma.

Corollary B.2 (Monotonicity) *For each $\alpha > -1$, $f_\alpha(x)$ is strictly increasing in x and unbounded, and $g_\alpha(y)$ is strictly increasing in y and unbounded.*

Corollary B.3 (Convexity) *For each $\alpha \geq 0$, $f_\alpha(x)$ is convex in x and $g_\alpha(y)$ is concave in y. For each $-1 < \alpha < 0$, $f_\alpha(x)$ is concave in x and $g_\alpha(y)$ is convex in y.*

Here is a proof of Theorem 2.5.

Proof of Theorem 2.5: Fix $\boldsymbol{\mu} \in \mathbb{R}^n$, $S \subseteq \mathcal{N}$, $k \in S$, and $i \in S \setminus \{k\}$. Let $y_0 = \sum_{\ell \in S \setminus \{k\}} e^{\mu_\ell}$ and $y_1 = e^{\mu_k}$. It follows from Theorem 2.3 that

$$\frac{\phi_i(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_i(S \,|\, \alpha, \boldsymbol{\mu})} \Big/ \frac{\phi_0(S \setminus \{k\} \,|\, \alpha, \boldsymbol{\mu})}{\phi_0(S \,|\, \alpha, \boldsymbol{\mu})} = \frac{g_\alpha(y_1 + y_0)}{g_\alpha(y_0)} \Big/ \left(\frac{g_\alpha(y_1 + y_0)}{g_\alpha(y_0)}\right)^{1/(1+\alpha)} = \left(\frac{g_\alpha(y_1 + y_0)}{g_\alpha(y_0)}\right)^{\alpha/(1+\alpha)},$$

and thus, to prove the desired result, it suffices to show that the function $H(\alpha) \equiv \frac{\alpha}{1+\alpha} \log\left(\frac{g_\alpha(y_1+y_0)}{g_\alpha(y_0)}\right)$ is strictly increasing for $\alpha > -1$. We will do this by taking the derivative with respect to α. Using the product rule and the derivative formula from Lemma B.1, we have that

$$\begin{aligned}
H'(\alpha) &= \frac{\alpha}{1+\alpha}\left(\frac{\frac{\partial}{\partial \alpha} g_\alpha(y_1 + y_0)}{g_\alpha(y_1 + y_0)} - \frac{\frac{\partial}{\partial \alpha} g_\alpha(y_0)}{g_\alpha(y_0)}\right) + \frac{1}{(1+\alpha)^2} \log\left(\frac{g_\alpha(y_1 + y_0)}{g_\alpha(y_0)}\right) \\
&= \frac{\alpha}{1+\alpha}\left(\frac{\log g_\alpha(y_1 + y_0)}{(1+\alpha)^2\left[g_\alpha(y_1 + y_0)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}\right]} - \frac{\log g_\alpha(y_0)}{(1+\alpha)^2\left[g_\alpha(y_0)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}\right]}\right) \\
&\qquad + \frac{1}{(1+\alpha)^2} \log\left(\frac{g_\alpha(y_1 + y_0)}{g_\alpha(y_0)}\right) \\
&= \frac{1}{(1+\alpha)^2}\left(\frac{\alpha/(1+\alpha)}{g_\alpha(y_1 + y_0)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}} + 1\right) \log g_\alpha(y_1 + y_0) \\
&\qquad - \frac{1}{(1+\alpha)^2}\left(\frac{\alpha/(1+\alpha)}{g_\alpha(y_0)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}} + 1\right) \log g_\alpha(y_0) \\
&= \frac{1}{(1+\alpha)^2}\Big(G(g_\alpha(y_1 + y_0)) - G(g_\alpha(y_0))\Big),
\end{aligned}$$

where the function $G : [1, \infty) \to \mathbb{R}$ is defined as follows: letting $a = \alpha/(1+\alpha)$, for all $x \geq 1$, $G(x) = \left(\frac{a}{x^{1-a} - a} + 1\right) \log x$. To show that $H'(\alpha) > 0$, it suffices to show that G is strictly increasing in x because $g_\alpha(y_1 + y_0) > g_\alpha(y_0)$. Note that $a < 1$ and

$$\begin{aligned}
G'(x) &= \frac{1}{x}\left(\frac{a}{x^{1-a} - a} + 1\right) - \frac{a(1-a)x^{-a} \log x}{(x^{1-a} - a)^2} = \frac{a(x^{1-a} - a) + (x^{1-a} - a)^2 - a(1-a)x^{1-a} \log x}{x(x^{1-a} - a)^2} \\
&= \frac{ax^{1-a} - a^2 + x^{2(1-a)} - 2ax^{1-a} + a^2 - a(1-a)x^{1-a} \log x}{x(x^{1-a} - a)^2} \\
&= \frac{x^{1-a}\left(x^{1-a} - a - a(1-a)\log x\right)}{x(x^{1-a} - a)^2} = \frac{x^{1-a} - a - a(1-a)\log x}{x^a (x^{1-a} - a)^2},
\end{aligned}$$

and observe that the denominator is always nonnegative. We will now show that the numerator is strictly positive by showing that $x^{1-a} > a + a(1-a)\log x$ for all $x \geq 1$. The inequality is trivially true if $a \leq 0$, so consider the case where $0 < a < 1$. The inequality is true at $x = 1$ because $1 > a$. Moreover, the derivative of the lefthand side is $\frac{1-a}{x^a}$ and the the derivative of the righthand side is $\frac{a(1-a)}{x}$. Because $x \geq 1$ and $0 < a < 1$, we have $\frac{1-a}{x^a} > \frac{a(1-a)}{x} > 0$. This shows that $x^{1-a} > a + a(1-a)\log x$ for all $x \geq 1$, which implies that $G'(x) > 0$ for all $x \geq 1$, completing the proof. ∎

Appendix C: Proof of Theorem 3.3

The proof of Theorem 3.3 makes use of a series of lemmas. In this section, we are interested in assortment optimization problem where $\alpha = -1/2$ and there is no cardinality constraint ($K = n$), so $Z^* = \max_{S \subseteq \mathcal{N}} \sum_{i \in S} r_i \phi_i(S) = \max_{S \subseteq \mathcal{N}} \left(\sum_{i \in S} r_i v_i(S) / \left(1 + \sum_{i \in S} v_i(S)\right)\right)$, where the last equality follows from Theorem 2.2. The first lemma establishes an important property of an optimal solution.

Lemma C.1 *For each $\alpha \in (-1, 0)$, there is an optimal assortment that contains the product with the highest revenue.*

Proof: Let $h^* \in \mathcal{N}$ denote the product with the highest revenue; that is, $r_{h^*} \geq r_i$ for all $i \in \mathcal{N}$. For each assortment $S \subseteq \mathcal{N}$ such that $h^* \notin S$, we will show that adding h^* to S does not decrease the revenue. Because $1 = \phi_0(S \cup \{h^*\}) + \phi_{h^*}(S \cup \{h^*\}) + \sum_{i \in S} \phi_i(S \cup \{h^*\})$, we have

$$\begin{aligned}
\sum_{i \in S \cup \{h^*\}} r_i \phi_i(S \cup \{h^*\}) &= r_{h^*}\Big(1 - \phi_0(S \cup \{h^*\}) - \sum_{i \in S} \phi_i(S \cup \{h^*\})\Big) + \sum_{i \in S} r_i \phi_i(S \cup \{h^*\}) \\
&= r_{h^*}\Big(1 - \phi_0(S \cup \{h^*\})\Big) + \sum_{i \in S} (r_i - r_{h^*}) \phi_i(S \cup \{h^*\}) \\
&\geq r_{h^*}\Big(1 - \phi_0(S)\Big) + \sum_{i \in S} (r_i - r_{h^*}) \phi_i(S) \;=\; \sum_{i \in S} r_i \phi_i(S) \;,
\end{aligned}$$

where the inequality follows from the substitutability of the Context-MNL model in Corollary 2.4, so $\phi_0(S \cup \{h^*\}) \leq \phi_0(S)$, which implies that $r_{h^*}\big(1 - \phi_0(S \cup \{h^*\})\big) \;\geq\; r_{h^*}\Big(1 - \phi_0(S)\Big)$. By substitutability, we also have $\phi_i(S \cup \{h^*\}) \leq \phi_i(S)$ and because $r_i - r_{h^*} \leq 0$, it follows that $(r_i - r_{h^*})\phi_i(S \cup \{h^*\}) \geq (r_i - r_{h^*})\phi_i(S)$ for all $i \in S$. This completes the proof. ■

The next lemma focuses on a one-dimensional function that will show up in our proof.

Lemma C.2 *Let A and B be two positive real numbers such that $B < A$ and $A^2 < B^2 + 4$. Consider a function $q : \mathbb{R}_+ \to \mathbb{R}_+$ defined as follows: for each $x \geq 0$,*

$$q(x) = \frac{A + x}{B + x + \sqrt{4 + (B + x)^2}} \;.$$

Then, q has a unique maximizer at $x^* = \frac{2}{A-B} - \frac{A+B}{2} = \frac{4+B^2-A^2}{2(A-B)}$ and $q(x^*) = \frac{4+(A-B)^2}{8}$.

Proof: By definition,

$$\begin{aligned}
q'(x) &= \frac{\left[B + x + \sqrt{4 + (B+x)^2}\right] - (A + x)\left[1 + \frac{B+x}{\sqrt{4+(B+x)^2}}\right]}{\left[B + x + \sqrt{4 + (B+x)^2}\right]^2} \\
&= \frac{\sqrt{4 + (B+x)^2}\left[B + x + \sqrt{4 + (B+x)^2}\right] - (A + x)\left[B + x + \sqrt{4 + (B+x)^2}\right]}{\sqrt{4 + (B+x)^2}\left[B + x + \sqrt{4 + (B+x)^2}\right]^2}
\end{aligned}$$

$$= \frac{\sqrt{4+(B+x)^2}-(A+x)}{\sqrt{4+(B+x)^2}\left[B+x+\sqrt{4+(B+x)^2}\right]} .$$

Because the denominator is strictly positive, the sign of $q'(x)$ is determined by the numerator. Note that

$$\begin{aligned} \sqrt{4+(B+x)^2}-(A+x)\geq 0 \quad &\Leftrightarrow \quad 4+(B+x)^2\geq (A+x)^2 \quad \Leftrightarrow \quad 4+B^2+2Bx\geq A^2+2Ax \\ &\Leftrightarrow \quad x\leq \frac{4+B^2-A^2}{2(A-B)}=x^* .\end{aligned}$$

This shows that $q'(x)>0$ for all $x<x^*$, $q'(x^*)=0$, and $q'(x)<0$ for all $x>x^*$. Therefore, q achieves a unique maximum over the nonnegative real line at x^*. Moreover, note that

$$\begin{aligned} A+x^* &= A+\frac{4+B^2-A^2}{2(A-B)} = \frac{4+(A-B)^2}{2(A-B)} \\ B+x^* &= B+\frac{4+B^2-A^2}{2(A-B)} = \frac{4-(A-B)^2}{2(A-B)} \\ \sqrt{4+(B+x^*)^2} &= \sqrt{4+\frac{\left[4-(A-B)^2\right]^2}{4(A-B)^2}} = \sqrt{\frac{\left[4+(A-B)^2\right]^2}{4(A-B)^2}} = \frac{4+(A-B)^2}{2(A-B)} \\ q(x^*) &= \frac{4+(A-B)^2}{4-(A-B)^2+4+(A-B)^2} = \frac{4+(A-B)^2}{8} ,\end{aligned}$$

which is the desired result. ■

We are now ready to give a proof of Theorem 3.3.

Proof: Fix $\alpha=-1/2$ and $K=n$. For ease of exposition, we will drop the subscript α from f_α and g_α, and write f and g. Recall that

$$Z^* = \max_{S\subseteq\mathcal{N}}\sum_{i\in S} r_i\phi_i(S) = \max_{S\subseteq\mathcal{N}}\frac{\sum_{i\in S} r_i e^{\mu_i}}{g\left(\sum_{i\in S} e^{\mu_i}\right)} = \max_{S\subseteq\mathcal{N}}\frac{\sum_{i\in S} 2\, r_i\, e^{\mu_i}}{\sum_{i\in S} e^{\mu_i} + \sqrt{4+\left(\sum_{i\in S} e^{\mu_i}\right)^2}} , \tag{9}$$

where the second equality follows from Theorem 2.3 and the last equality follows because for $\alpha=-1/2$, we have $f(x)=x-\frac{1}{x}$, which implies that $g(y)=(\,y+\sqrt{4+y^2}\,)/2$.

To show that Z^* is NP-hard, we will use a reduction from PARTITION, which is a well known NP-complete problem (Garey and Johnson 1979). Consider an arbitrary instance of the PARTITION problem given by

PARTITION

Inputs: A collection of n positive integers $w_1,w_2,\ldots,w_n$.

Question: Is there a subset $X\subseteq\{1,2,\ldots,n\}$ such that $\sum_{i\in X} w_i=\sum_{i\notin X} w_i$?

Let $T=\frac{1}{2}\sum_{i=1}^n w_i$. Note that $\sum_{i\in X} w_i=\sum_{i\notin X} w_i$ if and only if $\sum_{i\in X} w_i=T$. Thus, we can assume without loss of generality that T is a positive integer. Given an arbitrary instance of the

PARTITION problem, we will construct the following instance of the Context-MNL model. We will have $n+1$ products indexed by $1,2,\ldots,n+1$, and the parameters of these products are given as follows:

$$r_i = \begin{cases} \frac{1}{2} & \text{if} \quad i = 1,2,\ldots,n, \\ \frac{1}{2} \times \frac{3T+\frac{1}{4T}}{3T-\frac{1}{4T}} & \text{if} \quad i = n+1 \end{cases} \qquad \text{and} \qquad e^{\mu_i} = \begin{cases} w_i & \text{if} \quad i = 1,2,\ldots,n, \\ 3T - \frac{1}{4T} & \text{if} \quad i = n+1\ . \end{cases}$$

Because T is a positive integer, $3T - \frac{1}{4T} > 0$, so all the parameters are well defined. Let Z^* denote the maximum expected revenue under the above Context-MNL model. Letting $R^* \equiv \frac{1}{2} + \frac{1}{32T^2}$, to prove the theorem, it suffices to prove the following claim.

Claim: There is a subset $X \subseteq \{1,2,\ldots,n\}$ such that $\sum_{i\in X} w_i = T$ if and only if $Z^* \geq R^*$.

To prove this claim, let $q : \mathbb{R}_+ \to \mathbb{R}_+$ be defined by: for all $x \geq 0$,

$$q(x) = \frac{3T + \frac{1}{4T} + x}{3T - \frac{1}{4T} + x + \sqrt{4 + \left(3T - \frac{1}{4T} + x\right)^2}}\ .$$

Then, it follows from Equation (9) that

$$\begin{aligned}
Z^* &= \max_{S\subseteq\{1,2,\ldots,n,n+1\}} \frac{\sum_{i\in S} 2\,r_i\,e^{\mu_i}}{\sum_{i\in S} e^{\mu_i} + \sqrt{4 + \left(\sum_{i\in S} e^{\mu_i}\right)^2}} \\
&\overset{(a)}{=} \max_{S\subseteq\{1,2,\ldots,n\}} \frac{2\,r_{n+1}\,e^{\mu_{n+1}} + \sum_{i\in S} 2\,r_i\,e^{\mu_i}}{e^{\mu_{n+1}} + \sum_{i\in S} e^{\mu_i} + \sqrt{4 + \left(e^{\mu_{n+1}} + \sum_{i\in S} e^{\mu_i}\right)^2}} \\
&\overset{(b)}{=} \max_{S\subseteq\{1,2,\ldots,n\}} \frac{3T + \frac{1}{4T} + \sum_{i\in S} w_i}{3T - \frac{1}{4T} + \sum_{i\in S} w_i + \sqrt{4 + \left(3T - \frac{1}{4T} + \sum_{i\in S} w_i\right)^2}} \\
&\overset{(c)}{=} \max_{S\subseteq\{1,2,\ldots,n\}} q\left(\sum_{i\in S} w_i\right)\ ,
\end{aligned} \tag{10}$$

where (a) follows because $r_{n+1} = \frac{1}{2} \times \frac{3T+\frac{1}{4T}}{3T-\frac{1}{4T}} > \frac{1}{2} = r_i$ for all $i = 1,\ldots,n$, so product $n+1$ has the highest revenue, and by Lemma C.1, there is an optimal assortment that contains product $n+1$. Equality (b) follows from our construction of the parameters so that

$$2r_{n+1}e^{\mu_{n+1}} = 3T + \frac{1}{4T}, \quad e^{\mu_{n+1}} = 3T - \frac{1}{4T}, \quad 2r_ie^{\mu_i} = w_i \ \ \forall\, i = 1,\ldots,n, \quad \text{and} \quad e^{\mu_i} = w_i \ \ \forall\, i = 1,\ldots,n.$$

The last equality (c) follows from the definition of $q(\cdot)$.

Let $A = 3T + \frac{1}{4T}$ and $B = 3T - \frac{1}{4T}$. By definition, we have that $q(x) = \frac{A+x}{B+x+\sqrt{4+(B+x)^2}}$ for all $x \geq 0$. Note that $B < A$ and $A^2 = 9T^2 + \frac{6}{4} + \frac{1}{16T^2} < 4 + 9T^2 - \frac{6}{4} + \frac{1}{16T^2} = 4 + B^2$ because $\frac{6}{4} < 4 - \frac{6}{4}$. Therefore, the hypothesis of Lemma C.2 is satisfied. Note that $A - B = \frac{1}{2T}$ and $A + B = 6T$. By Lemma C.2, q has a unique maximizer at $x^* = \frac{2}{A-B} - \frac{A+B}{2} = 4T - 3T = T$ and

$$q(x^*) = \max_{x\geq 0} q(x) = \frac{4 + (A-B)^2}{8} = \frac{4 + \left(\frac{1}{2T}\right)^2}{8} = \frac{1}{2} + \frac{1}{32\,T^2} = R^*\ .$$

From Equation (10), we have $Z^* = \max_{S\subseteq\{1,2,\ldots,n\}} \; q\Big(\sum_{i\in S} w_i\Big)$. Therefore, $Z^* \geq R^*$ if and only if there exists a subset $S \in \{1,2,\ldots,n\}$ such that $\sum_{i\in S} w_i = T$, which is the desired claim. ■

Appendix D: Proof of Lemma 3.4

Fix $\epsilon > 0$ and $(\nu,\xi) \in \mathbb{R}_+ \times \mathbb{R}_+$. For each $i \in \mathcal{N}$, let $\tilde{r}_i = \left\lfloor \dfrac{r_i\, e^{\mu_i}}{\epsilon\,\nu/n} \right\rfloor$ and $\tilde{u}_i = \left\lceil \dfrac{e^{\mu_i}}{\epsilon\,\xi/n} \right\rceil$. Noting that $\tilde{r}_i \in \mathbb{Z}_+$ and $\tilde{u}_i \in \mathbb{Z}_+$ for all i and letting $G = \lfloor \frac{n}{\epsilon} \rfloor - n$ and $H = \lceil \frac{n}{\epsilon} \rceil + n$, we will develop a dynamic programming algorithm for solving the following optimization problem:

$$Y^* \;\equiv\; \min_{S\subseteq\mathcal{N}} \left\{ |S| : |S| \leq K,\; \sum_{i\in S} \tilde{r}_i \geq G, \text{ and } \sum_{i\in S} \tilde{u}_i \leq H \right\} ,$$

and if there is no feasible solution, we set $Y^* = \infty$. For each $(g,h,p) \in \{0,1,...,G\} \times \{0,1,...,H\} \times \{1,2,\ldots,n\}$, let $F(g,h,p)$ denote the minimum cardinality among all sets $S \subseteq \{1,2,\ldots,p\}$ such that $\sum_{\ell\in S} \tilde{r}_\ell \geq g$ and $\sum_{\ell\in S} \tilde{u}_\ell \leq h$; that is, $F(g,h,p) = \min_{S\subseteq\{1,2,\ldots,p\}}\{|S| : \sum_{\ell\in S} \tilde{r}_\ell \geq g \text{ and } \sum_{\ell\in S} \tilde{u}_\ell \leq h\}$. We can compute $F(i,j,p)$ using the following dynamic programming recursion:

$$\begin{aligned}
F(g,h,1) \;&=\; \begin{cases} 1 & \text{if} \quad g \leq \tilde{r}_1 \text{ and } \tilde{u}_1 \leq h \\ 0 & \text{if} \quad g = 0 \\ \infty & \text{otherwise} \end{cases} \\
F(g,h,p+1) \;&=\; \min\left\{F(g,h,p), 1 + F(g - \tilde{r}_{p+1},\; h - \tilde{u}_{p+1},\; p)\right\} ,
\end{aligned}$$

with boundary conditions that for each $g < 0$, $F(g,h,p) = F(0,h,p)$, and for each $h < 0$, $F(\cdot,h,\cdot) = \infty$. It is easy to check that $F(G,H,n) = Y^*$.

If $F(G,H,n) \leq K$, then output the corresponding subset; otherwise, if $F(G,H,n) > K$, then output an empty set. Note that computing $F(G,H,n)$ requires $\mathcal{O}(GHn) = \mathcal{O}(n^3/\epsilon^2)$ operations and it is independent of ν and ξ.

Recall that $\mathcal{F}_\epsilon(\nu,\xi) = \{S \subseteq \mathcal{N} : |S| \leq K,\; \sum_{i\in S} r_i e^{\mu_i} \geq \nu,\; \sum_{i\in S} e^{\mu_i} \leq \xi\}$. To prove the first part of Lemma 3.4, assume that $\mathcal{F}_\epsilon(\nu,\xi) \neq \varnothing$. Pick an arbitrary $S \in \mathcal{F}_\epsilon(\nu,\xi)$. For each $S \in \mathcal{F}_\epsilon(\nu,\xi)$, we have that

$$\begin{aligned}
\sum_{\ell\in S} \tilde{r}_\ell \;&=\; \sum_{\ell\in S} \left\lfloor \frac{r_\ell e^{\mu_\ell}}{\epsilon\,\nu/n} \right\rfloor \;\geq\; \sum_{\ell\in S} \frac{r_\ell e^{\mu_\ell}}{\epsilon\,\nu/n} - n \;\geq\; \left\lfloor \sum_{\ell\in S} \frac{r_\ell e^{\mu_\ell}}{\epsilon\,\nu/n} \right\rfloor - n \;\overset{(a)}{\geq}\; \left\lfloor \frac{n}{\epsilon} \right\rfloor - n \;=\; G \\
\sum_{\ell\in S} \tilde{u}_\ell \;&=\; \sum_{\ell\in S} \left\lceil \frac{e^{\mu_\ell}}{\epsilon\,\xi/n} \right\rceil \;\leq\; \sum_{\ell\in S} \frac{e^{\mu_\ell}}{\epsilon\,\xi/n} + n \;\leq\; \left\lceil \sum_{\ell\in S} \frac{e^{\mu_\ell}}{\epsilon\,\xi/n} \right\rceil + n \;\overset{(b)}{\leq}\; \left\lceil \frac{n}{\epsilon} \right\rceil + n \;=\; H ,
\end{aligned}$$

where inequalities (a) and (b) follow because $S \in \mathcal{F}_\epsilon(\nu,\xi)$. The above chains of inequalities imply that $F(G,H,n) \leq K$. Because $F(G,H,n) \leq K$, let $\tilde{S}$ be the corresponding subset determined by the dynamic program. It follows that $|\tilde{S}| \le K$, $\sum_{\ell \in \tilde{S}} \tilde{r}_\ell \ge G$, and $\sum_{\ell \in \tilde{S}} \tilde{u}_\ell \le H$, and thus, using the definition of $\tilde{r}_\ell$ and $\tilde{u}_\ell$, we have

$$\sum_{\ell \in \tilde{S}} \frac{r_\ell e^{\mu_\ell}}{\epsilon \nu / n} \;\ge\; \sum_{\ell \in \tilde{S}} \left\lfloor \frac{r_\ell e^{\mu_\ell}}{\epsilon \nu / n} \right\rfloor \;=\; \sum_{\ell \in \tilde{S}} \tilde{r}_\ell \;\ge\; G \quad \text{and} \quad \sum_{\ell \in \tilde{S}} \frac{e^{\mu_\ell}}{\epsilon \xi / n} \;\le\; \sum_{\ell \in \tilde{S}} \left\lceil \frac{e^{\mu_\ell}}{\epsilon \xi / n} \right\rceil \;=\; \sum_{\ell \in \tilde{S}} \tilde{u}_\ell \;\le\; H \;,$$

which implies that

$$\sum_{\ell \in \tilde{S}} r_\ell e^{\mu_\ell} \ge \frac{\epsilon \nu}{n} G \ge \frac{\epsilon \nu}{n} \left(\frac{n}{\epsilon} - 1 - n \right) \ge \nu(1 - 2\epsilon) \text{ and } \sum_{\ell \in \tilde{S}} e^{\mu_\ell} \le \frac{\epsilon \xi}{n} H \le \frac{\epsilon \xi}{n} \left(\frac{n}{\epsilon} + 1 + n \right) \le \xi(1 + 2\epsilon) \;,$$

which is the desired result.

Appendix E: Proof of Theorem 4.1(a)

It follows from the definition of $Q(x)$ that

$$\begin{aligned}
Q'(x) \;&=\; -\left[\frac{f(x)}{x} \frac{A}{f(x)} \frac{f'(x)}{A} \;+\; \log\left(\frac{f(x)}{A}\right) \left(\frac{x f'(x) - f(x)}{x^2} \right) \right] \\
&=\; -\left[\frac{f'(x)}{x} \;+\; \log\left(\frac{f(x)}{A}\right) \left(\frac{x f'(x) - f(x)}{x^2} \right) \right] \\
&\overset{(a)}{=}\; -\left[\frac{1}{x} - \frac{\alpha/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \;+\; \log\left(\frac{f(x)}{A}\right) \frac{1/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \right] \\
&=\; -\frac{1/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \left[(1+\alpha) x^{1/(1+\alpha)} - \Big(\alpha \;+\; \log A - \log f(x) \Big) \right] \\
&=\; \frac{1/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \left[\Big(\alpha \;+\; \log A - \log f(x) \Big) \;-\; (1+\alpha) x^{1/(1+\alpha)} \right] \;,
\end{aligned}$$

where (a) follows because $f'(x) = 1 - \frac{\alpha/(1+\alpha)}{x^{1/(1+\alpha)}}$, so

$$\frac{f'(x)}{x} \;=\; \frac{1}{x} - \frac{\alpha/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \quad \text{and} \quad \frac{x f'(x) - f(x)}{x^2} \;=\; \frac{\frac{1}{1+\alpha} x^{\alpha/(1+\alpha)}}{x^2} = \frac{1/(1+\alpha)}{x^{(2+\alpha)/(1+\alpha)}} \;.$$

So, the sign of $Q'(x)$ is determined by the sign of $\big(\alpha \;+\; \log A - \log f(x)\big) \;-\; (1+\alpha) x^{1/(1+\alpha)}$. Note that the function $x \mapsto \big(\alpha \;+\; \log A - \log f(x)\big)$ is a strictly decreasing function and the value as $x \downarrow 1$ is equal to ∞, and the value as $x \uparrow \infty$ is $-\infty$. On the other hand, the function $x \mapsto (1+\alpha) x^{1/(1+\alpha)}$ is strictly increasing with a value of $(1+\alpha)$ at $x = 1$, and its value as $x \uparrow \infty$ is ∞. So, these two functions intersect each other at exactly a single point, say h^*, and we have $Q'(x) > 0$ for all $x < h^*$, $Q'(h^*) = 0$, and $Q'(x) < 0$ for all $x > h^*$. This shows that Q is quasi-concave and has a unique maximizer at h^*. Moreover, note that $Q'(f^{-1}(A)) = \frac{1/(1+\alpha)}{\left(f^{-1}(A)\right)^{(2+\alpha)/(1+\alpha)}} \left[\alpha \;-\; (1+\alpha) \left(f^{-1}(A)\right)^{1/(1+\alpha)} \right] < 0$ because $f^{-1}(A) \ge 1$. Therefore, $h^* < f^{-1}(A)$, which completes part (a).

Appendix F: Proof of Lemma 5.2

The proof of Lemma 5.2 uses the following lemma that establishes an important property of the derivatives of g_α.

Lemma F.1 (Derivatives of g_α) *For each $\alpha > -1$ and $y \in \mathbb{R}_+$,* $y\left(\dfrac{g'_\alpha(y)}{g_\alpha(y)} - \dfrac{g''_\alpha(y)}{g'_\alpha(y)}\right) \le 1$.

Proof: Fix an arbitrary $\alpha > -1$. For ease of exposition, we will drop the reference to α and just write f and g. Because $f(x) = x - x^{\alpha/(1+\alpha)}$ and $f(g(y)) = y$, it follows from the the Inverse Function Theorem and the Chain Rule that

$$\begin{aligned} g'(y) &= \frac{1}{f'(g(y))} = \frac{1}{1 - \frac{\alpha/(1+\alpha)}{g(y)^{1/(1+\alpha)}}} \\ g''(y) &= -\frac{f''(g(y))g'(y)}{\left[f'(g(y))\right]^2} = -\frac{\alpha/(1+\alpha)^2}{g(y)^{(2+\alpha)/(1+\alpha)}} \times \frac{g'(y)}{\left(1 - \frac{\alpha/(1+\alpha)}{g(y)^{1/(1+\alpha)}}\right)^2}, \end{aligned}$$

where we use the fact that $f'(x) = 1 - \frac{\alpha/(1+\alpha)}{x^{1/(1+\alpha)}}$ and $f''(x) = \frac{\alpha/(1+\alpha)^2}{x^{(2+\alpha)/(1+\alpha)}}$. Therefore, it follows that

$$\begin{aligned} \frac{g'(y)}{g(y)} &= \frac{1}{g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}} \\ \frac{g''(y)}{g'(y)} &= -\frac{\alpha/(1+\alpha)^2}{g(y)^{(2+\alpha)/(1+\alpha)}} \times \frac{1}{\left(1 - \frac{\alpha/(1+\alpha)}{g(y)^{1/(1+\alpha)}}\right)^2} \\ &= -\frac{\alpha/(1+\alpha)^2}{g(y)^{1/(1+\alpha)}} \times \frac{1}{\left(g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}\right)\left(1 - \frac{\alpha/(1+\alpha)}{g(y)^{1/(1+\alpha)}}\right)} \\ &= -\frac{\alpha/(1+\alpha)^2}{\left(g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}\right)\left(g(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}\right)}. \end{aligned}$$

Because $y = f(g(y)) = g(y) - g(y)^{\alpha/(1+\alpha)}$, it follows that

$$\begin{aligned} y\left(\frac{g'(y)}{g(y)} - \frac{g''(y)}{g'(y)}\right) &= \frac{g(y) - g(y)^{\alpha/(1+\alpha)}}{g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}} \left(1 + \frac{\alpha/(1+\alpha)^2}{g(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}}\right) \\ &= \frac{g(y) - g(y)^{\alpha/(1+\alpha)}}{g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}} \times \frac{g(y)^{1/(1+\alpha)} - \left[\alpha/(1+\alpha)\right]^2}{g(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}}. \end{aligned} \tag{11}$$

Because $g(y) \ge 1$ and $\alpha > -1$, the denominator in the expression on the righthand side of Equation (11) is always positive. So,

$$\begin{aligned} & y\left(\frac{g'(y)}{g(y)} - \frac{g''(y)}{g'(y)}\right) \le 1 \\ \Leftrightarrow \quad & \left(g(y) - g(y)^{\alpha/(1+\alpha)}\right)\left(g(y)^{1/(1+\alpha)} - \frac{\alpha^2}{(1+\alpha)^2}\right) \le \left(g(y) - \frac{\alpha}{1+\alpha} g(y)^{\alpha/(1+\alpha)}\right)\left(g(y)^{1/(1+\alpha)} - \frac{\alpha}{1+\alpha}\right) \\ \overset{(a)}{\Leftrightarrow} \quad & -g(y)\left(1 + \frac{\alpha^2}{(1+\alpha)^2}\right) \le -g(y)\frac{2\alpha}{1+\alpha} \end{aligned}$$

$$\Leftrightarrow \quad 0 \leq 1 - \frac{2\alpha}{1+\alpha} + \frac{\alpha^2}{(1+\alpha)^2}$$
$$\Leftrightarrow \quad 0 \leq \left(1 - \frac{\alpha}{1+\alpha}\right)^2 ,$$

where (a) follows by multiplying out the expressions on both sides and canceling common terms. This establishes the desired result. ■

Here is the proof of Lemma 5.2.

Proof of Lemma 5.2: Fix an arbitrary $\alpha > -1$. For ease of exposition, we will drop the reference to α and just write g. It follows from the definition that

$$h'(y) = \frac{g'(y)}{g(y)} \qquad \text{and} \qquad h''(y) = \frac{g(y)g''(y) - [g'(y)]^2}{g(y)^2} ,$$

which implies that

$$\begin{aligned} h''(y) + \frac{h'(y)}{y} \; \geq \; 0 \qquad &\Leftrightarrow \qquad \frac{g(y)g'(y) + y\,(\,g(y)g''(y) \;-\; [g'(y)]^2\,)}{y\,[g(y)]^2} \;\geq\; 0 \\ &\overset{(a)}{\Leftrightarrow} \qquad g(y)g'(y) \geq y\,(\,[g'(y)]^2 \;-\; g(y)g''(y)\,) \\ &\overset{(b)}{\Leftrightarrow} \qquad 1 \geq y\left(\frac{g'(y)}{g(y)} \;-\; \frac{g''(y)}{g'(y)}\right) , \end{aligned}$$

where (a) follows because $g(y) \geq 1$ and $y > 0$, and (b) follows because $g(y)$ is strictly increasing so $g'(y) > 0$. The desired result then follows from Lemma F.1. ■

Appendix G: Preprocessing the Dataset from Expedia

Our approach for preprocessing the dataset is similar to the method used by Gao et al. (2021), and we provide the details here for completeness. The raw dataset includes about ten million rows and 54 columns. In some of the search queries, the price is given as the total amount over the whole length of the stay, whereas in some others, the price is given as the amount per night. It is not possible to reliably tell which approach is used in each search query. To avoid ambiguity, we focused our attention on the search queries for a single night stay and dropped the remaining search queries. Furthermore, we dropped the columns for which the entries are missing for more than 25% of the rows. Considering the remaining columns, we dropped the search queries for which the entries were missing in one of the remaining columns. Lastly, some rows in the dataset included entries that are too large or too small. We dropped all search queries which had an entry in a column that falls outside the 0.5^{th} and 99.5^{th} percentile band of the entries in the corresponding column. After preprocessing the dataset, we end up with 595,965 rows representing 34,561 search queries and 15 columns.

Appendix H: Testing the Statistical Significance of α

In each of the 90 datasets that we generate, we test the null hypothesis that $\alpha = 0$ using both the Likelihood-Ratio (Severini 2000) and Quasi-t tests (Ben-Akiva and Lerman 1985, pages 23-26). We describe the two tests below.

Under the Likelihood-Ratio test, the test statistic is $-2(LogLik^*_{SML} - LogLik^*_{CML})$ where $LogLik^*_{CML}$ is the maximum log-likelihood of our Context-MNL model under the training dataset and $LogLik^*_{SML}$ is the maximum log-likelihood value of the standard multinomial logit model under the same training dataset. Because the standard multinomial logit model corresponds to $\alpha = 0$ in our model, under the null hypothesis, the test statistic $-2(LogLik^*_{SML} - LogLik^*_{CML})$ has a chi-squared distribution with one degree of freedom. We can then reject the null hypothesis by comparing the value of the test statistics and critical value of chi-squared distribution with one degree of freedom.

In our estimation, we represent $\alpha = e^{\gamma} - 1$ where $\gamma \in \mathbb{R}$. The null hypothesis that $\alpha = 0$ is thus equivalent to $\gamma = 0$. Under the Quasi-t test, the test statistic is $\hat{\gamma}/\sqrt{\mathsf{Var}(\hat{\gamma})}$, where $\hat{\gamma}$ is the maximum likelihood estimates of γ and $\mathsf{Var}(\hat{\gamma})$ denotes its variance. Under fairly general conditions, maximum likelihood estimators are asymptotically normal, thus we use test statistic $\hat{\gamma}/\sqrt{\mathsf{Var}(\hat{\gamma})}$ and t-test with $N - K$ degrees, where N is number of training samples and K is total number of parameters we estimate. Since $N - K$ is very large under our case, we directly use a normal test. We can then reject the null hypothesis by comparing the value of the test statistics and critical value of standard normal distribution. To compute $\sqrt{\mathsf{Var}(\hat{\gamma})}$, we use the Cramer-Rao bound to estimate it, by numerically computing the Hessian matrix of log-likelihood function of the Context-MNL model at maximum likelihood estimates, and the inverse of negative Hessian matrix gives an approximation to the covariance matrix. In this case, the element in the corresponding diagonal entry would be our estimated $\mathrm{Var}(\hat{\gamma})$. In our experiments, we use two ways to compute the Hessian matrix: (1) we compute the Hessian directly using the finite difference method, and (2) we use the estimated Hessian obtained from the fminunc function in MATLAB. Both ways of computing the Hessian matrix give very similar results.