Technical Note: Multi-Product Pricing Under the Generalized Extreme Value Models with Homogeneous Price Sensitivity Parameters

Heng Zhang, Paat Rusmevichientong
Marshall School of Business, University of Southern California, Los Angeles, CA 90089,
hengz@usc.edu, rusmevic@marshall.usc.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell Tech, New York, NY 10011,
topaloglu@orie.cornell.edu

January 9, 2018

We consider unconstrained and constrained multi-product pricing problems when customers choose according to an arbitrary generalized extreme value (GEV) model and the products have the same price sensitivity parameter. In the unconstrained problem, there is a unit cost associated with the sale of each product. The goal is to choose the prices for the products to maximize the expected profit obtained from each customer. We show that the optimal prices of the different products have a constant markup over their unit costs. We provide an explicit formula for the optimal markup in terms of the Lambert-W function. In the constrained problem, motivated by the applications with inventory considerations, the expected sales of the products are constrained to lie in a convex set. The goal is to choose the prices for the products to maximize the expected revenue obtained from each customer, while making sure that the constraints for the expected sales are satisfied. If we formulate the constrained problem by using the prices of the products as the decision variables, then we end up with a non-convex program. We give an equivalent market-share-based formulation, where the purchase probabilities of the products are the decision variables. We show that the market-share-based formulation is a convex program, the gradient of its objective function can be computed efficiently, and we can recover the optimal prices for the products by using the optimal purchase probabilities from the market-share-based formulation. Our results for both unconstrained and constrained problems hold for any arbitrary GEV model.

Key words: customer choice modeling, generalized extreme value models, price optimization

1. Introduction

In most revenue management settings, customers make a choice among the set of products that are offered for purchase. While making their choices, customers substitute among the products based on attributes such as price, quality, and richness of features. In these situations, increasing the price for one product may shift the demand of other products, and such substitutions create complex interactions among the demands for the different products. There is a growing body of literature pointing out that capturing the choice process of customers and the interactions among the demands for different products through discrete choice models can significantly improve operational decisions; see, for example, Talluri and van Ryzin (2004), Gallego et al. (2004), and Vulcano et al. (2010). Nevertheless, as the discrete choice models become more complex, finding the optimal prices to charge for the products becomes more difficult as well. This challenge reflects the fundamental tradeoff between choice model complexity and operational tractability.

In this paper, we study unconstrained and constrained multi-product pricing problems when customers choose according to an arbitrary choice model from the generalized extreme value (GEV) family. The GEV family is a rather broad family of discrete choice models, as it encapsulates many widely studied discrete choice models as special cases, including the multinomial logit (Luce 1959, McFadden 1974, McFadden 1980), nested logit (Williams 1977, McFadden 1978), d-level logit (Daganzo and Kusnic 1993, Li et al. 2015, Li and Huh 2015), and paired combinatorial logit (Koppelman and Wen 2000, Chen et al. 2003, Li and Webster 2015). Throughout this paper, when we refer to a GEV model, we refer to an arbitrary choice model within the GEV family. For both unconstrained and constrained multi-product pricing problems studied in this paper, we consider the case where different products share the same price sensitivity parameter. We present results that hold simultaneously for all GEV models.

Our Contributions: In the unconstrained problem, there is a unit cost associated with the sale of each product. The goal is to set the prices for the products to maximize the expected profit from each customer. We show that the optimal prices of the different products have the same markup, which is to say that the optimal price of each product is equal to its unit cost plus a constant markup that does not depend on the product; see Theorem 3.1. We provide an explicit formula for the optimal markup in terms of the Lambert-W function; see Proposition 3.2. These results greatly simplify the computation of the optimal prices and they hold under any GEV model. We give comparative statistics that describe how the optimal prices change as a function of the unit costs; see Corollary 3.3. In particular, if the unit cost of a product increases, then its optimal price increases and the optimal prices of the other products decreases. If the unit costs of all products increase by the same amount, then the optimal prices of all products increase.

In the constrained problem, motivated by the applications with inventory considerations, the expected sales of the products are constrained to lie in a convex set. The goal is to set the prices for the products to maximize the expected revenue obtained from each customer while satisfying the constraints on the expected sales. A natural formulation of the constrained problem, which uses the prices of the products as the decision variables, is a non-convex program. We give an equivalent market-share-based formulation, where the purchase probabilities of the products are the decision variables. We show that the market-share-based formulation is a convex program that can be solved efficiently. In particular, for any given purchase probabilities for the products, we can recover the unique prices that achieve these purchase probabilities; see Theorem 4.1. Also, the objective function of the market-share-based formulation is concave in the purchase probabilities and its gradient can be computed efficiently; see Theorem 4.3. Thus, we can solve the market-share-based formulation and recover the optimal prices by using the optimal purchase probabilities.

The solution methods that we provide for the unconstrained and constrained problems are applicable to any GEV model. This generality comes at the expense of requiring homogeneous price sensitivity parameters. As discussed shortly in our literature review, there is a significant amount of work that studies pricing problems for specific instances of the GEV models, such as the multinomial logit and nested logit, under the assumption that the price sensitivities for the products are the same. Also, in many applications, customers choose among products that are in the same product category. In such cases, it is reasonable to expect similar price sensitivities across products. Cotterill (1994), Hausman et al. (1994), Chidmi and Lopez (2007), and Mumbower et al. (2014) estimate the price sensitivities for the products within the categories of soft drinks, domestic beer, breakfast cereal, and flights with the same change restrictions in an origin-destination market. They report similar price sensitivities for the products in each category.

Even when the price sensitivities of the products are the same, the GEV models can provide significant modeling flexibility, as they include many other parameters. Consider the generalized nested logit model, which is a GEV model. Let N be the set of all products and β be the price sensitivity of the products. Besides the price sensitivity β, the generalized nested logit model has the parameters $\{\alpha_i : i \in N\}$, $\{\tau_k : k \in L\}$, and $\{\sigma_{ik} : i \in N,\ k \in L\}$ for a generic index set L. If the prices of the products are $\boldsymbol{p} = (p_i : i \in N)$, then the choice probability of product i is

$$\Theta_i^{\text{GenNest}}(\boldsymbol{p}) = \frac{\sum_{k\in L}(\sigma_{ik}\, e^{\alpha_i - \beta\, p_i})^{1/\tau_k}\left(\sum_{j\in N}(\sigma_{jk}\, e^{\alpha_j-\beta\, p_j})^{1/\tau_k}\right)^{\tau_k-1}}{1+\sum_{k\in L}\left(\sum_{j\in N}(\sigma_{jk}\, e^{\alpha_j-\beta\, p_j})^{1/\tau_k}\right)^{\tau_k}}.$$

Letting c_i be the unit cost for product i, if we charge the prices $\boldsymbol{p}$, then the expected profit from a customer is $\sum_{i\in N}(p_i - c_i)\, \Theta_i^{\text{GenNest}}(\boldsymbol{p})$. This expected profit function is rather complicated when the purchase probabilities are as above, but our results show that we can efficiently find the prices that maximize this expected profit function. Also, Swait (2003), Daly and Bierlaire (2006) and Newman (2008) give general approaches to combine GEV models to generate new ones. The purchase probabilities under such new GEV models can be even more complicated.

Literature Review: There is a rich vein of literature on unconstrained multi-product pricing problems under specific members of the GEV family, including the multinomial logit, nested logit, and paired combinatorial logit, but these results make use of the specific form of the purchase probabilities under each specific GEV model. Hopp and Xu (2005) and Dong et al. (2009) consider the pricing problem under the multinomial logit model, Anderson and de Palma (1992) and Li and Huh (2011) consider the pricing problem under the nested logit model, and Li and Webster (2015) consider the pricing problem under the paired combinatorial logit model. Under each of these choice models, the authors show that if the price sensitivities of the products are the same, then the optimal prices for the products have a constant markup. We extend the constant markup result established in these papers from the multinomial logit, nested logit, and paired combinatorial logit models to an arbitrary choice model within the GEV family. Furthermore, the constant markup results established in these papers often exploit the structure of the specific choice model to find an explicit formula for the price of each product as a function of the purchase probabilities. This approach fails for general GEV models, as there is no explicit formula for the prices as a function of the choice probabilities, but it turns out that we can still establish that the optimal prices have a constant markup under any GEV model.

With the exception of Li and Huh (2011) and Li and Webster (2015), the papers mentioned in the paragraph above exclusively assume that the price sensitivities of the products are the same. Li and Huh (2011) also go one step beyond to study the pricing problem under the nested logit model when the products in each nest have the same price sensitivity. In this case, they show that the optimal prices for the products in each nest have a constant markup. In addition to the case with homogeneous price sensitivities for the products, Li and Webster (2015) also consider the pricing problem under the paired combinatorial logit model with arbitrary price sensitivities. The authors establish sufficient conditions on the price sensitivities to ensure unimodality of the expected profit function and give an algorithm to compute the optimal prices.

Other work on unconstrained multi-product pricing problems under specific GEV models includes Wang (2012), where the author considers joint assortment planning and pricing problems under the multinomial logit model with arbitrary price sensitivities. Gallego and Wang (2014) show that the expected profit function under the nested logit model can have multiple local maxima when the price sensitivities are arbitrary and give sufficient conditions on the price sensitivities to ensure unimodality of the expected profit function. Rayfield et al. (2015) study the pricing problem under the nested logit model with arbitrary price sensitivities and provide heuristics with performance guarantees. Li et al. (2015) and Li and Huh (2015) study pricing problems under the d-level nested logit model with arbitrary price sensitivities.

Our study of constrained multi-product pricing problems is motivated by the applications with inventory considerations. Gallego and van Ryzin (1997) study a network revenue management model where the sale of each product consumes a combination of resources and the resources have limited inventories. The goal is to find the prices for the products to maximize the expected revenue from each customer, while making sure that the expected consumptions of the resources do not exceed their inventories. The authors use their pricing problem to give heuristics for the case where customers arrive sequentially over time to make product purchases subject to resource availability. We show that their pricing problem is tractable under GEV models with homogeneous price sensitivities. Song and Xue (2007) and Zhang and Lu (2013) show that the expected revenue function under the multinomial logit model is concave in the market shares when the products have the same price sensitivity. Keller (2013) considers pricing problems under the multinomial logit and nested logit models when there are linear constraints on the expected sales of the products. The author establishes sufficient conditions to ensure that the expected revenue is concave in the market shares. Song and Xue (2007) and Zhang and Lu (2013) focus on the multinomial logit model with homogeneous price sensitivities for the products. Thus, our work generalizes theirs to an arbitrary GEV model. Keller (2013) works with non-homogeneous price sensitivities. In that sense, his work is more general than ours. However, Keller (2013) works with specific GEV models. In that sense, our work is more general than his.

Each GEV model is uniquely defined by a generating function. McFadden (1978) gives sufficient conditions on the generating function to ensure that the corresponding GEV model is compatible with the random utility maximization principle, where each customer associates random utilities with the available alternatives and chooses the alternative that provides the largest utility. McFadden (1980) discusses the connections between GEV models and other choice models. Train (2002) cover the theory and application of GEV models. Swait (2003), Daly and Bierlaire (2006) and Newman (2008) show how to combine generating functions from different GEV models to create a new GEV model. The GEV family offers a rich class of choice models. As discussed above, there is work on pricing problems under the multinomial logit, nested logit, paired combinatorial logit, and d-level nested logit models, but applications in numerous areas indicate that using other members of the GEV family can provide useful modeling flexibility. In particular, Small (1987) uses the ordered GEV model, Bresnahan et al. (1997) use the principles of differentiation GEV model, Vovsha (1997) uses the cross-nested logit model, Wen and Koppelman (2001) use the generalized nested logit model, Swait (2001) uses the choice set generation logit model, and Papola and Marzano (2013) use the network GEV model in applications including scheduling trips, route selection, travel mode choice, and purchasing computers.

Organization: The paper is organized as follows. In Section 2, we explain how we can characterize a GEV model by using a generating function. In Section 3, we study the unconstrained problem. In Section 4, we study the constrained problem. In Section 5, we conclude.

2. Generalized Extreme Value Models

A general approach to construct discrete choice models is based on the random utility maximization (RUM) principle. Under the RUM principle, each product, including the no-purchase option, has a random utility associated with it. The realizations of these random utilities are drawn from a particular probability distribution and they are known only to the customer. The customer chooses the alternative that provides the largest utility. We index the products by $N=\{1,\ldots,n\}$. We use 0 to denote the no-purchase option. For each $i\in N\cup\{0\}$, we let $U_i=\mu_i+\epsilon_i$ be the utility associated with alternative i, where μ_i is the deterministic utility component and ϵ_i is the random utility component. Under the RUM principle, the probability that a customer chooses alternative i is given by $\Pr\{U_i > U_\ell \ \ \forall\, \ell\in N\cup\{0\},\ \ell\neq i\}$. The family of GEV models allows us to construct discrete choice models that are compatible with the RUM principle. A GEV model is characterized by a generating function G that maps the vector $\boldsymbol{Y}=(Y_1,\ldots,Y_n)\in\mathbb{R}^n_+$ to a scalar $G(\boldsymbol{Y})$. The function G satisfies the following four properties.

(*i*) $G(\boldsymbol{Y})\geq 0$ for all $\boldsymbol{Y}\in\mathbb{R}^n_+$.

(*ii*) The function G is homogeneous of degree one. In other words, we have $G(\lambda\boldsymbol{Y})=\lambda G(\boldsymbol{Y})$ for all $\lambda\in\mathbb{R}_+$ and $\boldsymbol{Y}\in\mathbb{R}^n_+$.

(*iii*) For all $i\in N$, we have $G(\boldsymbol{Y})\to\infty$ as $Y_i\to\infty$.

(*iv*) Using $\partial G_{i_i,\ldots,i_k}(\boldsymbol{Y})$ to denote the cross partial derivative of the function G with respect to $Y_{i_1},\ldots,Y_{i_k}$ evaluated at $\boldsymbol{Y}$, if $i_1,\ldots,i_k$ are distinct from each other, then $\partial G_{i_1,\ldots,i_k}(\boldsymbol{Y})\geq 0$ when k is odd, whereas $\partial G_{i_1,\ldots,i_k}(\boldsymbol{Y})\leq 0$ when k is even.

Then, for any fixed vector $\boldsymbol{Y}\in\mathbb{R}^n_+$, under the GEV model characterized by the generating function G, the probability that a customer chooses product $i\in N$ is given by

$$\Theta_i(\boldsymbol{Y})=\frac{Y_i\,\partial G_i(\boldsymbol{Y})}{1+G(\boldsymbol{Y})}. \tag{1}$$

With probability $\Theta_0(\boldsymbol{Y})=1-\sum_{i\in N}\Theta_i(\boldsymbol{Y})$, a customer leaves without purchasing anything. Thus, the choice probabilities depend on the function G and the fixed vector $\boldsymbol{Y}\in\mathbb{R}^n_+$.

McFadden (1978) shows that if the function G satisfies the four properties described above, then for any fixed vector $\boldsymbol{Y}\in\mathbb{R}^n_+$, the choice probability in (1) is compatible with the RUM principle, where the deterministic utility components $(\mu_1,\ldots,\mu_n)$ are given by $\mu_i=\log Y_i$ for all $i\in N$, the deterministic utility component for the no-purchase option is fixed at $\mu_0=0$, and the random utility components $(\epsilon_0,\epsilon_1,\ldots,\epsilon_n)$ have a generalized extreme value distribution with the cumulative distribution function $F(x_0,x_1,\ldots,x_n)=\exp(-e^{-x_0}-G(e^{-x_1},\ldots,e^{-x_n}))$. The GEV models allow for correlated utilities and we can use different generating functions to model different correlation patterns among the random utilities. In the next example, we show that numerous choice models that are commonly used in the operations management and economics literature are specific instances of the GEV models.

Example 2.1 (Specific Instances of GEV Models) The multinomial logit, nested logit, and paired combinatorial logit models are all instances of the GEV models. For some generic index set L, consider the function G given by

$$G(\boldsymbol{Y}) = \sum_{k \in L} \left(\sum_{i \in N} (\sigma_{ik} \, Y_i)^{1/\tau_k} \right)^{\tau_k},$$

where for all $i \in N$, $k \in L$, $\tau_k \in (0,1]$, $\sigma_{ik} \geq 0$, and for all $i \in N$, $\sum_{k \in L} \sigma_{ik} = 1$. The function G above satisfies the four properties described at the beginning of this section. Thus, the expression in (1) with this choice of the function G yields a choice model that is consistent with the RUM principle. The choice model that we obtain by using the function G given above is called the generalized nested logit model. Train (2002) discusses how specialized choices of the index set L and the scalars $\{\tau_j : j \in L\}$ and $\{\sigma_{ik} : i \in N,\ k \in L\}$ result in well-known choice models. If the set L is the singleton $L = \{1\}$ and $\tau_1 = 1$, then $G(\boldsymbol{Y}) = \sum_{i \in N} Y_i$, and the expression in (1) yields the choice probabilities under the multinomial logit model. If, for each product $i \in N$, there exists a unique $k_i \in L$ such that $\sigma_{i,k_i} = 1$, then $G(\boldsymbol{Y}) = \sum_{g \in L} \left(\sum_{i \in N_g} Y_i^{1/\tau_g} \right)^{\tau_g}$ where $N_g = \{i \in N : k_i = g\}$, in which case, the expression in (1) yields the choice probabilities under the nested logit model, and k_i is known as the nest of product i. If the set L is given by $\{(i,j) \in N^2 : i \neq j\}$ and $\sigma_{ik} = 1/(2(n-1))$ whenever $k = (i,j)$ or (j,i) for some $j \neq i$, then $G(\boldsymbol{Y}) = \sum_{(i,j) \in N^2 : i \neq j} \left(Y_i^{1/\tau(i,j)} + Y_j^{1/\tau(i,j)} \right)^{\tau(i,j)} / (2(n-1))$, and the expression in (1) yields the choice probabilities under the paired combinatorial logit model.

The discussion in Example 2.1 indicates that the multinomial logit, nested logit, and paired combinatorial logit models are special cases of the generalized nested logit model. As discussed by Wen and Koppelman (2001), the ordered GEV, principles of differentiation GEV, and cross-nested logit are special cases of the generalized nested logit model as well. However, although the nested logit model is a special case of the generalized nested logit model, the d-level nested logit model is not a special case of the generalized nested logit model. In Figure 1, we show the relationship between well-known GEV models. In this figure, an arc between two GEV models indicates that the GEV model at the destination is a special case of the GEV model at the origin. In the next lemma, we give two properties of functions that are homogeneous of degree one. These properties are a consequence of a more general result, known as Euler's formula, but we provide a self-contained proof for completeness. We will use these properties extensively.

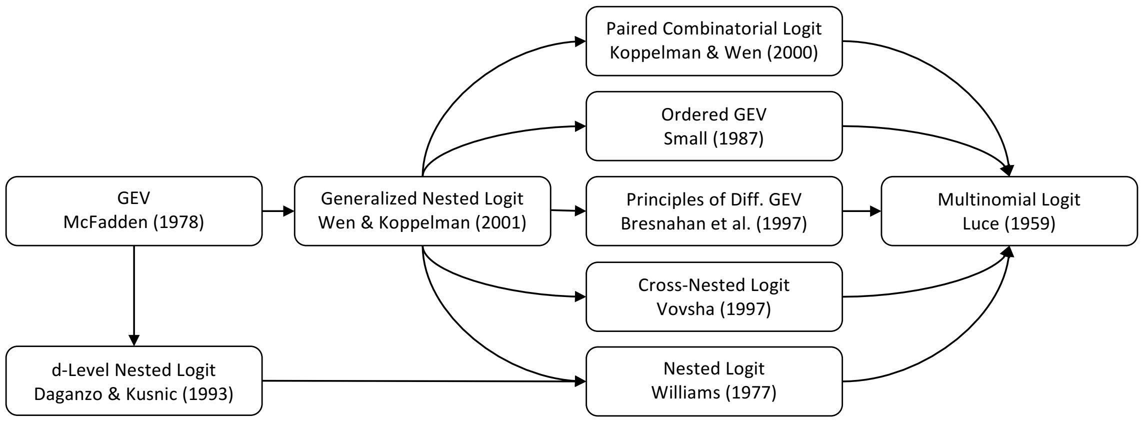

Figure 1 Relationship between well-known GEV models.

Lemma 2.2 (Properties of Generating Functions) *If G is a homogeneous function of degree one, then we have $G(\boldsymbol{Y}) = \sum_{i\in N} Y_i\, \partial G_i(\boldsymbol{Y})$ and $\sum_{j\in N} Y_j\, \partial G_{ij}(\boldsymbol{Y}) = 0$ for all $i \in N$.*

Proof: Since the function G is homogeneous of degree one, we have $G(\lambda \boldsymbol{Y}) = \lambda\, G(\boldsymbol{Y})$. Differentiating both sides of this equality with respect to λ, we obtain $\sum_{i\in N} Y_i\, \partial G_i(\lambda \boldsymbol{Y}) = G(\boldsymbol{Y})$. Using the last equality with $\lambda = 1$, we obtain $G(\boldsymbol{Y}) = \sum_{i\in N} Y_i\, \partial G_i(\boldsymbol{Y})$, which is the first desired equality. Also, differentiating both sides of this equality with respect to Y_j, we obtain $\partial G_j(\boldsymbol{Y}) = \partial G_j(\boldsymbol{Y}) + \sum_{i\in N} Y_i\, \partial G_{ij}(\boldsymbol{Y})$, in which case, canceling $\partial G_j(\boldsymbol{Y})$ on both sides and noting that $\partial G_{ij}(\boldsymbol{Y}) = G_{ji}(\boldsymbol{Y})$, we obtain $\sum_{i\in N} Y_i\, \partial G_{ji}(\boldsymbol{Y}) = 0$, which is the second desired equality. $\square$

3. Unconstrained Pricing

We consider unconstrained pricing problems where the mean utility of a product is a linear function of its price and we want to find the product prices that maximize the expected profit obtained from a customer. For each product $i \in N$, let $p_i \in \mathbb{R}$ denote the price charged for product i, and c_i denote its unit cost. As a function of the price of product i, the deterministic utility component of product i is given by $\mu_i = \alpha_i - \beta\, p_i$, where $\alpha_i \in \mathbb{R}$ and $\beta \in \mathbb{R}_+$ are constants. Anderson et al. (1992) interpret the parameter α_i as a measure of the quality of product i, while the parameter β is the price sensitivity that is common to all of the products. Throughout the paper, we focus on the case where all of the products share the same price sensitivity.

Noting the connection of the GEV models to the RUM principle discussed in the previous section, the deterministic utility component $\alpha_i - \beta\, p_i$ of product i is given by $\log Y_i$. So, let $Y_i(p_i) = e^{\alpha_i - \beta p_i}$ for all $i \in N$, and let $\boldsymbol{Y}(\boldsymbol{p}) = (Y_1(p_1), \ldots, Y_n(p_n))$. If we charge the prices $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n$, then it follows from the selection probability in (1) that a customer purchases product i with probability

$\Theta_i(\boldsymbol{p}) = Y_i(p_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))/(1+G(\boldsymbol{Y}(\boldsymbol{p})))$. Our goal is to find the prices for the products to maximize the expected profit from each customer, yielding the problem

$$\max_{\boldsymbol{p}\in\mathbb{R}^n} R(\boldsymbol{p}) \stackrel{\text{def}}{=} \sum_{i\in N}(p_i-c_i)\,\Theta_i(\boldsymbol{p}) = \sum_{i\in N}(p_i-c_i)\,\frac{Y_i(p_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}. \qquad (\textsc{Unconstrained})$$

Since the function G satisfies the four properties at the beginning of Section 2, we have $\partial G_i(\boldsymbol{Y})\geq 0$ for all $\boldsymbol{Y}\in\mathbb{R}^n_+$. We impose a rather mild additional assumption that $\partial G_i(\boldsymbol{Y})>0$ for all $\boldsymbol{Y}\in\mathbb{R}^n_+$ satisfying $Y_i>0$ for all $i\in N$; so the partial derivative is strictly positive whenever every entry of $\boldsymbol{Y}$ is positive. This assumption holds for all of the GEV models we are aware of, including those in Section 1, Example 2.1 and Figure 1. We shortly point out where this assumption becomes critical.

Let $\boldsymbol{p}^*$ denote the optimal solution to the Unconstrained problem. In Theorem 3.1, we will show that $\boldsymbol{p}^*$ has a constant markup, so $p_i^*-c_i=m^*$ for all $i\in N$ for some constant m^*. In other words, the optimal price of each product is equal to its unit cost plus a constant markup that does not depend on the product. In Proposition 3.2, we will also give an explicit formula for the optimal markup m^* in terms of the Lambert-W function. Since the Lambert-W function is available in most mathematical computation packages, this proposition greatly simplifies the computation of the optimal prices. Recall that the Lambert-W function is defined as follows: for all $x\in\mathbb{R}_+$, $W(x)$ is the unique value such that $W(x)\,e^{W(x)}=x$. Using standard calculus, it can be verified that $W(x)$ is increasing and concave in $x\in\mathbb{R}_+$; see Corless et al. (1996). The starting point for our discussion is the expression for the partial derivative of the expected profit function. Since $Y_i(p_i)=e^{\alpha_i-\beta p_i}$, we have that $dY_i(p_i)/dp_i=-\beta\,Y_i(p_i)$, in which case, using the definition of $R(\boldsymbol{p})$ in the Unconstrained problem, we have

$$\begin{aligned}
\frac{\partial R(\boldsymbol{p})}{\partial p_i} &= \Big\{Y_i(p_i)-\beta\,Y_i(p_i)\,(p_i-c_i)\Big\}\,\frac{\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\\
&\qquad -\sum_{j\in N}(p_j-c_j)\,Y_j(p_j)\,\frac{\partial G_{ji}(\boldsymbol{Y}(\boldsymbol{p}))\,(1+G(\boldsymbol{Y}(\boldsymbol{p})))-\partial G_j(\boldsymbol{Y}(\boldsymbol{p}))\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{(1+G(\boldsymbol{Y}(\boldsymbol{p})))^2}\,\beta\,Y_i(p_i)\\
&= \Big\{Y_i(p_i)-\beta\,Y_i(p_i)\,(p_i-c_i)\Big\}\,\frac{\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\\
&\qquad -\beta\,\frac{Y_i(p_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\left\{\sum_{j\in N}(p_j-c_j)\,\frac{Y_j(p_j)\,\partial G_{ji}(\boldsymbol{Y}(\boldsymbol{p}))}{\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}-\sum_{j\in N}(p_j-c_j)\,\frac{Y_j(p_j)\,\partial G_j(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\right\}\\
&= \beta\,\frac{Y_i(p_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\left\{\frac{1}{\beta}-(p_i-c_i)-\sum_{j\in N}(p_j-c_j)\,\frac{Y_j(p_j)\,\partial G_{ji}(\boldsymbol{Y}(\boldsymbol{p}))}{\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}+R(\boldsymbol{p})\right\}.
\end{aligned}$$

In the next theorem, we use the above derivative expression to show that the optimal prices for the Unconstrained problem involves a constant markup for all of the products.

Theorem 3.1 (Constant Markup is Optimal) *For all $i\in N$, $p_i^*-c_i=\frac{1}{\beta}+R(\boldsymbol{p}^*)$.*

Proof: Note that there exist optimal prices that are finite; the proof is straightforward but tedious, and we defer the details to Appendix A. Since the optimal prices are finite, they satisfy the first order conditions: $\frac{\partial R(\boldsymbol{p})}{\partial p_i}\Big|_{\boldsymbol{p}=\boldsymbol{p}^*} = 0$ for all i. The finiteness also implies that $Y_i(p_i^*) = e^{\alpha_i - \beta\, p_i^*} > 0$ for all $i \in N$. Since $\partial G_i(\boldsymbol{Y}) > 0$ for all $\boldsymbol{Y} \in \mathbb{R}^n_+$ with $Y_i > 0$ for all $i \in N$, we have $\partial G_i(\boldsymbol{Y}(\boldsymbol{p}^*)) > 0$ as well. Thus, if the prices $\boldsymbol{p}^*$ satisfy the first order conditions $\frac{\partial R(\boldsymbol{p})}{\partial p_i}\Big|_{\boldsymbol{p}=\boldsymbol{p}^*} = 0$ for all i, then by the expression for the partial derivative $\frac{\partial R(\boldsymbol{p})}{\partial p_i}$ right before the statement of the theorem[1],

$$p_i^* - c_i = \frac{1}{\beta} - \frac{1}{\partial G_i(\boldsymbol{Y}(\boldsymbol{p}^*))} \sum_{j \in N} (p_j^* - c_j)\, Y_j(p_j^*)\, \partial G_{ji}(\boldsymbol{Y}(\boldsymbol{p}^*)) + R(\boldsymbol{p}^*).$$

For notational brevity, define $m_i^* = p_i^* - c_i$. Without loss of generality, we index the products such that $m_1^* \geq \ldots \geq m_n^*$. By the discussion in Section 2, the function G satisfies the property $\partial G_{ji}(\boldsymbol{Y}) \leq 0$ for any $\boldsymbol{Y} \in \mathbb{R}^n_+$ and $i \neq j$. In this case, using the equality above for $i = 1$ and noting that we have $m_1^* \geq p_i^* - c_i$ for all $i \in N$, we obtain

$$m_1^* \leq \frac{1}{\beta} - \frac{1}{\partial G_1(\boldsymbol{Y}(\boldsymbol{p}^*))} \sum_{j \in N} m_1^*\, Y_j(p_j^*)\, \partial G_{j1}(\boldsymbol{Y}(\boldsymbol{p}^*)) + R(\boldsymbol{p}^*) = \frac{1}{\beta} + R(\boldsymbol{p}^*),$$

where the equality follows from Lemma 2.2. Therefore, we obtain $m_1^* \leq 1/\beta + R(\boldsymbol{p}^*)$. A similar argument also yields $m_n^* \geq 1/\beta + R(\boldsymbol{p}^*)$, in which case, we have $m_1^* \leq 1/\beta + R(\boldsymbol{p}^*) \leq m_n^*$. Noting the assumption that $m_1^* \geq \ldots \geq m_n^*$, we must have $1/\beta + R(\boldsymbol{p}^*) = m_1^* = \ldots = m_n^*$ and the desired result follows by noting that $m_i^* = p_i^* - c_i$. □

Noting Theorem 3.1, let $m^* = \frac{1}{\beta} + R(\boldsymbol{p}^*)$ denote the optimal markup. In the next proposition, we give an explicit formula for m^* in terms of the Lambert-W function.

Proposition 3.2 (Explicit Formula for the Optimal Markup) *Let the scalar γ be defined as $\gamma = G(Y_1(c_1), \ldots, Y_n(c_n)) = G(e^{\alpha_1 - \beta\, c_1}, \ldots, e^{\alpha_n - \beta\, c_n})$. Then,*

$$m^* = \frac{1 + W(\gamma\, e^{-1})}{\beta} \quad \text{and} \quad R(\boldsymbol{p}^*) = \frac{W(\gamma\, e^{-1})}{\beta}.$$

Proof: The optimal prices have a constant markup. So, we focus on price vectors $\boldsymbol{p}$ such that $p_i - c_i = m$ for all $i \in N$ for some $m \in \mathbb{R}_+$. Let $\boldsymbol{c} = (c_1, \ldots, c_n)$, and $\boldsymbol{Y}(m\,\boldsymbol{e} + \boldsymbol{c}) = (Y_1(m + c_1), \ldots, Y_n(m + c_n))$, where $\boldsymbol{e} \in \mathbb{R}^n$ is the vector with all entries of one. In this case, we can write the objective function of the Unconstrained problem as a function of m, which is given by

$$R(m) = \sum_{i \in N} m\, \frac{Y_i(m + c_i)\ \partial G_i(\boldsymbol{Y}(m\,\boldsymbol{e} + \boldsymbol{c}))}{1 + G(\boldsymbol{Y}(m\,\boldsymbol{e} + \boldsymbol{c}))} = m\, \frac{G(\boldsymbol{Y}(m\,\boldsymbol{e} + \boldsymbol{c}))}{1 + G(\boldsymbol{Y}(m\,\boldsymbol{e} + \boldsymbol{c}))},$$

[1] This step in the proof requires our assumption that $\partial G_i(\boldsymbol{Y}) > 0$ whenever $Y_i > 0$ for all i.

where the second equality relies on the fact that $\sum_{i\in N} Y_i(m+c_i)\,\partial G_i(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c})) = G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))$ by Lemma 2.2. Thus, we can compute the optimal objective value of the UNCONSTRAINED problem by maximizing $R(m)$ over all possible values of m. Since $dY_i(m+c_i)/dm = -\beta\,Y_i(m+c_i)$, differentiating the objective function above with respect to m, we get

$$\begin{aligned}\frac{dR(m)}{dm} &= \frac{G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}{1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))} - m\sum_{i\in N}\frac{\partial G_i(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}{(1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c})))^2}\,\beta\,Y_i(m+c_i)\\ &= \left(\frac{G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}{1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}\right)\left(1-\frac{\beta\,m}{1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}\right),\end{aligned}$$

where the second equality once again uses the fact that $\sum_{i\in N} Y_i(m+c_i)\,\partial G_i(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c})) = G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))$. Because $Y_i(m+c_i)$ is decreasing in m, and $\partial G_i(\boldsymbol{Y}) \geq 0$ for all $\boldsymbol{Y}\in\mathbb{R}^n_+$, it follows that $G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))$ is decreasing in m. Therefore, in the expression for $\frac{dR(m)}{dm}$, the term $1-\frac{\beta\,m}{1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))}$ is decreasing in m; this implies that the derivative $\frac{dR(m)}{dm}$ can change sign from positive to negative only once as the value of m increases, so $R(m)$ is quasiconcave in m. Thus, setting the derivative with respect to m to zero provides a maximizer of $R(m)$. By the derivative expression above, if $\frac{dR(m)}{dm}=0$, then $\beta\,m = 1+G(\boldsymbol{Y}(m\,\boldsymbol{e}+\boldsymbol{c}))$, so the optimal markup m^* satisfies

$$\begin{aligned}\beta\,m^* \;&=\; 1+G(\boldsymbol{Y}(m^*\,\boldsymbol{e}+\boldsymbol{c})) \;=\; 1+G(e^{\alpha_1-\beta\,(m^*+c_1)},\ldots,e^{\alpha_n-\beta\,(m^*+c_n)})\\ &=\; 1+e^{-\beta\,m^*}\,G(e^{\alpha_1-\beta\,c_1},\ldots,e^{\alpha_n-\beta\,c_n}) \;=\; 1+\gamma\,e^{-\beta\,m^*} \;=\; 1+\gamma\,e^{-1}e^{-(\beta\,m^*-1)},\end{aligned}$$

where the third equality uses the fact that G is homogeneous of degree one. The last chain of equalities implies that $(\beta\,m^*-1)\,e^{\beta\,m^*-1} = \gamma\,e^{-1}$, so that $W(\gamma\,e^{-1}) = \beta\,m^*-1$. Solving for m^*, we obtain $m^* = (1+W(\gamma\,e^{-1}))/\beta$, which is the desired expression for the optimal markup. Furthermore, since $p_i^*-c_i = m^*$ for all $i\in N$, Theorem 3.1 implies that the optimal objective value of the UNCONSTRAINED problem is $R(p^*) = m^*-1/\beta = W(\gamma\,e^{-1})/\beta$. □

By Proposition 3.2, to obtain the optimal prices, we can simply compute γ as in the proposition and set $m^* = (1+W(\gamma\,e^{-1}))/\beta$, in which case, the optimal price for product i is m^*+c_i. When the price sensitivities of the products are the same, the fact that the optimal prices have constant markup is shown in Proposition 1 in Hopp and Xu (2005) for the multinomial logit model, in Lemma 1 in Anderson and de Palma (1992) for the nested logit model, and in Lemma 3 in Li and Webster (2015) for the paired combinatorial logit model. Theorem 3.1 generalizes these results to an arbitrary GEV model. Explicit formulas for the optimal markup are given in Theorem 1 in Dong et al. (2009) for the multinomial logit model and in Theorem 1 in Li and Webster (2015) for the paired combinatorial logit model. Proposition 3.2 generalizes these results to an arbitrary GEV model. Theorem 3.1 also allows us to give comparative statistics that describe how the optimal prices change as a function of the unit costs. As a function of the unit product costs

$\boldsymbol{c} = (c_1, \ldots, c_n)$ in the UNCONSTRAINED problem, let $\boldsymbol{p}^*(\boldsymbol{c}) = (p_1^*(\boldsymbol{c}), \ldots, p_n^*(\boldsymbol{c}))$ denote the optimal prices. To facilitate our exposition, we use $\boldsymbol{e}_i \in \mathbb{R}_+^n$ for the vector with one in the i-th entry and zeros everywhere else, and designate $\boldsymbol{e} \in \mathbb{R}_+^n$ as the vector of all ones. In the next corollary, which is a corollary to Theorem 3.1, we show that if the unit cost of a product increases, then its optimal price increases and the optimal prices of the other products decreases, whereas if the unit costs of all products increase by the same amount, then the optimal prices of all products increase as well. We defer the proof to Appendix B.

Corollary 3.3 (Comparative Statistics) *For all* $\delta \geq 0$,

(a) *For all* $i \in N$, $p_i^*(\boldsymbol{c} + \delta\, \boldsymbol{e}_i) \geq p_i^*(\boldsymbol{c})$, *and for all* $j \neq i$, $p_j^*(\boldsymbol{c} + \delta\, \boldsymbol{e}_i) \leq p_j^*(\boldsymbol{c})$;

(b) *For all* $i \in N$, $p_i^*(\boldsymbol{c} + \delta\, \boldsymbol{e}) \geq p_i^*(\boldsymbol{c})$.

We can give somewhat more general versions of the results in this section. In particular, we partition the set of products N into the disjoint subsets $N^1, \ldots, N^m$ such that $N = \cup_{k=1}^m N^k$ and $N_k \cap N_{k'} = \varnothing$ for $k \neq k'$. Similarly, we partition the vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \in \mathbb{R}_+^n$ into the subvectors $\boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^m$ such that each subvector $\boldsymbol{Y}^k$ is given by $\boldsymbol{Y}^k = (Y_i : i \in N^k)$. Assume that the products in each partition N^k share the same price sensitivity β^k, and the generating function G is a separable function of the form $G(\boldsymbol{Y}) = \sum_{k=1}^m G^k(\boldsymbol{Y}^k)$, where the functions $G^1, \ldots, G^m$ satisfy the four properties discussed at the beginning of Section 2. In Appendix C, we use an approach similar to the one used in this section to show that the optimal prices for the products in the same partition have a constant markup and give a formula to compute the optimal markups. Considering unconstrained pricing problems under the nested logit model, when the products in each nest have the same price sensitivity, Theorem 2 in Li and Huh (2011) shows that the optimal prices for the products in each nest have a constant markup and gives a formula that can be used to compute the optimal markup. The generating function for the nested logit model is a separable function of the form $\sum_{k=1}^m \gamma^k\, G^k(\boldsymbol{Y}^k)$, where the products in a partition N^k correspond to the products in a nest. Thus, our results in Appendix C generalize Theorem 2 in Li and Huh (2011) to an arbitrary GEV model with a separable generating function. Throughout the paper, we do not explicitly work with separable generating functions to minimize notational burden.

4. Constrained Pricing

We consider constrained pricing problems where the expected sales of the products are constrained to lie in a convex set. Similar to the previous section, the products have the same price sensitivity parameter β. The goal is to find the product prices that maximize the expected revenue obtained from each customer, while satisfying the constraints on the expected sales. To formulate the constrained pricing problem, we define the vector $\boldsymbol{\Theta}(\boldsymbol{p}) = (\Theta_1(\boldsymbol{p}), \ldots, \Theta_n(\boldsymbol{p}))$, which includes the purchase probabilities of the products. To capture the constraints on the expected sales, let M denote some generic index set. For each $\ell \in M$, we let F_ℓ be a convex function that maps the vector $\boldsymbol{q} = (q_1, \ldots, q_n) \in \mathbb{R}^n_+$ to a scalar. We are interested in solving the problem

$$\max_{\boldsymbol{p}\in\mathbb{R}^n} \left\{ \sum_{i\in N} p_i\, \Theta_i(\boldsymbol{Y}(\boldsymbol{p})) \ : \ F_\ell(\boldsymbol{\Theta}(\boldsymbol{p})) \leq 0 \ \ \forall\, \ell \in M \right\}. \qquad \text{(CONSTRAINED)}$$

The objective function above accounts for the expected revenue from each customer. Interpreting $\Theta_i(\boldsymbol{p})$ as the expected sales for product i, the constraints ensure that the expected sales for the products lie in the convex set $\{\boldsymbol{q} \in \mathbb{R}^n_+ \ : \ F_\ell(\boldsymbol{q}) \leq 0 \ \ \forall\, \ell \in M\}$. The CONSTRAINED problem finds applications in the network revenue management setting, where the sale of each product consumes a combination of resources (Gallego and van Ryzin 1997). In this setting, the set M indexes the set of resources. The sale of product i consumes $a_{\ell i}$ units of resource ℓ. There are C_ℓ units of resource ℓ. The expected number of customer arrivals is T. We want to find the product prices to maximize the expected revenue from each customer, while ensuring that the expected consumption of each resource does not exceed its availability. If we charge the prices $\boldsymbol{p}$, then the expected sales for product i is $T\,\Theta_i(\boldsymbol{p})$. Thus, the constraint $\sum_{i\in N} a_{\ell i}\, T\,\Theta_i(\boldsymbol{p}) \leq C_\ell$ ensures that the total expected consumption of resource ℓ does not exceed its inventory. In this case, defining F_ℓ as $F_\ell(\boldsymbol{q}) = \sum_{i\in N} a_{\ell i}\, T\, q_i - C_\ell$, the constraints in the CONSTRAINED problem ensure that the expected capacity consumption of each resource does not exceed its inventory.

In the CONSTRAINED problem, the objective function is generally not concave in the prices $\boldsymbol{p}$. Also, although F_ℓ is convex, $F_\ell(\boldsymbol{\Theta}(\boldsymbol{p}))$ is not necessarily convex is $\boldsymbol{p}$. Thus, the CONSTRAINED problem is not a convex program[2]. However, by expressing the CONSTRAINED problem in terms of the purchase probabilities or market shares, we will reformulate the problem into a convex program. In our reformulation, the decision variables $\boldsymbol{q} = (q_1, \ldots, q_n)$ correspond to the purchase probabilities of the products. We let $\boldsymbol{p}(\boldsymbol{q}) = (p_1(\boldsymbol{q}), \ldots, p_n(\boldsymbol{q}))$ denote the prices that achieve the purchase probabilities $\boldsymbol{q}$. Our reformulation of the CONSTRAINED problem is

$$\max_{\boldsymbol{q}\in\mathbb{R}^n_+} \left\{ \sum_{i\in N} p_i(\boldsymbol{q})\, q_i \ : \ F_\ell(\boldsymbol{q}) \leq 0 \ \ \forall\, \ell \in M, \ \sum_{i\in N} q_i \leq 1 \right\}. \quad \text{(MARKET-SHARE-BASED)}$$

[2] The objective function of the CONSTRAINED problem is not quasi-concave. As an example, consider the multinomial logit choice model with $N = \{1, 2\}$ and $\alpha_1 = \alpha_2 = 10$ and $\beta = 1$. Then, the objective function is given by

$$f(p_1, p_2) = \frac{p_1 e^{10-p_1} + p_2 e^{10-p_2}}{1 + e^{10-p_1} + e^{10-p_2}} \qquad \forall\ (p_1, p_2) \in \mathbb{R}^2.$$

If $(x_1, x_2) = (10, 20)$ and $(y_1, y_2) = (20, 10)$, then $f(x_1, x_2) = f(y_1, y_2) \approx 5.0$ but $f(0.5(x_1, x_2) + 0.5(y_1, y_2)) = f(15, 15) \approx 0.2 < \min\{f(x_1, x_2), f(y_1, y_2)\}$. So, the objective function is not quasi-concave.

The interpretations of the objective function and the first constraint in the Market-Share-Based formulation are similar to those of the Constrained problem. The last constraint in the Market-Share-Based formulation ensures that the total purchase probability of all products does not exceed one. We will establish the following results for the Market-Share-Based formulation. In Theorem 4.1 in Section 4.1, we show that for each market share vector $\boldsymbol{q}$, there exists the unique price vector $\boldsymbol{p}(\boldsymbol{q})$ that achieves the market shares in the vector $\boldsymbol{q}$. Furthermore, the price vector $\boldsymbol{p}(\boldsymbol{q})$ is the solution of an unconstrained minimization problem with a strictly convex objective function. Therefore, computing $\boldsymbol{p}(\boldsymbol{q})$ is tractable. Then, in Theorem 4.3 in Section 4.2, we show that the objective function in the Market-Share-Based formulation $\boldsymbol{q} \mapsto \sum_{i\in N} p_i(\boldsymbol{q})\, q_i$ is concave in $\boldsymbol{q}$ and we give an expression for its gradient. Since the constraints in the Market-Share-Based formulation are convex $\boldsymbol{q}$, we have a convex program. Thus, we can efficiently solve the Market-Share-Based formulation and obtain the optimal purchase probabilities $\boldsymbol{q}^*$ by using standard convex optimization methods (Boyd and Vandenberghe 2004). Once we compute the optimal purchase probabilities $\boldsymbol{q}^*$, we can also compute the corresponding optimal prices $\boldsymbol{p}(\boldsymbol{q}^*)$.

4.1 Prices as a Function of Purchase Probabilities

We focus on the question of how to compute the unique prices $\boldsymbol{p}(\boldsymbol{q}) = (p_1(\boldsymbol{q}), \ldots, p_n(\boldsymbol{q}))$ that are necessary to achieve the given purchase probabilities $\boldsymbol{q} = (q_1, \ldots, q_n)$. The main result of this section is stated in the following theorem.

Theorem 4.1 (Inverse Mapping) *For each $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all $i \in N$ and $\sum_{i\in N} q_i < 1$, there exists a unique price vector $\boldsymbol{p}(\boldsymbol{q})$ such that $q_i = \Theta_i(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))$ for all $i \in N$. Moreover, $\boldsymbol{p}(\boldsymbol{q})$ is the finite and unique solution to the strictly convex minimization problem*

$$\min_{\boldsymbol{s}\in\mathbb{R}^n} \left\{ \frac{1}{\beta} \log(1 + G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i\in N} q_i\, s_i \right\}.$$

The proof of Theorem 4.1 makes use of the lemma given below. Throughout this section, all vectors are assumed to be column vectors. For any vector $\boldsymbol{s} \in \mathbb{R}^n$, $\boldsymbol{s}^\top$ denotes its transpose and will be always be a row vector, whereas $\mathrm{diag}(\boldsymbol{s})$ denotes an n-by-n diagonal matrix whose diagonal entries correspond to the vector $\boldsymbol{s}$. Also, let $\nabla G(\boldsymbol{Y}(\boldsymbol{s}))$ denote the gradient vector of the generator function G evaluated at $\boldsymbol{Y}(\boldsymbol{s})$ and $\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))$ denote the Hessian matrix of G evaluated at $\boldsymbol{Y}(\boldsymbol{s})$. Last but not least, we use $\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s})) \in \mathbb{R}^n$ to denote an n-dimensional vector whose entries are the selection probabilities $\Theta_1(\boldsymbol{Y}(\boldsymbol{s})),\ldots,\Theta_n(\boldsymbol{Y}(\boldsymbol{s}))$. Fix an arbitrary $\boldsymbol{q}\in\mathbb{R}^n_+$ such that $q_i>0$ for all i and $\sum_{i\in N} q_i<1$, and let $f:\mathbb{R}^n\to\mathbb{R}$ be defined by: for all $\boldsymbol{s}\in\mathbb{R}^n$,

$$f(\boldsymbol{s})=\frac{1}{\beta}\log(1+G(\boldsymbol{Y}(\boldsymbol{s})))+\sum_{i\in N} q_i\, s_i.$$

In the next lemma, we give the expressions for the gradient $\nabla f(\boldsymbol{s})$ and the Hessian $\nabla^2 f(\boldsymbol{s})$. The proof of this lemma directly follows by differentiating the function f and using the definition of the choice probabilities in (1). We defer the proof to Appendix D.

Lemma 4.2 (Gradient and Hessian) *For all $\boldsymbol{s}\in\mathbb{R}^n$, $\nabla f(\boldsymbol{s})=\boldsymbol{q}\;-\;\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))$ and*

$$\frac{1}{\beta}\nabla^2 f(\boldsymbol{s})=\mathrm{diag}\left(\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\right)\;-\;\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))^\top\;+\;\frac{\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\,\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)}{1+G(\boldsymbol{Y}(\boldsymbol{s}))}.$$

In the proof of Theorem 4.1, we will also use two results in linear algebra. First, if the vector $\boldsymbol{v}\in\mathbb{R}^n_+$ satisfies $v_i>0$ for all i and $\sum_{i=1}^n v_i<1$, then the matrix $\mathrm{diag}(\boldsymbol{v})-\boldsymbol{v}\boldsymbol{v}^\top$ is positive definite. To see this result, since $1-\boldsymbol{v}^\top\mathrm{diag}(\boldsymbol{v})^{-1}\boldsymbol{v}=1-\sum_{i=1}^n v_i>0$, by the Sherman-Morrison formula, the inverse of $\mathrm{diag}(\boldsymbol{v})-\boldsymbol{v}\boldsymbol{v}^\top$ exists and it is given by $\mathrm{diag}(\boldsymbol{v})^{-1}+(\mathrm{diag}(\boldsymbol{v}))^{-1}\boldsymbol{v}\,\boldsymbol{v}^\top\mathrm{diag}(\boldsymbol{v})^{-1}/(1-\boldsymbol{e}^\top\boldsymbol{v})=\mathrm{diag}(\boldsymbol{v})^{-1}+\boldsymbol{e}\,\boldsymbol{e}^\top/(1-\sum_{i=1}^n v_i)$; see Section 0.7.4 in Horn and Johnson (2012). The last matrix is clearly positive definite, which implies that $\mathrm{diag}(\boldsymbol{v})-\boldsymbol{v}\boldsymbol{v}^\top$ is also positive definite. Second, if $\mathbf{A}$ is a symmetric matrix such that each row sums to zero and all off-diagonal entries are non-positive, then $\mathbf{A}$ is positive semidefinite. To see this result, by our assumption, $\mathbf{A}$ is a symmetric and diagonally dominant matrix with non-negative diagonal entries, and such a matrix is known to be positive semidefinite; see Theorem A.6 in de Klerk (2004). Here is the proof of Theorem 4.1.

Proof of Theorem 4.1: Note that the objective function of the minimization problem in the theorem is $f(\boldsymbol{s})$. We claim that f is strictly convex. For any $\boldsymbol{s}\in\mathbb{R}^n$, let $Y_i(s_i)=e^{\alpha_i-\beta\,s_i}>0$ for all $i\in N$, so that $\Theta_i(\boldsymbol{Y}(\boldsymbol{s}))=Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))/(1+G(\boldsymbol{Y}(\boldsymbol{s}))>0$, where the inequality is by the assumption that $\partial G_i(\boldsymbol{Y})>0$ when $Y_i>0$ for all $i\in N$. Using Lemma 2.2, we also have

$$\sum_{i\in N}\Theta_i(\boldsymbol{Y}(\boldsymbol{s}))\;=\;\sum_{i\in N}\frac{Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))}{1+G(\boldsymbol{Y}(\boldsymbol{s}))}\;=\;\frac{G(\boldsymbol{Y}(\boldsymbol{s}))}{1+G(\boldsymbol{Y}(\boldsymbol{s}))}\;<\;1.$$

In this case, by the first linear algebra result, the matrix $\mathrm{diag}\left(\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\right)-\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\,\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))^\top$ is positive definite. Next, consider the matrix $\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\,\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)$, which is symmetric and its (i,j)-th component is given by $Y_i(s_i)\,\partial G_{ij}(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)$. For $i\neq j$, we have $\partial G_{ij}(\boldsymbol{Y}(\boldsymbol{s}))\le 0$ by the property of the generating function G, so all off-diagonal entries of the matrix are non-positive. Furthermore, by Lemma 2.2, we have $\sum_{j\in N} Y_i(s_i)\,\partial G_{ij}(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)=0$, so that each row of the matrix sums the zero. In this case, by the second linear algebra result, the matrix $\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\,\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)$ is positive semidefinite.

By the discussion in the previous paragraph, the matrix $\operatorname{diag}(\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))) - \boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\,\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))^\top$ is positive definite and the matrix $\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{s}))\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{s}))$ is positive semidefinite. Adding a positive definite matrix to a positive semidefinite matrix gives a positive definite matrix. In this case, noting the expression for the Hessian of f given in Lemma 4.2, f is strictly convex, which establishes the claim. Therefore, f has a unique minimizer. Furthermore, we can show that for any $L \geq 0$, there exists an $M \geq 0$, such that having $\|\boldsymbol{s}\| \geq M$ implies that $f(\boldsymbol{s}) \geq L$; the proof is straightforward but tedious, and we defer the details to Appendix E. Therefore, given some $\boldsymbol{s}_0$ with $f(\boldsymbol{s}_0) \geq 0$, there exists $M_0 \geq 0$ such that having $\|\boldsymbol{s}\| \geq M_0$ implies that $f(\boldsymbol{s}) \geq f(\boldsymbol{s}_0)$. In this case, the minimizer of f must lie in the set $\{\boldsymbol{s} \in \mathbb{R}^n : \|\boldsymbol{s}\| \leq M_0\}$, which implies that f has a finite minimizer. Since f is strictly convex and it has a finite minimizer, its minimizer $\boldsymbol{p}(\boldsymbol{q})$ is the solution to the first-order condition $\nabla f(\boldsymbol{p}(\boldsymbol{q})) = \mathbf{0}$, where $\mathbf{0}$ is the vector of all zeros. In this case, by the expression for the gradient of f given in Lemma 4.2, we must have $\nabla f(\boldsymbol{p}(\boldsymbol{q})) = \boldsymbol{q} - \boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q}))) = \mathbf{0}$, which implies that $q_i = \Theta_i(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))$ for all $i \in N$, as desired. □

To summarize, given a vector of purchase probabilities $\boldsymbol{q}$, the unique price vector $\boldsymbol{p}(\boldsymbol{q})$ that achieves these purchase probabilities is the unique optimal solution to the minimization problem $\min_{\boldsymbol{s}\in\mathbb{R}^n}\left\{\frac{1}{\beta}\log(1+G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i=1}^n q_i\, s_i\right\}$. Because the objective function in this problem is strictly convex, with its gradient given in Lemma 4.2, and there are no constraints on the decision variables, we can compute $\boldsymbol{p}(\boldsymbol{q})$ efficiently using standard convex optimization methods. We emphasize that one might be tempted to set $q_i = \Theta_i(\boldsymbol{Y}(\boldsymbol{p}))$ for all $i \in N$ and solve for $\boldsymbol{p}$ in terms of $\boldsymbol{q}$ in order to compute $\boldsymbol{p}(\boldsymbol{q})$. However, solving this system of equations directly is difficult. Even showing that there is a unique solution to this system of equations is not straightforward. Theorem 4.1 shows that there is a unique solution to this system of equations, and we can compute the solution by solving an unconstrained convex optimization problem.

4.2 Concavity of the Expected Revenue Function and its Gradient

Let $R(\boldsymbol{q}) = \sum_{i\in N} p_i(\boldsymbol{q})\, q_i$ denote the expected revenue function that is defined in terms of the market shares. The main result of this section is stated in the following theorem, which shows that $R(\boldsymbol{q})$ is concave in $\boldsymbol{q}$ and provides an expression for its gradient.

Theorem 4.3 (Concavity of the Revenue Function in terms of Market Shares) *For all $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all i and $\sum_{i\in N} q_i < 1$, the Hessian matrix $\nabla^2 R(\boldsymbol{q})$ is negative definite and $\nabla R(\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{q}) - \dfrac{1}{\beta\,(1-\boldsymbol{e}^\top \boldsymbol{q})}\,\boldsymbol{e}$.*

Before we proceed to the proof, we discuss the significance of Theorem 4.3. As noted at the beginning of Section 4, the function $\boldsymbol{p} \mapsto \sum_{i\in N} p_i \Theta_i(\boldsymbol{Y}(\boldsymbol{p}))$ is not necessarily concave in the prices $\boldsymbol{p}$. However, the theorem above shows that when we express the problem in terms of market shares $\boldsymbol{q}$, the expected revenue function $R(\boldsymbol{q})$ is concave in $\boldsymbol{q}$. Using the gradient of the expected revenue function in the theorem, we can then immediately solve the MARKET-SHARE-BASED problem using standard tools from convex programming.

Also, we note that the restriction that $q_i > 0$ for all i and $\sum_{i\in N} q_i < 1$ is necessary for the expected revenue function $R(\boldsymbol{q})$ and its derivatives to be well-defined. To give an example, we consider the multinomial logit model. Under this choice model, the selection probability of product i is $\Theta_i^{\mathrm{MNL}}(\boldsymbol{p}) = e^{\alpha_i - \beta p_i}/(1+\sum_{k\in N} e^{\alpha_k-\beta p_k})$. We can check that $p_i(\boldsymbol{q}) = \frac{1}{\beta}(\alpha_i + \log(1-\sum_{k\in N} q_k) - \log q_i)$ so that $R(\boldsymbol{q}) = \sum_{i\in N} \frac{1}{\beta}\,(\alpha_i + \log(1-\sum_{k\in N} q_k) - \log q_i)\, q_i$. This expected revenue function and its derivatives is well-defined only when $q_i > 0$ for all $i \in N$ and $\sum_{i\in N} q_i < 1$. A key ingredient in the proof of Theorem 4.3 is the Jacobian matrix $\mathbf{J}(\boldsymbol{q})$ associated with the vector-valued mapping $\boldsymbol{q} \mapsto \boldsymbol{p}(\boldsymbol{q})$, which is given in the following lemma. To characterize this Jacobian, we define the n-by-n matrix $\mathbf{B}(\boldsymbol{q}) = (B_{ij}(\boldsymbol{q}) : i,j\in N)$ as

$$\mathbf{B}(\boldsymbol{q}) = \frac{\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))\ \nabla^2 G(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))}{1+G(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))}.$$

The proof of Lemma 4.4 is given in Appendix F.

Lemma 4.4 (Jacobian) *The Jacobian matrix* $\mathbf{J}(\boldsymbol{q}) = \left(\frac{\partial p_i(\boldsymbol{q})}{\partial q_j} : i,j\in N\right)$ *is given by*

$$\mathbf{J}(\boldsymbol{q}) = -\frac{1}{\beta}\left(\operatorname{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})\right)^{-1}.$$

We are ready to give the proof of Theorem 4.3.

Proof of Theorem 4.3: First, we show the expression for $\nabla R(\boldsymbol{q})$. Since $R(\boldsymbol{q}) = \sum_{i\in N} p_i(\boldsymbol{q})\, q_i$, it follows that

$$\nabla R(\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{q}) + \mathbf{J}(\boldsymbol{q})^\top \boldsymbol{q} = \boldsymbol{p}(\boldsymbol{q}) - \frac{1}{\beta}\left(\operatorname{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})\right)^{-1}\boldsymbol{q},$$

where the last equality follows from Lemma 4.4. Consider the matrix $(\operatorname{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q}))^{-1}$ on the right side above. In the proof of Theorem 4.1, we show that $\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{s}))\ \nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\operatorname{diag}(\boldsymbol{Y}(\boldsymbol{s}))$ is positive semidefinite and each of its rows sums to zero. Noting the definition of $\mathbf{B}(\boldsymbol{q})$, we can use precisely the same argument to show that $\mathbf{B}(\boldsymbol{q})$ is positive semidefinite and each of its rows sums to zero as well. Since $\operatorname{diag}(\boldsymbol{q})$ is positive definite and $\mathbf{B}(\boldsymbol{q})$ is positive semidefinite, $\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q})$ is invertible, in which case, we get $(\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}(\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))\,\boldsymbol{e} = \boldsymbol{e}$. We have $\mathbf{B}(\boldsymbol{q})\,\boldsymbol{e} = \mathbf{0}$ because the rows of $\mathbf{B}(\boldsymbol{q})$ sum to zero. Noting also that $\operatorname{diag}(\boldsymbol{q})\boldsymbol{e} = \boldsymbol{q}$, the last equality implies that $(\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}\boldsymbol{q} = \boldsymbol{e}$. Using the fact that $\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q})$ is symmetric, taking the transpose, we have $\boldsymbol{q}^\top(\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1} = \boldsymbol{e}^\top$ as well. In this case, since we have $1 - \boldsymbol{q}^\top(\operatorname{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}\boldsymbol{q} =$

$1-\boldsymbol{q}^\top \boldsymbol{e} = 1-\sum_{i=1}^n q_i > 0$, by the Sherman-Morrison formula, the inverse of $\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})$ exists and it is given by

$$\begin{aligned}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q}))^{-1} \;&=\; (\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1} \;+\; \frac{(\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}\boldsymbol{q}\,\boldsymbol{q}^\top (\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}}{1-\boldsymbol{q}^\top (\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}\boldsymbol{q}} \\ &=\; (\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1} \;+\; \frac{1}{1-\boldsymbol{e}^\top \boldsymbol{q}}\,\boldsymbol{e}\,\boldsymbol{e}^\top.\end{aligned}$$

Using the equality above in the expression for $\nabla R(\boldsymbol{q})$ at the beginning of the proof, together with the fact that $(\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1}\boldsymbol{q} = \boldsymbol{e}$, we get

$$\nabla R(\boldsymbol{q}) \;=\; \boldsymbol{p}(\boldsymbol{q}) - \frac{1}{\beta}\left[\boldsymbol{e} + \left(\frac{\boldsymbol{e}^\top \boldsymbol{q}}{1-\boldsymbol{e}^\top \boldsymbol{q}}\right)\boldsymbol{e}\right] \;=\; \boldsymbol{p}(\boldsymbol{q}) - \frac{1}{\beta\,(1-\boldsymbol{e}^\top \boldsymbol{q})}\boldsymbol{e},$$

which is the desired expression for $\nabla R(\boldsymbol{q})$. Second, we show that $\nabla^2 R(\boldsymbol{q})$ is negative definite. By the discussion at the beginning of the proof, $\mathbf{B}(\boldsymbol{q})$ is positive semidefinite. By the first linear algebra result discussed right after Lemma 4.2, $\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top$ is positive definite. Thus, $(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q}))^{-1}$ is positive definite. Writing the gradient expression above componentwise, we get $\partial R(\boldsymbol{q})/\partial q_i = p_i(\boldsymbol{q}) - 1/(\beta\,(1-\sum_{k\in N} q_k))$; differentiating it with respect to q_j, we obtain $\partial^2 R(\boldsymbol{q})/\partial q_i \partial q_j = \partial p_i(\boldsymbol{q})/\partial q_j - 1/(\beta\,(1-\sum_{k\in N} q_k)^2)$. The last equality in matrix notation is

$$\nabla^2 R(\boldsymbol{q}) = \mathbf{J}(\boldsymbol{q}) - \frac{1}{\beta\,(1-\sum_{k\in N} q_k)^2}\boldsymbol{e}\,\boldsymbol{e}^\top \;=\; -\frac{1}{\beta}\left[\left(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})\right)^{-1} + \frac{1}{(1-\sum_{k\in N} q_k)^2}\boldsymbol{e}\,\boldsymbol{e}^\top\right],$$

where the last equality uses Lemma 4.4. The above equality shows that $\nabla^2 R(\boldsymbol{q})$ is negative definite, because $\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})$ is positive definite, so its inverse is also positive definite. □

Note that all of the results in this section continue to hold when products have unit costs. In that case, the revenue function is given by $R(\boldsymbol{q}) = \sum_{i\in N} (p_i(\boldsymbol{q}) - c_i)\, q_i$, where c_i denotes the unit cost of product i. The statements of all theorems and lemmas remain the same, except in Theorem 4.3, where the expression of the gradient $\nabla R(\boldsymbol{q})$ will change to $\nabla R(\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{q}) - \dfrac{1}{\beta\,(1-\boldsymbol{e}^\top \boldsymbol{q})}\,\boldsymbol{e} - \boldsymbol{c}$ to include the unit cost vector $\boldsymbol{c} = (c_1, \ldots, c_n)$.

The results that we present in this section demonstrate that any optimization problem that maximizes the expected revenue subject to constraints that are convex in the market shares of the products can be solved efficiently, as long as customers choose under a member of the GEV family with homogeneous price sensitivities. As discussed at the beginning of Section 4, our formulation of the constrained pricing problem finds applications in network revenue management settings, where the goal is to set the prices for the products to maximize the expected revenue obtained from each customer, the sale of a product consumes a combination of resources, and the resources have limited inventories (Gallego and van Ryzin 1997, Bitran and Caldentey 2003). Our formulation also becomes useful when the products have limited inventories and we set the prices for the products to maximize the expected revenue obtained from each customer, subject to the constraint that the expected sales of the products do not exceed their inventories (Gupta et al. 2006, Gallego and Hu 2014). There may be other applications, where the goal is to set the prices for the products so that the market shares of the products deviate from fixed desired market shares by at most a given margin. We can handle such constraints through convex functions of the market shares as well. In Appendix G, we provide a short numerical study in the network revenue management setting, where the choices of the customers are governed by the paired combinatorial logit model with homogeneous price sensitivities. When we have as many as 200 products and 80 resources, we can use our results to solve the corresponding constrained multi-product pricing problem in less than 20 seconds.

Under the multinomial logit model with homogeneous price sensitivities, Proposition 2 in Song and Xue (2007) and Section 3.3 in Zhang and Lu (2013) give a formula for the prices that achieve given market shares and show that the expected revenue function is concave in the market shares. Our Theorems 4.1 and 4.3 generalize these results to an arbitrary GEV model. When the customers choose according to the nested logit model and the products in each nest have the same price sensitivity, the discussion at the beginning of Section 2.1 and Theorem 1 in Li and Huh (2011) give a formula for the prices that achieve given market shares and show that the expected revenue function is concave in the market shares. Naturally, if the price sensitivities of all products are the same, then these results apply and imply that the expected revenue function under the nested logit model is concave in the market shares. Also, considering the same setup discussed at the end of Section 3, where the set of products are partitioned into the subsets and the generating function is separable by the partitions, in Appendix H, we show that if the products in each partition share the same price sensitivity, then we can compute the prices that achieve given market shares efficiently and the expected revenue function is concave in the market shares. The generating function for the nested logit model is separable by the products in each nest, so our results in Appendix H generalize the discussion at the beginning of Section 2.1 and Theorem 1 in Li and Huh (2011) to an arbitrary GEV model with a separable generating function.

5. Conclusions

This paper unifies and extends some of the pricing results that were discovered under special cases of the GEV model, such as the multinomial logit, nested logit, and paired combinatorial logit models. The value of our results derives from the fact that they hold under any arbitrary GEV model. For instance, to our knowledge, there has been no attempt to solve multi-product constrained pricing problems under the paired combinatorial logit model. The generality of our results comes at the cost of assuming that the price sensitivity parameters of the products are identical. Existing research has shown that this assumption is reasonable for certain product categories, including soft drinks, domestic beer, breakfast cereal, and flights with the same change restrictions in an origin-destination market. An important avenue for research is to investigate to what extent our results can be extended to non-homogeneous price sensitivities. In Appendices C and H, we extend our results to the case where the set of products are partitioned into subsets, the generating function is separable by the products in different partitions, and the products in a partition share the same price sensitivity. This extension is a step towards non-homogeneous price sensitivities, but addressing completely general price sensitivities seems rather nontrivial. Another interesting research direction is to investigate a unified approach for product assortment problems under GEV models. Product assortment problems have a combinatorial nature, and thus, they appear to need an entirely new line of attack.

Acknowledgements: We thank Professor Dan Adelman, the associate editor, and two anonymous referees whose comments substantially helped us streamline our proofs and focus our exposition.

References

Anderson, S. P., A. de Palma. 1992. Multiproduct firms: A nested logit approach. *The Journal of Industrial Economics* **40**(3) 261–276.

Anderson, S. P., A. de Palma, J.-F. Thisse. 1992. *Discrete choice theory of product differentiation*. MIT Press.

Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–229.

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

Bresnahan, T. F., S. Stern, M. Trajtenberg. 1997. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *The RAND Journal of Economics* **28** S17–S44.

Chen, A., P. Kasikitwiwat, Z. Ji. 2003. Solving the overlapping problem in route choice with paired combinatorial logit model. *Transportation Research Record* **1857**(1) 65–73.

Chidmi, Benaissa, Rigoberto A Lopez. 2007. Brand-supermarket demand for breakfast cereals and retail competition. *American Journal of Agricultural Economics* **89**(2) 324–337.

Corless, R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth. 1996. On the Lambert-W function. *Advances in Computational Mathematics* **5**(1) 329–359.

Cotterill, Ronald W. 1994. Scanner Data: New opportunities for demand and competitive strategy analysis. *Agricultural and Resource Economics Review* **23**(2) 125–139.

Daganzo, C. F., M. Kusnic. 1993. Technical note – two properties of the nested logit model. *Transportation Science* **27**(4) 395–400.

Daly, A., M. Bierlaire. 2006. A general and operational representation of Generalised Extreme Value models. *Transportation Research Part B: Methodological* **40**(4) 285–305.

de Klerk, E. 2004. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, New York, NY.

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

Gallego, G., M. Hu. 2014. Dynamic pricing of perishable assets under competition. *Management Science* **60**(5) 1241–1259.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. Working Paper, Columbia University.

Gallego, G., G. J. van Ryzin. 1997. A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research* **45**(1) 24–41.

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

Gupta, D., A. V. Hill, T. Bouzdine-Chameeva. 2006. A pricing model for clearing end-of-season retail inventory. *European Journal of Operational Research* **170** 518–540.

Hausman, Jerry, Gregory Leonard, J Douglas Zona. 1994. Competitive analysis with differenciated products. *Annales d'Economie et de Statistique* (34) 159–180.

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

Horn, R. A., C. R. Johnson. 2012. *Matrix Analysis*. Cambridge University Press, New York, NY.

Keller, P. W. 2013. Tractable multi-product pricing under discrete choice models. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Koppelman, F. S., C.-H. Wen. 2000. The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological* **34**(2) 75–89.

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

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

Li, H., S. Webster. 2015. Optimal pricing of correlated product options under the pairwise combinatorial logit model. Working Paper, University of Arizona.

Luce, R. D. 1959. *Individual choice behavior: a theoretical analysis*. Wiley, New York, NY.

McFadden, D. 1978. Modeling the choice of residential location. A. Karlqvist, L. Lundqvist, F. Snickars, J. Weibull, eds., *Spatial Interaction Theory and Planning Models*. North Holland, 531–551.

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

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., Frontiers in Economics. Academic Press, New York, NY, 105–142.

Mumbower, S., L. A. Garrow, M. J. Higgins. 2014. Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. *Transportation Research Part A: Policy and Practice* **66** 196–212.

Newman, J. P. 2008. Normalization of network generalized extreme value models. *Transportation Research Part B: Methodological* **42**(10) 958–969.

Papola, A., V. Marzano. 2013. A network generalized extreme value model for route choice allowing implicit route enumeration. *Computer-Aided Civil and Infrastructure Engineering* **28**(8) 560–580.

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

Small, K. A. 1987. A discrete choice model for ordered alternatives. *Econometrica* **55**(2) 409–424.

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

Swait, J. 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological* **35**(7) 643–666.

Swait, J. 2003. Flexible covariance structures for categorical dependent variables through finite mixtures of generalized extreme value models. *Journal of Business & Economic Statistics* **21**(1) 80–87.

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

Train, K. 2002. *Discrete choice methods with simulation*. Cambridge University Press.

Vovsha, P. 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board* **1607** 6–15.

Vulcano, G., G. J. van Ryzin, W. Chaar. 2010. Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management* **12**(3) 371–392.

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

Wen, C.-H., F. S. Koppelman. 2001. The generalized nested logit model. *Transportation Research Part B: Methodological* **35**(7) 627–641.

Williams, H. C. W. L. 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A* **9**(3) 285–344.

Zhang, D., Z. Lu. 2013. Assessing the value of dynamic pricing in network revenue management. *INFORMS Journal on Computing* **25**(1) 102 –115.

Appendix A: Finiteness of the Optimal Prices

We show that the optimal prices in the UNCONSTRAINED problem are finite. In the next lemma, we begin by showing that if we increase the prices of some products, then the purchase probabilities of the remaining products, as well as the probability of no-purchase, increase.

Lemma A.1 *For some $M \subseteq N$, assume that the prices $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_n)$ and $\tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_n)$ satisfy $\hat{p}_i \geq \tilde{p}_i$ for all $i \in M$ and $\hat{p}_i = \tilde{p}_i$ for all $i \in N \setminus M$. Then, we have $\Theta_i(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \Theta_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$ for all $i \in N \setminus M$ and $\Theta_0(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \Theta_0(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$.*

Proof: First, we show that $\Theta_i(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \Theta_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$ for all $i \in N \setminus M$. Fix $i \in N \setminus M$. Noting that $Y_j(p_j) = \exp(\alpha_j - \beta\, p_j)$, since $\hat{p}_j \geq \tilde{p}_j$ for all $j \in M$ and $\hat{p}_j = \tilde{p}_j$ for all $j \in N \setminus M$, we have $Y_j(\hat{p}_j) \leq Y_j(\tilde{p}_j)$ for all $j \in M$ and $Y_j(\hat{p}_j) = Y_j(\tilde{p}_j)$ for all $j \in N \setminus M$. Furthermore, noting that the function G satisfies $\partial G_{ij}(\boldsymbol{Y}) \leq 0$ for all $j \in M$, $\partial G_i(\boldsymbol{Y})$ is decreasing in Y_j for all $j \in M$. In this case, having $Y_j(\hat{p}_j) \leq Y_j(\tilde{p}_j)$ for all $j \in M$ and $Y_j(\hat{p}_j) = Y_j(\tilde{p}_j)$ for all $j \in N \setminus M$ implies that $\partial G_i(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \partial G_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$. Similarly, since the function G satisfies $\partial G_j(\boldsymbol{Y}) \geq 0$ for all $j \in N$, $G(\boldsymbol{Y})$ is increasing Y_j for all $j \in N$, in which case, using the fact that $Y_j(\hat{p}_j) \leq Y_j(\tilde{p}_j)$ for all $j \in M$ and $Y_j(\hat{p}_j) = Y_j(\tilde{p}_j)$ for all $j \in N \setminus M$, we obtain $G(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \leq G(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$. Lastly, since $i \in N \setminus M$, we have $Y_i(\hat{p}_i) = Y_i(\tilde{p}_i)$. Because $\partial G_i(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \partial G_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$ and $G(Y(\hat{\boldsymbol{p}})) \leq G(Y(\tilde{\boldsymbol{p}}))$, using the definition of the choice probability under the GEV model, we get

$$\Theta_i(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \;=\; \frac{Y_i(\hat{p}_i)\,\partial G_i(\boldsymbol{Y}(\hat{\boldsymbol{p}}))}{1 + G(\boldsymbol{Y}(\hat{\boldsymbol{p}}))} \;\geq\; \frac{Y_i(\tilde{p}_i)\,\partial G_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))}{1 + G(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))} \;=\; \Theta_i(\boldsymbol{Y}(\tilde{\boldsymbol{p}})).$$

Second, we show that $\Theta_0(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \geq \Theta_0(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$. By the discussion at the beginning of the proof, we have $G(\boldsymbol{Y}(\hat{\boldsymbol{p}})) \leq G(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$. The no-purchase probability at prices $\boldsymbol{p}$ is given by

$$\Theta_0(\boldsymbol{Y}(\boldsymbol{p})) = 1 - \sum_{i \in N} \Theta_i(\boldsymbol{Y}(\boldsymbol{p})) \;=\; 1 - \frac{\sum_{i \in N} Y_i(p_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{p}))}{1 + G(\boldsymbol{Y}(\boldsymbol{p}))} \;=\; \frac{1}{1 + G(\boldsymbol{Y}(\boldsymbol{p}))},$$

where the last equality above is by Lemma 2.2. In this case, since $G(Y(\hat{\boldsymbol{p}})) \leq G(Y(\tilde{\boldsymbol{p}}))$, we obtain $\Theta_0(\boldsymbol{Y}(\hat{\boldsymbol{p}})) = 1/(1 + G(\boldsymbol{Y}(\hat{\boldsymbol{p}}))) \geq 1/(1 + G(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))) = \Theta_0(\boldsymbol{Y}(\tilde{\boldsymbol{p}}))$. $\square$

In the next proposition, we use the lemma above to show that the optimal prices in the UNCONSTRAINED problem are finite.

Proposition A.2 *There exists an optimal solution $\boldsymbol{p}^*$ to the* UNCONSTRAINED *problem such that $p_i^* \in [c_i, \infty)$ for all $i \in N$.*

Proof: Let $\boldsymbol{p}^*$ be an optimal solution to the UNCONSTRAINED problem and set $N^- = \{i \in N : p_i^* < c_i\}$ to capture the set of products whose prices are below their unit costs. We define the prices $\boldsymbol{p}^-$ as

$p_i^- = c_i$ for all $i \in N^-$ and $p_i^- = p_i^*$ for all $i \in N \setminus N^-$. Note that we have $p_i^- \geq p_i^*$ for all $i \in N^-$ and $p_i^- = p_i^*$ for all $i \in N \setminus N^-$, in which case, by Lemma A.1, we get $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^-)) \geq \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*))$ for all $i \in N \setminus N^-$. Therefore, we have $\sum_{i\in N}(p_i^* - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \leq \sum_{i\in N\setminus N^-}(p_i^* - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \leq \sum_{i\in N\setminus N^-}(p_i^- - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^-)) = \sum_{i\in N}(p_i^- - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^-))$, where the first inequality uses the fact that $p_i^* < c_i$ for all $i \in N^-$, the second inequality uses the fact that $p_i^- = p_i^* \geq c_i$ and $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^-)) \geq \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*))$ for all $i \in N \setminus N^-$, and the last equality uses the fact that $p_i^- = c_i$ for all $i \in N^-$. By the last chain of inequalities, the objective value of the UNCONSTRAINED problem corresponding to the prices $\boldsymbol{p}^-$ is at least as large as the one corresponding to the prices $\boldsymbol{p}^*$, which implies that there exists an optimal solution $\boldsymbol{p}^*$ such that $p_i^* \geq c_i$ for all $i \in N$. Thus, there exists an optimal solution such that the prices of all of the products are lower bounded by their corresponding unit costs. In the rest of the proof, we can assume that $p_i^* \geq c_i$ for all $i \in N$.

Let $N^+ = \{i \in N : p_i^* = \infty\}$ to capture the set of products whose prices are infinite. Noting that $Y_i(p_i) = \exp(\alpha_i - \beta\, p_i)$, we have $\lim_{p_i\to\infty} Y_i(p_i) = 0$ and $\lim_{p_i\to\infty} p_i\, Y_i(p_i) = 0$, in which case, by the definition of the purchase probabilities in (1), we have $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = 0$ and $p_i^*\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = 0$ for all $i \in N^+$. Letting $\bar{p} = \max\{p_i^* : i \in N \setminus N^+\} < \infty$, define the prices $\boldsymbol{p}^+$ as $p_i^+ = \bar{p} + c_i$ for all $i \in N^+$ and $p_i^+ = p_i^*$ for all $i \in N \setminus N^+$. Note that we have $p_i^* \geq p_i^+$ for all $i \in N^+$ and $p_i^* = p_i^+$ for all $i \in N \setminus N^+$. In this case, by Lemma A.1, we obtain $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \geq \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+))$ for all $i \in N \setminus N^+$ and $\Theta_0(\boldsymbol{Y}(\boldsymbol{p}^*)) \geq \Theta_0(\boldsymbol{Y}(\boldsymbol{p}^+))$. Using the last two inequalities, we get $\sum_{i\in N^+}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) + \sum_{i\in N\setminus N^+}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) = \sum_{i\in N}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) = 1 - \Theta_0(\boldsymbol{Y}(\boldsymbol{p}^+)) \geq 1 - \Theta_0(\boldsymbol{Y}(\boldsymbol{p}^*)) = \sum_{i\in N}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = \sum_{i\in N\setminus N^+}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*))$, where the last equality uses the fact that $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = 0$ for all $i \in N^+$. Focusing on the first and last expressions in the last chain of inequalities, we have $\sum_{i\in N^+}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) \geq \sum_{i\in N\setminus N^+}(\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) - \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)))$, which implies that

$$
\begin{aligned}
\sum_{i\in N}(p_i^* - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \;&=\; \sum_{i\in N\setminus N^+}(p_i^* - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \\
&= \sum_{i\in N\setminus N^+}(p_i^+ - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) \;+\; \sum_{i\in N\setminus N^+}(p_i^* - c_i)\,(\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) - \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+))) \\
&\leq \sum_{i\in N\setminus N^+}(p_i^+ - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) \;+\; \sum_{i\in N\setminus N^+}\bar{p}\,(\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) - \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+))) \\
&\leq \sum_{i\in N\setminus N^+}(p_i^+ - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) \;+\; \sum_{i\in N^+}\bar{p}\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) = \sum_{i\in N}(p_i^+ - c_i)\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)),
\end{aligned}
$$

where the first equality holds because $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = 0$ and $p_i^*\,\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) = 0$ for all $i \in N^+$, the first inequality is by the fact that $\bar{p} \geq p_i^*$ for all $i \in N \setminus N^+$ and $\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) \geq \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+))$ for all $i \in N \setminus N^+$, and the second inequality holds as $\sum_{i\in N^+}\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)) \geq \sum_{i\in N\setminus N^+}(\Theta_i(\boldsymbol{Y}(\boldsymbol{p}^*)) - \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^+)))$. Thus, the expected profit at the prices $\boldsymbol{p}^+$ is at least as large as the one at the prices $\boldsymbol{p}^*$, in which case, there exists an optimal solution $\boldsymbol{p}^*$ such that $p_i^* < \infty$ for all $i \in N$. □

Appendix B: Proof of Corollary 3.3

As a function of the unit product costs $\boldsymbol{c}$, let $R^*(\boldsymbol{c})$ be the optimal expected profit in the Unconstrained problem. We claim that $R^*(\boldsymbol{c}) \geq R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i) \geq R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}) \geq R^*(\boldsymbol{c})-\delta$. Note that the chain of inequalities $R^*(\boldsymbol{c}) \geq R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i) \geq R^*(\boldsymbol{c}+\delta\,\boldsymbol{e})$ follows immediately because we have $\boldsymbol{c} \leq \boldsymbol{c}+\delta\,\boldsymbol{e}_i \leq \boldsymbol{c}+\delta\,\boldsymbol{e}$, and as the unit costs increase, the optimal expected profits decrease. To establish that $R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}) \geq R^*(\boldsymbol{c})-\delta$, observe that

$$\begin{aligned} R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}) &= \sum_{i\in N} (p_i^*(\boldsymbol{c}+\delta\,\boldsymbol{e})-c_i-\delta)\,\Theta_i(\boldsymbol{p}^*(\boldsymbol{c}+\delta\,\boldsymbol{e})) \;\geq\; \sum_{i\in N}(p_i^*(\boldsymbol{c})-c_i-\delta)\,\Theta_i(\boldsymbol{p}^*(\boldsymbol{c})) \\ &= \sum_{i\in N}(p_i^*(\boldsymbol{c})-c_i)\,\Theta_i(\boldsymbol{p}^*(\boldsymbol{c}))-\delta\sum_{i\in N}\Theta_i(\boldsymbol{p}^*(\boldsymbol{c})) \;\geq\; R^*(\boldsymbol{c})-\delta, \end{aligned}$$

where the first inequality follows because $\boldsymbol{p}^*(\boldsymbol{c})$ may not be optimal when the unit costs are $\boldsymbol{c}\,+\,\delta\boldsymbol{e}$ and the last inequality follows from the fact that $R^*(\boldsymbol{c})=\sum_{i\in N}(p_i^*(\boldsymbol{c})-c_i)\,\Theta_i(\boldsymbol{p}^*(\boldsymbol{c}))$ and $\sum_{i\in N}\Theta_i(\boldsymbol{p}^*(\boldsymbol{c}))\leq 1$. Thus, the claim holds. To show the first part of the corollary, as a function of the unit costs of the products, we let $m^*(\boldsymbol{c})$ be the optimal markup in the Unconstrained problem. Noting that $R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i) \geq R^*(\boldsymbol{c})-\delta$ by the discussion at the beginning of the proof, by Theorem 3.1, we obtain $m^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)=1/\beta+R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)\geq 1/\beta+R^*(\boldsymbol{c})-\delta=m^*(\boldsymbol{c})-\delta$, which implies that $p_i^*(\boldsymbol{c}+\delta\,e_i)=m^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)+c_i+\delta\geq m^*(\boldsymbol{c})+c_i=p_i^*(\boldsymbol{c})$. Similarly, we also have $m^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)=1/\beta+R^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)\leq 1/\beta+R^*(\boldsymbol{c})=m^*(\boldsymbol{c})$. Thus, for all $j\neq i$, we get $p_j^*(\boldsymbol{c}+\delta\,e_i)=m^*(\boldsymbol{c}+\delta\,\boldsymbol{e}_i)+c_j\leq m^*(\boldsymbol{c})+c_j=p_j^*(\boldsymbol{c})$, which completes the first part. Considering the second part of the corollary, by Theorem 3.1 and the discussion at the beginning of the proof, we have $m^*(\boldsymbol{c}+\delta\,\boldsymbol{e})=1/\beta+R^*(\boldsymbol{c}+\delta\,\boldsymbol{e})\geq 1/\beta+R^*(\boldsymbol{c})-\delta=m^*(\boldsymbol{c})-\delta$. Thus, we get $p_i^*(\boldsymbol{c}+\delta\,\boldsymbol{e})=m_i^*(\boldsymbol{c}+\delta\,\boldsymbol{e})+c_i+\delta\geq m_i^*(\boldsymbol{c})+c_i=p_i^*(\boldsymbol{c})$, which completes the second part.

Appendix C: Optimal Markup Under Separable Generating Functions

We consider the Unconstrained problem when the products are partitioned into disjoint subsets, the generating function is separable by the partitions, and the products in each partition share the same price sensitivity parameter. We partition the set of products N into the subsets $N^1,\ldots,N^m$ such that $N=\cup_{k=1}^m N^k$ and $N^k\cap N^{k'}=\varnothing$ for $k\neq k'$. Similarly, we partition the vector $\boldsymbol{Y}=(Y_1,\ldots,Y_n)$ into the subvectors $\boldsymbol{Y}^1,\ldots,\boldsymbol{Y}^m$ such that each subvector $\boldsymbol{Y}^k$ is given by $\boldsymbol{Y}^k=(Y_i\,:\,i\in N^k)$. We assume that the generating function G is separable by the partitions so that $G(\boldsymbol{Y})=\sum_{k=1}^m G^k(\boldsymbol{Y}^k)$, where the functions $G^1,\ldots,G^m$ satisfy the four properties discussed at the beginning of Section 2. Furthermore, we assume that the products in each partition N^k share the same price sensitivity β^k. In the next theorem, we show that the optimal prices for the products in each partition have a constant markup and we give a formula to compute the optimal markup. In this theorem, we let $\boldsymbol{Y}^k(\boldsymbol{p})=(Y_i(p_i)\,:\,i\in N^k)=(e^{\alpha_i-\beta^k p_i}\,:\,i\in N^k)$, $\boldsymbol{p}^*$ be the optimal solution to the Unconstrained problem, and $\boldsymbol{c}=(c_1,\ldots,c_n)$ be the vector of unit product costs.

Theorem C.1 *For all $i \in N^k$, $p_i^* - c_i = \frac{1}{\beta^k} + R(\boldsymbol{p}^*)$, so that the products in N^k have a constant markup. Furthermore, letting R^* be the unique value of R satisfying*

$$R = \sum_{k=1}^{m} \frac{1}{\beta^k}\, e^{-(\beta^k R+1)}\, G^k(\boldsymbol{Y}^k(\boldsymbol{c})),$$

the optimal expected profit in the UNCONSTRAINED *problem is $R(\boldsymbol{p}^*) = R^*$, in which case, the optimal markup for the products in N^k is $p_i^* - c_i = \frac{1}{\beta^k} + R(\boldsymbol{p}^*) = \frac{1}{\beta^k} + R^*$.*

Proof: For notational brevity, let $R^k(\boldsymbol{p}) = \sum_{j\in N^k}(p_j - c_j)\, Y_j(p_j)\, \partial G_j^k(\boldsymbol{Y}^k(\boldsymbol{p}))/(1+G(\boldsymbol{Y}(\boldsymbol{p})))$, so that $R(\boldsymbol{p}) = \sum_{k=1}^m R^k(\boldsymbol{p})$. Fix product i. Let k be such that $i \in N^k$. Noting that the definition of $R^k(\boldsymbol{p})$ is similar to that of $R(\boldsymbol{p})$, following precisely the same approach that we use right before Theorem 3.1, we can verify that

$$\frac{\partial R^k(\boldsymbol{p})}{\partial p_i} = \beta^k\, \frac{Y_i(p_i)\, \partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))} \left\{ \frac{1}{\beta^k} - (p_i - c_i) - \sum_{j\in N^k} (p_j - c_j)\, \frac{Y_j(p_j)\, \partial G_{ji}^k(\boldsymbol{Y}^k(\boldsymbol{p}))}{\partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))} + R^k(\boldsymbol{p}) \right\}.$$

One the other hand, consider any ℓ such that $i \notin N^\ell$. In the definition of $R^\ell(\boldsymbol{p})$, which is given by $\sum_{j\in N^\ell}(p_j - c_j)\, Y_j(p_j)\, \partial G_j^\ell(\boldsymbol{Y}^\ell(\boldsymbol{p}))/(1+G(\boldsymbol{Y}(\boldsymbol{p})))$, since $i \notin N^\ell$, only the expression $1+G(\boldsymbol{Y}(\boldsymbol{p}))$ in the denominator depends on the price of product i. In this case, differentiating $R^\ell(\boldsymbol{p})$ with respect to p_i, it follows that

$$\frac{\partial R^\ell(\boldsymbol{p})}{\partial p_i} = \sum_{j\in N^\ell} (p_j - c_j)\, \frac{Y_j(p_j)\, \partial G_j^\ell(\boldsymbol{Y}^\ell(\boldsymbol{p}))}{(1+G(\boldsymbol{Y}(\boldsymbol{p})))^2}\, \partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))\, \beta^k\, Y_i(p_i) = \beta^k\, \frac{Y_i(p_i)\, \partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))}\, R^\ell(\boldsymbol{p}).$$

Since $R(\boldsymbol{p}) = R^k(\boldsymbol{p}) + \sum_{\ell\neq k} R^\ell(\boldsymbol{p})$, we have $\partial R(\boldsymbol{p})/\partial p_i = \partial R^k(\boldsymbol{p})/\partial p_i + \sum_{\ell \neq k} \partial R^\ell(\boldsymbol{p})/\partial p_i$. Therefore, the two equalities above yield

$$\frac{\partial R(\boldsymbol{p})}{\partial p_i} = \beta^k\, \frac{Y_i(p_i)\, \partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))}{1+G(\boldsymbol{Y}(\boldsymbol{p}))} \left\{ \frac{1}{\beta^k} - (p_i - c_i) - \sum_{j\in N^k} (p_j - c_j)\, \frac{Y_j(p_j)\, \partial G_{ji}^k(\boldsymbol{Y}^k(\boldsymbol{p}))}{\partial G_i^k(\boldsymbol{Y}^k(\boldsymbol{p}))} + R(\boldsymbol{p}) \right\}.$$

Once we have the expression above for the derivative of the expected profit function, we can follow precisely the same argument in the proof of Theorem 3.1 to get $p_i^* - c_i = \frac{1}{\beta^k} + R(\boldsymbol{p}^*)$.

The right side of the equation in the theorem is decreasing in R, whereas the left side is strictly increasing. Furthermore, the right side of the equation evaluated at $R=0$ is non-negative, whereas the left side evaluated at $R=0$ is zero. Thus, there exists a unique value of R satisfying the equation in the theorem. We proceed to showing that the value of R satisfying this equation corresponds to the optimal expected profit in the UNCONSTRAINED problem. The optimal price of each product $i \in N^k$ is of the form $p_i^* = c_i + \frac{1}{\beta^k} + R(\boldsymbol{p}^*)$, so that we have $Y_i(p_i^*) = e^{\alpha_i - \beta^k\,(c_i + \frac{1}{\beta^k} + R(\boldsymbol{p}^*))} = e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, Y_i(c_i)$. In this case, noting that $p_i^* - c_i = \frac{1}{\beta^k} + R^*(\boldsymbol{p}^*)$ for all $i \in N^k$ and plugging the optimal prices into the expected profit function in the UNCONSTRAINED problem, the optimal expected profit satisfies the equation

$$
\begin{aligned}
R(\boldsymbol{p}^*) &= \frac{\sum_{k=1}^m \sum_{i\in N^k} \left(\frac{1}{\beta^k} + R(\boldsymbol{p}^*)\right) e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, Y_i(c_i)\, \partial G_i^k(e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, \boldsymbol{Y}^k(\boldsymbol{c}))}{1+\sum_{k=1}^m G^k(e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, \boldsymbol{Y}^k(\boldsymbol{c}))} \\
&= \frac{\sum_{k=1}^m \left(\frac{1}{\beta^k} + R(\boldsymbol{p}^*)\right) G^k(e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, \boldsymbol{Y}^k(\boldsymbol{c}))}{1+\sum_{k=1}^m G^k(e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, \boldsymbol{Y}^k(\boldsymbol{c}))} \\
&= \frac{\sum_{k=1}^m \left(\frac{1}{\beta^k} + R(\boldsymbol{p}^*)\right) e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, G^k(\boldsymbol{Y}^k(\boldsymbol{c}))}{1+\sum_{k=1}^m e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, G^k(\boldsymbol{Y}^k(\boldsymbol{c}))},
\end{aligned}
$$

where the second equality uses Lemma 2.2 and G^k is a generating function, and the third equality uses the fact that $G^1,\dots,G^m$ satisfy the properties of a generating function so that these functions are homogeneous of degree one. Focusing on the first and last expressions in the chain of equalities above and rearranging the terms, we get $R(\boldsymbol{p}^*) = \sum_{k=1}^m \frac{1}{\beta^k}\, e^{-(\beta^k R(\boldsymbol{p}^*)+1)}\, G^k(\boldsymbol{Y}^k(\boldsymbol{c}))$. $\square$

There are instances of existing GEV models with a separable generating function. Under the nested logit model, the generating function is $G(\boldsymbol{Y}) = \sum_{k=1}^m (\sum_{j\in N^k} Y_j^{1/\lambda^k})^{\lambda^k}$, where $(\lambda^1,\dots,\lambda^m)$ are constants, each in the interval $(0,1]$. The products in each partition N^k correspond to the products in a particular nest. Letting $G^k(\boldsymbol{Y}^k) = (\sum_{j\in N^k} Y_j^{1/\lambda^k})^{\lambda^k}$, this generating function is a separable function of the form $G(\boldsymbol{Y}) = \sum_{k=1}^m G^k(\boldsymbol{Y}^k)$. Also, under the d-level nested logit model, the products are organized in a tree. The tree starts at a root node. The products are at the leaf nodes. The degree of overlap between the paths from the root node to the leaf nodes corresponding to different products captures the degree of substitution between the different products. Figure 2 shows a sample tree. Li et al. (2015) show that the generating function under the d-level nested logit model is of the form $G(\boldsymbol{Y}) = \sum_{k=1}^m G^k(\boldsymbol{Y}^k)$, where $1,\dots,m$ correspond to the children nodes of the root node, and the partition N^k corresponds to the subset of products such that if $i \in N^k$, then the path from the root node to the leaf node corresponding to product i passes through the k-th child node of the root node. Thus, if the products in each partition N^k share the same price sensitivity, then the optimal prices for these products have a constant markup. To our knowledge, this result was not known earlier. Lastly, we note that if G is a generating function, then for $\theta > 0$, θG is a generating function as well. Therefore, Theorem C.1 applies when G is a separable function of the form $G(\boldsymbol{Y}) = \sum_{k=1}^m \theta^k G^k(\boldsymbol{Y}^k)$ for positive scalars $\theta^1,\dots,\theta^m$.

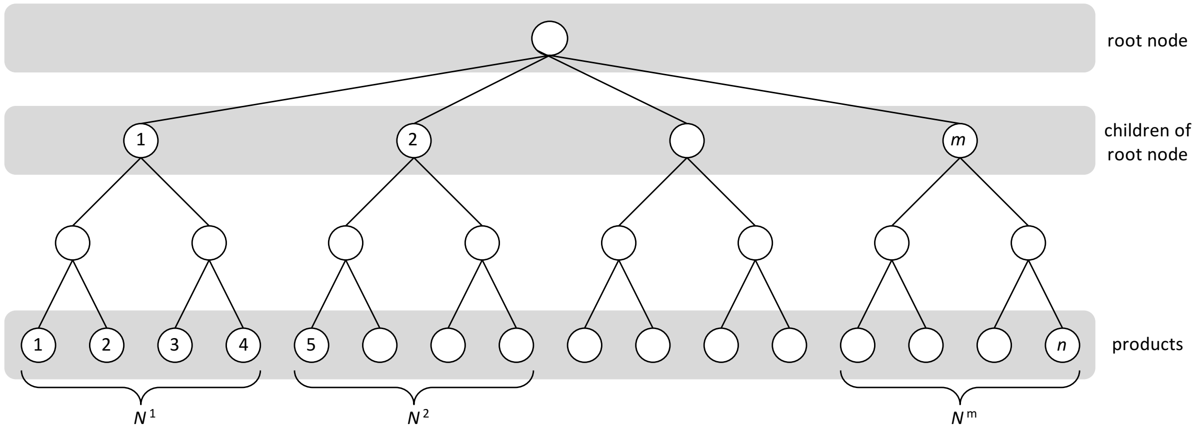

Figure 2 A sample tree for the d-level nested logit model with $n = 16$ and $m = 4$.

Appendix D: Proof of Lemma 4.2

Note that $Y_i(s_i) = e^{\alpha_i - \beta s_i}$, so $dY_i(s_i)/ds_i = -\beta\, Y_i(s_i)$. In this case, using the definition of f and differentiating, for all $i \in N$, we obtain

$$\frac{\partial f(\boldsymbol{s})}{\partial s_i} \;=\; -\frac{Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))}{1+G(\boldsymbol{Y}(\boldsymbol{s}))} + q_i \;=\; q_i - \Theta_i\left(\boldsymbol{Y}(\boldsymbol{s})\right),$$

where the second equality uses the definition of the selection probabilities in (1). The equality above establishes the expression for the gradient f. Letting $\mathbb{1}_{\{\cdot\}}$ be the indicator function and differentiating the middle expression above once more, we get

$$\frac{1}{\beta}\frac{\partial^2 f(\boldsymbol{s})}{\partial s_i \partial s_j} \;=\; \frac{\mathbb{1}_{\{i=j\}}\, Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))}{1+G(\boldsymbol{Y}(\boldsymbol{s}))} + \frac{Y_i(s_i)\,\partial G_{ij}(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)}{1+G(\boldsymbol{Y}(\boldsymbol{s}))} - \frac{Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))\,\partial G_j(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)}{(1+G(\boldsymbol{Y}(\boldsymbol{s})))^2},$$

where we use the fact that the derivative of $h(x)/g(x)$ with respect to x is given by the formula $h'(x)/g(x) - h(x)\,g'(x)/g(x)^2$. By (1), we have

$$\frac{\mathbb{1}_{\{i=j\}}\, Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))}{1+G(\boldsymbol{Y}(\boldsymbol{s}))} = \mathbb{1}_{\{i=j\}}\,\Theta_i(\boldsymbol{Y}(\boldsymbol{s}))\,, \quad \frac{Y_i(s_i)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s}))\,\partial G_j(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)}{(1+G(\boldsymbol{Y}(\boldsymbol{s})))^2} = \Theta_i(\boldsymbol{Y}(\boldsymbol{s}))\,\Theta_j(\boldsymbol{Y}(\boldsymbol{s})),$$

and $Y_i(s_i)\,\partial G_{ij}(\boldsymbol{Y}(\boldsymbol{s}))\,Y_j(s_j)$ is the (i,j)-th entry of the matrix $\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\,\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)$. Putting everything together, we have

$$\frac{1}{\beta}\nabla^2 f(\boldsymbol{s}) = \mathrm{diag}\left(\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\right) + \frac{\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{s}))\,\mathrm{diag}\left(\boldsymbol{Y}(\boldsymbol{s})\right)}{1+G(\boldsymbol{Y}(\boldsymbol{s}))} - \boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))\,\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{s}))^\top\,,$$

which is the desired result.

Appendix E: Product Prices to Achieve Given Market Shares

For fixed $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all $i \in N$ and $\sum_{i \in N} q_i \;<\; 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by: for all $\boldsymbol{s} \in \mathbb{R}^n$, $f(\boldsymbol{s}) = \frac{1}{\beta}\,\log(1 + G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i \in N} q_i\, s_i$. In the next lemma, we show that the value $f(\boldsymbol{s})$ gets arbitrarily large as the norm of $\boldsymbol{s}$ gets arbitrarily large.

Lemma E.1 *Given any $L \geq 0$, there exists an $M \geq 0$, such if $\|\boldsymbol{s}\| \geq M$, then $f(\boldsymbol{s}) \geq L$.*

Proof: Let $\Delta = \min_{i \in N}\{Y_i(0)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{0}))\}$. Since $Y_i(0) > 0$ for all $i \in N$, noting the assumption that $\partial G_i(\boldsymbol{Y}) > 0$ when $Y_i > 0$ for all $i \in N$, we have $\Delta > 0$. Given any $L \geq 0$, we choose $M \geq 0$ such that it satisfies the equality

$$L = \min\left\{\frac{1}{\beta}\log\Delta + \min_{i\in N}\{q_i\}\left(1 - \sum_{i\in N} q_i\right)\frac{M}{2\sqrt{n}},\ \min_{i\in N}\{q_i\}\,\frac{M}{2\sqrt{n}}\right\}.$$

The right side of the equality above evaluated at $M = 0$ is non-positive. Furthermore, the right side is strictly increasing in M. Therefore, for any $L \geq 0$, there exists $M \geq 0$ that satisfies the equality above. We show that if $\|\boldsymbol{s}\| \geq M$, then $f(s) \geq L$. Consider two cases.

Case 1: Assume that there exists $j \in N$ such that $s_j \geq \frac{M}{\sqrt{n}}$, and we have $s_i \geq -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}}$ for all $i \in N$. In this case, since $\log(1 + G(\boldsymbol{Y}(\boldsymbol{s}))) \geq 0$, we obtain

$$\begin{aligned}
f(\boldsymbol{s}) &= \frac{1}{\beta}\log(1+G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i\in N} q_i\,s_i \;\geq\; \sum_{i\in N\setminus\{j\}} q_i\,s_i + q_j\,s_j \\
&\geq -\left(\sum_{i\in N\setminus\{j\}} q_i\right)\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}} + q_j\,\frac{M}{\sqrt{n}} \;\geq\; -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}} + \min_{\ell\in N}\{q_\ell\}\,\frac{M}{\sqrt{n}} \\
&= \min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}} \;\geq\; L,
\end{aligned}$$

where the third inequality uses the fact that $\sum_{i\in N} q_i < 1$ and the last inequality follows from the definition of L. Thus, we have $f(\boldsymbol{s}) \geq L$, as desired.

Case 2: This is the opposite of Case 1. Assume that we have $s_i < \frac{M}{\sqrt{n}}$ for all $i \in N$, or there exists $j \in N$ such that $s_j < -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}}$. In this case, we claim that there must exist $j \in N$ such that $s_j \leq -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}}$. To see the claim, assume on the contrary, we have $s_j > -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}}$ for all $j \in N$. Thus, noting the requirements of Case 2, we must have $s_i < \frac{M}{\sqrt{n}}$ for all $i \in N$, in which case, we get $-\frac{M}{\sqrt{n}} \leq -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}} < s_j < \frac{M}{\sqrt{n}}$, so $|s_j| < \frac{M}{\sqrt{n}}$ for all j, which contradicts the fact that $\|\boldsymbol{s}\| \geq M$. So, the claim holds and there exists $j \in N$ such that $s_j \leq -\min_{\ell\in N}\{q_\ell\}\,\frac{M}{2\sqrt{n}}$.

In the rest of the proof, we let $\ell = \arg\min_{i\in N}\{s_i\}$ and $t = \min_{i\in N}\{s_i\}$ so that we have $t \leq -\min_{i\in N}\{q_i\}\,\frac{M}{2\sqrt{n}}$, $s_i \geq t$ for all $i \in N$, and $s_\ell = t$. Using Lemma 2.2, we have

$$\begin{aligned}
G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) &= \sum_{i\in N} Y_i(s_i - t)\,\partial G_i(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) \;\geq\; Y_\ell(s_\ell - t)\,\partial G_\ell(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) \\
&\geq Y_\ell(0)\,\partial G_\ell(\boldsymbol{Y}(\boldsymbol{0})) \;\geq\; \Delta.
\end{aligned}$$

In the chain of inequalities above, the first inequality holds because, for any price vector $\boldsymbol{p}$, we have $Y_i(p_i) \geq 0$ and $\partial G_i(\boldsymbol{Y}(\boldsymbol{p})) \geq 0$ for all $i \in N$. To see the second inequality, note that $Y_i(p_i)$ is decreasing in p_i and $s_i - t \geq 0$, so that $Y_i(s_i - t) \leq Y_i(0)$ for all $i \neq \ell$. Furthermore, by the properties of a generating function discussed at the beginning of Section 2, since $\partial G_{\ell i}(\boldsymbol{Y}) \leq 0$ for all $i \neq \ell$, $\partial G_\ell(\boldsymbol{Y}) \geq 0$ is decreasing in Y_i for all $i \neq \ell$. Also noting that we have $Y_\ell(s_\ell - t) = Y_\ell(0)$, we get $\partial G_\ell(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) \geq \partial G_\ell(\boldsymbol{Y}(\boldsymbol{0}))$ because $Y_i(s_i - t) \leq Y_i(0)$ for all $i \neq \ell$. The last inequality uses the definition of Δ. Thus, the chain of inequalities above yields $G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) \geq \Delta$. In addition, observe that $\log(G(\boldsymbol{Y}(\boldsymbol{s}+t\,\boldsymbol{e}))) = \log(G(e^{-\beta\, t}\,\boldsymbol{Y}(\boldsymbol{s}))) = \log(e^{-\beta\, t}G(\boldsymbol{Y}(\boldsymbol{s}))) = -\beta\, t + \log(G(\boldsymbol{Y}(\boldsymbol{s})))$, where the first equality follows from the fact that $Y_i(p_i) = e^{\alpha_i - \beta\, p_i}$ and the second equality follows from the fact that G is homogeneous of degree one. Thus, we have

$$\begin{aligned} f(\boldsymbol{s}) \; &= \; \frac{1}{\beta}\,\log(1 + G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i\in N} q_i\, s_i \; \geq \; \frac{1}{\beta}\,\log(G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e} + t\,\boldsymbol{e}))) + \sum_{i\in N} q_i\, s_i \\ &= \; \frac{1}{\beta}\Big(-\beta\, t + \log(G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})))\Big) + \sum_{i\in N} q_i\, s_i \\ &= \; \frac{1}{\beta}\,\log(G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e}))) + \sum_{i\in N} q_i(s_i - t) - \Big(1 - \sum_{i\in N} q_i\Big)\, t \\ &\geq \frac{1}{\beta}\,\log\Delta - \Big(1 - \sum_{i\in N} q_i\Big)\, t \; \geq \; \frac{1}{\beta}\,\log\Delta + \min_{i\in N}\{q_i\}\,\Big(1 - \sum_{i\in N} q_i\Big)\,\frac{M}{2\sqrt{n}} \; \geq \; L, \end{aligned}$$

where the second equality is by the fact that $\log(G(\boldsymbol{Y}(\boldsymbol{s}+t\,\boldsymbol{e}))) = -\beta\, t + \log(G(\boldsymbol{Y}(\boldsymbol{s})))$, the second inequality holds because $G(\boldsymbol{Y}(\boldsymbol{s} - t\,\boldsymbol{e})) \geq \Delta$ and $s_i - t \geq 0$ for all $i \in N$, the third inequality follows since $t \;\leq\; -\min_{i\in N}\{q_i\}\,\frac{M}{2\sqrt{n}}$, and the last inequality is by the definition of L. So, $f(\boldsymbol{s}) \geq L$. $\square$

Appendix F: Proof of Lemma 4.4

Let $f(\boldsymbol{s}) = \frac{1}{\beta}\,\log(1 + G(\boldsymbol{Y}(\boldsymbol{s}))) + \sum_{i\in N} q_i\, s_i$ denote the objective function of the optimization problem in Theorem 4.1. By Lemma 4.2, we have $\nabla f(\boldsymbol{s}) = \boldsymbol{q} - \boldsymbol{\Theta}\,(\boldsymbol{Y}(\boldsymbol{s}))$ for all $\boldsymbol{s} \in \mathbb{R}^n$. Since $\boldsymbol{p}(\boldsymbol{q})$ is the unique minimizer of f, it follows that for all $i \in N$,

$$0 = \frac{\partial f(\boldsymbol{s})}{\partial s_i}\bigg|_{\boldsymbol{s}=\boldsymbol{p}(\boldsymbol{q})} \; = \; q_i - \Theta_i(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q}))).$$

Taking the derivative of the above equation with respect to q_j, we obtain that for all $i, j \in N$,

$$0 = \mathbb{1}_{\{i=j\}} - \sum_{\ell\in N} \frac{\partial \Theta_i(\boldsymbol{Y}(\boldsymbol{s}))}{\partial s_\ell}\bigg|_{\boldsymbol{s}=\boldsymbol{p}(\boldsymbol{q})} \times \frac{\partial p_\ell(\boldsymbol{q})}{\partial q_j}.$$

Since $\nabla f(\boldsymbol{s}) = \boldsymbol{q} - \Theta(\boldsymbol{Y}(\boldsymbol{s}))$ by Lemma 4.2, we have that $\nabla^2 f(\boldsymbol{s}) \;=\; \Big(-\frac{\partial\Theta_i(\boldsymbol{Y}(\boldsymbol{s}))}{\partial s_\ell} \;:\; i, \ell \in N\Big)$, and thus, the expression $-\sum_{\ell\in N} \frac{\partial\Theta_i(\boldsymbol{Y}(\boldsymbol{s}))}{\partial s_\ell}\Big|_{\boldsymbol{s}=\boldsymbol{p}(\boldsymbol{q})} \times \frac{\partial p_\ell(\boldsymbol{q})}{\partial q_j}$ is the inner product between the i-th row of the Hessian matrix $\nabla^2 f(\boldsymbol{p}(\boldsymbol{q}))$ and the j-th column of the Jacobian matrix $\mathbf{J}(\boldsymbol{q})$. Letting $\mathbf{I}$ denote the n-by-n identity matrix, we can write the above system of equations in matrix notation as

$$\mathbf{0} = \mathbf{I} + \nabla^2 f\,(\boldsymbol{p}(\boldsymbol{q}))\,\mathbf{J}(\boldsymbol{q}).$$

Note that, by definition, $\boldsymbol{\Theta}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q}))) = \boldsymbol{q}$, and thus, it follows from Lemma 4.2 that

$$\begin{aligned}\nabla^2 f(\boldsymbol{p}(\boldsymbol{q})) &= \beta\left(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \frac{\mathrm{diag}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))\,\nabla^2 G(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))\,\mathrm{diag}(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))}{1+G(\boldsymbol{Y}(\boldsymbol{p}(\boldsymbol{q})))}\right)\\ &= \beta\,(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q})),\end{aligned}$$

where the last equality follows from the definition of $\mathbf{B}(\boldsymbol{q})$. Since $\nabla^2 f(\cdot)$ is positive definite, and thus, invertible, the desired result follows by the fact that $\mathbf{0} = \mathbf{I} + \nabla^2 f(\boldsymbol{p}(\boldsymbol{q}))\,\mathbf{J}(\boldsymbol{q})$.

Appendix G: Numerical Study for Constrained Pricing

We provide a numerical study to understand how efficiently we can solve instances of the Constrained problem by using our results in Section 4. In our numerical study, we consider the network revenue management setting, where we have a set of resources with limited inventories and the sale of a product consumes a combination of resources. We index the set of resources by M. The sale of product i consumes $a_{\ell i}$ units of resource i. There are C_ℓ units of resource ℓ. The expected number of customer arrivals over the selling horizon is T. The goal is to find the prices to charge for the products to maximize the expected revenue obtained from each customer, while making sure that the expected consumption of each resource does not exceed its inventory. As discussed at the beginning of Section 4, we can formulate this problem as an instance of the Constrained problem, where the function F_ℓ is given by $F_\ell(\boldsymbol{q}) = \sum_{i\in N} a_{\ell i}\,T\,q_i - C_\ell$. However, the objective function of the Constrained problem is not necessarily concave in the prices $\boldsymbol{p}$ and its feasible space is not necessarily convex. To obtain the optimal prices $\boldsymbol{p}^*$ in the Constrained problem, we solve the Market-Share-Based formulation to obtain the optimal purchase probabilities $\boldsymbol{q}^*$. In Theorem 4.3, we show that the objective function of the Market-Share-Based formulation is concave in the purchase probabilities, and we give an expression that can be used to compute the gradient of the objective function. With the function F_ℓ as given above, the constraints in the Market-Share-Based formulation are linear in the purchase probabilities. Therefore, we can obtain the optimal purchase probabilities by solving the Market-Share-Based formulation through standard convex optimization methods. In our numerical study, we use the primal-dual algorithm for convex programs; see Chapter 11 in Boyd and Vandenberghe (2004). In Theorem 4.1, we show how to compute the prices that achieve given purchase probabilities. Once we obtain the optimal purchase probabilities $\boldsymbol{q}^*$ through the Market-Share-Based formulation, we compute the optimal prices $\boldsymbol{p}^* = \boldsymbol{p}(\boldsymbol{q}^*)$ that achieve the optimal purchase probabilities. In our numerical study, we generate a large number of test problems with varying sizes. We begin by describing how we generate our test problems. Next, we give our numerical results.

Generation of Test Problems: We generate our test problems as follows. In all of our test problems, the choice process of the customers is governed by the paired combinatorial logit model, whose generating function G is given by $G(\boldsymbol{Y}) = \sum_{(i,j)\in N^2: i\neq j} \left(Y_i^{1/\tau_{(i,j)}} + Y_j^{1/\tau_{(i,j)}}\right)^{\tau_{(i,j)}}$. This generating function differs from the one corresponding to the paired combinatorial logit model in Example 2.1 by a constant factor, but if we multiply a generating function by a constant, then the generating function still satisfies the properties at the beginning of Section 2. To come up with the parameters of the paired combinatorial logit model, we sample $\tau_{(i,j)}$ from the uniform distribution over $[0.2, 1]$ for all $i, j \in N$ with $i \neq j$. Recalling that the deterministic utility component for product i is $\alpha_i - \beta\, p_i$, to come up with $(\alpha_1, \ldots, \alpha_n)$ and β, we sample α_i from the uniform distribution over $[-2, 2]$ for all $i \in N$, and we sample β from the uniform distribution over $[1, 4]$.

We index the set of resources by $M = \{1, \ldots, m\}$. For each product i, we randomly choose a resource ν_i and set $a_{\nu_i, i} = 1$. For each other resource $\ell \in M \setminus \{\nu_i\}$, we set $a_{\ell i} = 1$ with probability ζ and $a_{\ell i} = 0$ with probability $1 - \zeta$, where ζ is a parameter that we vary. In this way, the expected number of resources used by a product is given by $1 + (m-1)\,\zeta$, and we vary ζ to control the expected number of resources used by a product. To come up with the capacities for the resources, we solve the unconstrained pricing problem $\max_{\boldsymbol{p}\in\mathbb{R}^n} \{\sum_{i\in N} p_i\, \Theta_i(\boldsymbol{Y}(\boldsymbol{p}))\}$. Using $\boldsymbol{p}^{UNC}$ to denote an optimal solution to the unconstrained problem, we set the capacity of resource ℓ as $C_\ell = \kappa \sum_{i\in N} T\, a_{\ell i}\, \Theta_i(\boldsymbol{Y}(\boldsymbol{p}^{UNC}))$, where κ is another parameter that we vary to control the tightness of the resource capacities. Thus, the capacity of resource ℓ is a κ fraction of the total expected capacity consumed when we charge the optimal prices $\boldsymbol{p}^{UNC}$ in the unconstrained problem. We fix the expected number of customers at $T = 100$.

In our test problems, we vary the number of resources m and the number of products n over $(m, n) \in \{(20, 50), (20, 100), (40, 50), (40, 100), (60, 100), (60, 200), (80, 100), (80, 200)\}$, yielding a total of eight combinations. We vary the parameters ζ and κ over $\zeta \in \{0.02, 0.2\}$ and $\kappa \in \{0.5, 0.8\}$. This setup yields 32 parameter combinations for (m, n, ζ, κ). In each parameter combination, we generate 100 test problems by using the approach described above.

Numerical Results: In Table 1, we show the average CPU seconds required to solve our test problems, where the average is taken over all 100 test problems in a parameter combination. The first column in the table shows the parameters of the test problem by using the tuple (m, n, ζ, κ). The second column shows the average CPU seconds. Our numerical results indicate that we can solve even the largest test problems with $n = 200$ products and $m = 80$ resources within 20 seconds on average. Over all of our test problems, the average CPU seconds is 5.44. Since the Market-Share-Based formulation is a convex program, our approach always obtains the optimal purchase probabilities and the corresponding optimal prices. To give some context to our numerical results, we also tried computing the optimal prices by directly solving the

Param. Combination (m, n, ζ, κ)	CPU Secs.	Param. Combination (m, n, ζ, κ)	CPU Secs.
(20, 50, 0.02, 0.5)	1.33	(60, 100, 0.02, 0.5)	4.23
(20, 50, 0.02, 0.8)	1.04	(60, 100, 0.02, 0.8)	3.32
(20, 50, 0.2, 0.5)	1.95	(60, 100, 0.2, 0.5)	6.21
(20, 50, 0.2, 0.8)	1.32	(60, 100, 0.2, 0.8)	4.98
(20, 100, 0.02, 0.5)	1.83	(60, 200, 0.02, 0.5)	8.27
(20, 100, 0.02, 0.8)	1.54	(60, 200, 0.02, 0.8)	6.64
(20, 100, 0.2, 0.5)	3.5	(60, 200, 0.2, 0.5)	12.62
(20, 100, 0.2, 0.8)	2.11	(60, 200, 0.2, 0.8)	9.72
(40, 50, 0.02, 0.5)	1.88	(80, 100, 0.02, 0.5)	6.81
(40, 50, 0.02, 0.8)	0.98	(80, 100, 0.02, 0.8)	4.31
(40, 50, 0.2, 0.5)	2.75	(80, 100, 0.2, 0.5)	5.58
(40, 50, 0.2, 0.8)	3.14	(80, 100, 0.2, 0.8)	5.6
(40, 100, 0.02, 0.5)	3.41	(80, 200, 0.02, 0.5)	11.54
(40, 100, 0.02, 0.8)	2.72	(80, 200, 0.02, 0.8)	10.88
(40, 100, 0.2, 0.5)	6.24	(80, 200, 0.2, 0.5)	18.3
(40, 100, 0.2, 0.8)	4.42	(80, 200, 0.2, 0.8)	15.08

Table 1 Numerical results for constrained pricing.

Constrained problem through the `fmincon` routine in `Matlab`. Since the objective function of the Constrained problem may not be concave in the prices and the feasible space may not be convex, directly solving the Constrained problem may not yield the optimal prices. For economy of space, we provide only summary statistics. In 35% of our test problems, directly solving the Constrained problem yields optimality gaps of 1% or more. In 27% of our test problems, directly solving the Constrained problem yields optimality gaps of 5% or more. Over all test problems, the average CPU seconds to directly solve the Constrained problem by using the `fmincon` routine in `Matlab` is 38.28 seconds. Thus, our approach, where we solve the Market-Share-Based formulation to find the optimal purchase probabilities and compute the corresponding optimal prices, provides advantages both in terms of solution quality and CPU seconds.

Appendix H: Concavity of Expected Revenue Under Separable Generating Functions

We consider the Constrained problem under the same setup discussed in Appendix C. In particular, we partition the set of products N into the subsets $N^1, \dots, N^m$ such that $N = \cup_{k=1}^m N^k$ and $N^k \cap N^{k'} = \varnothing$ for $k \neq k'$. Similarly, we partition the vector $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ into the subvectors $\boldsymbol{Y}^1, \dots, \boldsymbol{Y}^m$ such that each subvector $\boldsymbol{Y}^k$ is given by $\boldsymbol{Y}^k = (Y_i \,:\, i \in N^k)$. We assume that the generating function G is separable by the partitions such that $G(\boldsymbol{Y}) = \sum_{k=1}^m G^k(\boldsymbol{Y}^k)$, where the functions $G^1, \dots, G^m$ satisfy the four properties discussed at the beginning of Section 2. Also, we assume that the products in each partition N^k share the same price sensitivity β^k. Observe that we can carry out a change of variables to compute the prices that achieve given purchase probabilities. In particular, we let $\hat{p}_i = \beta^k p_i$ for all $i \in N^k$ and $k = 1, \dots, m$. Furthermore, we let $\hat{Y}_i(\hat{p}_i) = e^{\alpha_i - \hat{p}_i}$ and define vector $\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}) = (\hat{Y}_1(\hat{p}_1), \ldots, \hat{Y}_n(\hat{p}_n))$. For given purchase probabilities $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all $i \in N$ and $\sum_{i \in N} q_i \;<\; 1$, we consider the problem

$$\min_{\boldsymbol{s} \in \mathbb{R}^n} \left\{ \log(1 + G(\hat{\boldsymbol{Y}}(\boldsymbol{s}))) + \sum_{i \in N} q_i \, s_i \right\}.$$

The problem above is a specialized version of the problem given in Theorem 4.1, when the price sensitivities of all of the products are one. Thus, by Theorem 4.1, the objective function of the problem above is strictly convex and there exists a finite and unique solution to this problem. Letting $\hat{\boldsymbol{p}}(\boldsymbol{q}) = (\hat{p}_1(\boldsymbol{q}), \ldots, \hat{p}_n(\boldsymbol{q}))$ be the optimal solution to the problem above, by our change of variables, it follows that if we set $p_i(\boldsymbol{q}) = \frac{1}{\beta^k} \, \hat{p}_i(\boldsymbol{q})$ for all $i \in N^k$ and $k = 1, \ldots, m$, then the prices $\boldsymbol{p}(\boldsymbol{q})$ achieve the purchase probabilities $\boldsymbol{q}$. Note that our change of variables applies even when the price sensitivities of the products are completely arbitrary. In other words, even if the price sensitivities of the products are arbitrary, given purchase probabilities $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all $i \in N$ and $\sum_{i \in N} q_i \;<\; 1$, we can compute the prices $\boldsymbol{p}(\boldsymbol{q})$ that achieve these purchase probabilities by solving a convex optimization problem. However, if the price sensitivities are arbitrary, then the expected revenue may not be concave in the purchase probabilities.

Going back to the case where the generating function is separable by the partitions and the products in a partition have the same price sensitivity parameter, let $\boldsymbol{\beta} \in \mathbb{R}^n_+$ be the vector of price sensitivities, so that $\boldsymbol{\beta} = (b_1, \ldots, b_n)$, where $b_i = \beta^k$ for all $i \in N^k$ and $k = 1, \ldots, m$. In the following theorem, which is the main result of this section, we show that if the generating function is separable by the partitions and the products in each partition share the same price sensitivity parameter, then the expected revenue function is concave in the purchase probabilities. Also, we give an expression for the gradient of the expected revenue function.

Theorem H.1 *For all $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all i and $\sum_{i \in N} q_i < 1$, the Hessian matrix $\nabla^2 R(\boldsymbol{q})$ is negative definite and*

$$\nabla R(\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{q}) - \mathrm{diag}(\boldsymbol{\beta})^{-1} \, \boldsymbol{e} - \frac{\boldsymbol{e}^\top \mathrm{diag}(\boldsymbol{\beta})^{-1} \, \boldsymbol{q}}{1 - \boldsymbol{e}^\top \boldsymbol{q}} \, \boldsymbol{e}.$$

The proof of the theorem above uses a sequence of three lemmas. We begin by defining a matrix that will ultimately be useful to characterize the Jacobian matrix of the vector-valued mapping $\boldsymbol{q} \mapsto \boldsymbol{p}(\boldsymbol{q})$. Recalling that $\hat{Y}_i(\hat{p}) = e^{\alpha_i - \hat{p}_i}$, and $\hat{\boldsymbol{p}}(\boldsymbol{q})$ is the optimal solution to the optimization problem given at the beginning of this section, we define the matrix

$$\hat{\mathbf{B}}(\boldsymbol{q}) = \frac{\mathrm{diag}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) \, \nabla^2 G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) \, \mathrm{diag}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))}{1 + G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))}.$$

The matrix $\hat{\mathbf{B}}(\boldsymbol{q})$ is a specialized version of the matrix $\mathbf{B}(\boldsymbol{q})$ defined right before Lemma 4.4, when the price sensitivities of all of the products are one. So, by the discussion in the proof of Theorem 4.3, $\hat{\mathbf{B}}(\boldsymbol{q})$ is positive semidefinite. Since $\nabla^2 G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))$ is symmetric, $\hat{\mathbf{B}}(\boldsymbol{q})$ is symmetric as well. In the following lemma, we give other useful properties of $\hat{\mathbf{B}}(\boldsymbol{q})$.

Lemma H.2 *We have* $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = 0$*,* $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})\,\boldsymbol{e} = \mathbf{0}$*, and* $\hat{\mathbf{B}}(\boldsymbol{q})\,\boldsymbol{e} = \mathbf{0}$.

Proof: Note that the (i,j)-th entry of $\hat{\mathbf{B}}(\boldsymbol{q})$ is $\hat{Y}_i(\hat{p}_i(\boldsymbol{q}))\,\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))\,\hat{Y}_j(\hat{p}_j(\boldsymbol{q}))/(1+G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))))$. Furthermore, since the generating function G is a separable function of the form $G(\boldsymbol{Y}) = \sum_{k=1}^m G^k(\boldsymbol{Y}^k)$, letting $\hat{\boldsymbol{Y}}^k(\hat{\boldsymbol{p}}(\boldsymbol{q})) = (\hat{Y}_i(\hat{p}_i(\boldsymbol{q})) : i \in N^k)$, for $i \in N^k$, we have $\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) = \partial G^k_{ij}(\hat{\boldsymbol{Y}}^k(\hat{\boldsymbol{p}}(\boldsymbol{q})))$ when $j \in N^k$, whereas $\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) = 0$ when $j \notin N^k$. In this case, noting that we use b_i to denote the price sensitivity of product i, for $i \in N^k$, the i-th entry of the vector $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e}$ is given by

$$\frac{\sum_{j\in N} \hat{Y}_i(\hat{p}_i(\boldsymbol{q}))\,\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))\,\hat{Y}_j(\hat{p}_j(\boldsymbol{q}))\,\frac{1}{b_j}}{1+G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))} = \frac{\hat{Y}_i(\hat{p}_i(\boldsymbol{q}))\sum_{j\in N^k} \partial G^k_{ij}(\hat{\boldsymbol{Y}}^k(\hat{\boldsymbol{p}}(\boldsymbol{q})))\,\hat{Y}_j(\hat{p}_j(\boldsymbol{q}))\,\frac{1}{\beta^k}}{1+G(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q})))} = 0.$$

In the equalities above, the first equality is by the fact that $\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) = \partial G^k_{ij}(\hat{\boldsymbol{Y}}^k(\hat{\boldsymbol{p}}(\boldsymbol{q})))$ when $j \in N^k$, whereas $\partial G_{ij}(\hat{\boldsymbol{Y}}(\hat{\boldsymbol{p}}(\boldsymbol{q}))) = 0$ when $j \notin N^k$, along with the fact that $b_j = \beta^k$ for all $j \in N^k$. The second equality follows from Lemma 2.2 and the fact that G^k is a generating function. Since our choice of the entry i is arbitrary, it follows that $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = \mathbf{0}$, showing the first equality in the lemma. The other two equalities in the lemma follow by using the same approach after replacing $1/b_j$, respectively, with b_j and 1 in the chain of equalities above. □

In the following lemma, we use the matrix $\hat{\mathbf{B}}(\boldsymbol{q})$ to give an expression for the Jacobian matrix $\mathbf{J}(\boldsymbol{q})$ associated with the vector-valued mapping $\boldsymbol{q} \mapsto \boldsymbol{p}(\boldsymbol{q})$, which maps any purchase probabilities $\boldsymbol{q}$ to the prices $\boldsymbol{p}(\boldsymbol{q})$ that achieve these purchase probabilities.

Lemma H.3 *The Jacobian matrix* $\mathbf{J}(\boldsymbol{q}) = \left(\frac{\partial p_i(\boldsymbol{q})}{\partial q_j} : i,j \in N\right)$ *is given by*

$$\mathbf{J}(\boldsymbol{q}) = -\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}.$$

Proof: The optimization problem given at the beginning of this section is a specialized version of the one in Theorem 4.1, when the price sensitivities of all of the products are one. Noting that the unique optimal solution to this problem is denoted by $\hat{\boldsymbol{p}}(\boldsymbol{q})$, Lemma 4.4 immediately implies that the Jacobian matrix $\hat{\mathbf{J}}(\boldsymbol{q}) = \left(\frac{\partial \hat{p}_i(\boldsymbol{q})}{\partial q_j} : i,j \in N\right)$ associated with the vector-valued mapping $\boldsymbol{q} \mapsto \hat{\boldsymbol{p}}(\boldsymbol{q})$ is given by $\hat{\mathbf{J}}(\boldsymbol{q}) = -(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}$. By the discussion at the beginning of this section, we have $\boldsymbol{p}(\boldsymbol{q}) = \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\hat{\boldsymbol{p}}(\boldsymbol{q})$, in which case, by the chain rule, we obtain $\mathbf{J}(\boldsymbol{q}) = \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\hat{\mathbf{J}}(\boldsymbol{q}) = -\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}$, which is the desired result. □

In the following lemma, we consider a matrix that will appear in the computation of the Hessian matrix of the expected revenue function. We define

$$\mathbf{D}(\boldsymbol{q}) = \frac{1}{1-\boldsymbol{e}^\top \boldsymbol{q}}\,(\boldsymbol{e}\,\boldsymbol{e}^\top \mathrm{diag}(\boldsymbol{\beta})^{-1} + \mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{e}\,\boldsymbol{e}^\top) + \frac{\boldsymbol{e}^\top \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{q}}{(1-\boldsymbol{e}^\top \boldsymbol{q})^2}\,\boldsymbol{e}\,\boldsymbol{e}^\top.$$

One can verify that $\mathbf{D}(\boldsymbol{q})$ is not necessarily positive semidefinite, but as we show next, if we add the matrix $\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\mathrm{diag}(\boldsymbol{q})^{-1}$ to $\mathbf{D}(\boldsymbol{q})$, then we obtain a positive definite matrix.

Lemma H.4 *For all $\boldsymbol{q} \in \mathbb{R}^n_+$ such that $q_i > 0$ for all i and $\sum_{i\in N} q_i < 1$, the matrix $\mathbf{C}(\boldsymbol{q}) = \mathbf{D}(\boldsymbol{q}) + \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\mathrm{diag}(\boldsymbol{q})^{-1}$ is positive definite.*

Proof: Define $g(\boldsymbol{q}) = \sum_{k\in M} \frac{1}{\beta^k} \sum_{i\in N^k} q_i\,(\log q_i - \log(1 - \sum_{j\in N} q_j))$. By direct differentiation, it can be verified that $\nabla^2 g(\boldsymbol{q}) = \mathbf{C}(\boldsymbol{q})$. So, it is enough to show that g is strictly convex. Let $H : \mathbb{R}^{2n}_+ \to \mathbb{R}$ be defined by: for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_+$, $H(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k\in M} \frac{1}{\beta^k} \sum_{i\in N^k} x_i\,(\log x_i - \log y_i)$. The relative entropy function $f(x,y) = x\,(\log x - \log y)$ is strictly convex in $(x,y) \in \mathbb{R}^2_+$; see Section 3.1.5 in Boyd and Vandenberghe (2004). Therefore, $H(\boldsymbol{x}, \boldsymbol{y})$ is strictly convex in $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{2n}_+$. Let the vector-valued mapping $\boldsymbol{h} : \mathbb{R}^n_+ \to \mathbb{R}^{2n}_+$ be defined by: for all $\boldsymbol{q} \in \mathbb{R}^n_+$, $\boldsymbol{h}(\boldsymbol{q}) = (h_1(\boldsymbol{q}), \ldots, h_{2n}(\boldsymbol{q}))$ with $h_i(\boldsymbol{q}) = q_i$ for all $i = 1, \ldots, n$, and $h_i(\boldsymbol{q}) = 1 - \sum_{j\in N} q_j$ for all $i = n+1, \ldots, 2n$. Then, the definitions of g, H and $\boldsymbol{h}$ imply that $g(\boldsymbol{q}) = H(\boldsymbol{h}(\boldsymbol{q}))$. Thus, g is the composition of a strictly convex function with an affine mapping, which only implies that g is convex. To show that g is strictly convex, we let $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ be the Jacobian matrix of the vector-valued mapping $\boldsymbol{h}$. Note that $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ is a $2n$-by-n matrix. For $i = 1, \ldots, n$, (i,i)-th entry of $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ is 1. For $i = n+1, \ldots, 2n$ and $j = 1, \ldots, n$, the (i,j)-th entry of $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ is -1. The other entries are zero. Therefore, the Jacobian matrix $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ includes the n-by-n identity matrix as a submatrix. Noting that $g(\boldsymbol{q}) = H(\boldsymbol{h}(\boldsymbol{q}))$ and $\boldsymbol{h}$ is linear, by the chain rule, we have $\nabla^2 g(\boldsymbol{q}) = \mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})^\top \nabla^2 H(\boldsymbol{h}(\boldsymbol{q}))\,\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$. Since $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})$ includes the n-by-n identity matrix as a submatrix, for any $\boldsymbol{x} \in \mathbb{R}^n$ with $\boldsymbol{x} \neq \mathbf{0}$, we have $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})\,\boldsymbol{x} \neq 0$. In this case, for any $\boldsymbol{x} \in \mathbb{R}^n$ with $\boldsymbol{x} \neq \mathbf{0}$, we obtain $\boldsymbol{x}^\top \nabla^2 g(\boldsymbol{q})\,\boldsymbol{x} = (\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})\,\boldsymbol{x})^\top \nabla^2 H(\boldsymbol{h}(\boldsymbol{q}))\,(\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})\,\boldsymbol{x}) > 0$, where the inequality follows from the fact that H is strictly convex and $\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{q})\,\boldsymbol{x} \neq \mathbf{0}$. □

Here is the proof of Theorem H.1.

Proof of Theorem H.1: First, we show the expression for $\nabla R(\boldsymbol{q})$. In the proof of Theorem 4.3, we use the Sherman-Morrison formula to obtain $(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \mathbf{B}(\boldsymbol{q}))^{-1} = (\mathrm{diag}(\boldsymbol{q}) + \mathbf{B}(\boldsymbol{q}))^{-1} + \boldsymbol{e}\,\boldsymbol{e}^\top/(1 - \boldsymbol{e}^\top \boldsymbol{q})$. Using the same argument, it follows that we have $(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} = (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} + \boldsymbol{e}\,\boldsymbol{e}^\top/(1 - \boldsymbol{e}^\top \boldsymbol{q})$. Furthermore, since $\hat{\mathbf{B}}(\boldsymbol{q})$ is positive semidefinite, the inverse of $\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q})$ exists, in which case, we have $(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = \mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{e}$. Noting that $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = \mathbf{0}$ by Lemma H.2, the last equality implies that $(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e}$. Since $R(\boldsymbol{q}) = \sum_{i\in N} p_i(\boldsymbol{q})\,q_i$, we get

$$\begin{aligned}
\nabla R(\boldsymbol{q}) &= \boldsymbol{p}(\boldsymbol{q}) + \mathbf{J}(\boldsymbol{q})^\top \boldsymbol{q} \;=\; \boldsymbol{p}(\boldsymbol{q}) - (\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{q}\\
&= \boldsymbol{p}(\boldsymbol{q}) - \left[(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} + \frac{1}{1-\boldsymbol{e}^\top\boldsymbol{q}}\,\boldsymbol{e}\,\boldsymbol{e}^\top\right]\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{q}\\
&= \boldsymbol{p}(\boldsymbol{q}) - \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} - \frac{\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{q}}{1-\boldsymbol{e}^\top\boldsymbol{q}}\,\boldsymbol{e},
\end{aligned}$$

where the second equality follows from Lemma H.3 and the last equality uses the fact that $(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{q} = (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e} = \mathrm{diag}(\boldsymbol{\beta})^{-1}\,\boldsymbol{e}$. Thus, we have the desired expression for $\nabla R(\boldsymbol{q})$.

Second, we show that the inverse of $\nabla^2 R(\boldsymbol{q})$ is negative definite, which implies that $\nabla^2 R(\boldsymbol{q})$ is negative definite as well. Differentiating the last expression above for the gradient of the expected revenue function, we have

$$\begin{aligned}
\nabla^2 R(\boldsymbol{q}) &= \mathbf{J}(\boldsymbol{q}) - \frac{1}{1-\boldsymbol{e}^\top\boldsymbol{q}}\,\boldsymbol{e}\,\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1} - \frac{\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{q}}{(1-\boldsymbol{e}^\top\boldsymbol{q})^2}\,\boldsymbol{e}\,\boldsymbol{e}^\top\\
&= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} - \frac{1}{1-\boldsymbol{e}^\top\boldsymbol{q}}\,\boldsymbol{e}\,\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1} - \frac{\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{q}}{(1-\boldsymbol{e}^\top\boldsymbol{q})^2}\,\boldsymbol{e}\,\boldsymbol{e}^\top\\
&= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\\
&\qquad\qquad - \frac{1}{1-\boldsymbol{e}^\top\boldsymbol{q}}\left[(\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{e}\,\boldsymbol{e}^\top + \boldsymbol{e}\,\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1}\right] - \frac{\boldsymbol{e}^\top\mathrm{diag}(\boldsymbol{\beta})^{-1}\boldsymbol{q}}{(1-\boldsymbol{e}^\top\boldsymbol{q})^2}\,\boldsymbol{e}\,\boldsymbol{e}^\top\\
&= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} - \mathbf{D}(\boldsymbol{q})\\
&= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\left[\mathbf{I} + (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q})\right],
\end{aligned}$$

where the third equality follows from the fact that $(\mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\,\boldsymbol{q}^\top + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} = (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1} + \boldsymbol{e}\,\boldsymbol{e}^\top/(1-\boldsymbol{e}^\top\boldsymbol{q})$. Noting the definition of $\mathbf{D}(\boldsymbol{q})$, multiplying both sides of the equality in this definition by $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})$ from left, Lemma H.2 implies that $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q}) = \mathbf{0}$. In this case, considering the term $\mathbf{I} + (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q})$ on the right side of the chain of equalities above, we have $\mathbf{I} + (\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q}) = \mathbf{I} + \mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q})$, in which case, the last chain of equalities above implies that $\nabla^2 R(\boldsymbol{q})$ is given by

$$\begin{aligned}
\nabla^2 R(\boldsymbol{q}) &= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}(\mathbf{I} + \mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})\,\mathbf{D}(\boldsymbol{q}))\\
&= -\,\mathrm{diag}(\boldsymbol{\beta})^{-1}\,(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})\left[\mathrm{diag}(\boldsymbol{\beta})^{-1}\mathrm{diag}(\boldsymbol{q})^{-1} + \mathbf{D}(\boldsymbol{q})\right].
\end{aligned}$$

For notational brevity, we define the diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})$, in which case, we write the Hessian of $R(\boldsymbol{q})$ as $\nabla^2 R(\boldsymbol{q}) = -\mathrm{diag}(\boldsymbol{\beta})^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))^{-1}\,\boldsymbol{\Lambda}\,(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))$. Noting Lemma H.4, the inverse of $\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q})$ exists. By the Sherman-Morrison formula, it is given by

$$(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1} = \boldsymbol{\Lambda} - (\mathbf{I} + \boldsymbol{\Lambda}\,\mathbf{D}(\boldsymbol{q}))^{-1}\boldsymbol{\Lambda}\,\mathbf{D}(\boldsymbol{q})\,\boldsymbol{\Lambda}.$$

By Lemma H.2, noting the definition of $\mathbf{D}(\boldsymbol{q})$, we have $\hat{\mathbf{B}}(\boldsymbol{q})\,\mathbf{D}(\boldsymbol{q}) = \mathbf{0}$. Since $\hat{\mathbf{B}}(\boldsymbol{q})$ and $\mathbf{D}(\boldsymbol{q})$ are both symmetric, taking the transpose, it follows that $\mathbf{D}(\boldsymbol{q})\,\hat{\mathbf{B}}(\boldsymbol{q}) = \mathbf{0}$. Using the last expression for the Hessian matrix of $R(\boldsymbol{q})$ and taking the inverse, we obtain

$$\begin{aligned}
\nabla^2 R(\boldsymbol{q})^{-1} &= -\,(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1}\boldsymbol{\Lambda}^{-1}(\mathrm{diag}(\boldsymbol{q}) + \hat{\mathbf{B}}(\boldsymbol{q}))\,\mathrm{diag}(\boldsymbol{\beta}) \\
&= -\,(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1}\boldsymbol{\Lambda}^{-1}\mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta}) - (\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1}\boldsymbol{\Lambda}^{-1}\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta}) \\
&= -\,(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1} - \Big[\boldsymbol{\Lambda} - (\mathbf{I} + \boldsymbol{\Lambda}\,\mathbf{D}(\boldsymbol{q}))^{-1}\boldsymbol{\Lambda}\,\mathbf{D}(\boldsymbol{q})\,\boldsymbol{\Lambda}\Big]\boldsymbol{\Lambda}^{-1}\,\hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta}) \\
&= -\,(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1} - \hat{\mathbf{B}}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta}),
\end{aligned}$$

where third equality uses the fact that $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{q})\,\mathrm{diag}(\boldsymbol{\beta})$, whereas the last equality holds since $\mathbf{D}(\boldsymbol{q})\,\hat{\mathbf{B}}(\boldsymbol{q}) = \mathbf{0}$. By Lemma H.4, $(\boldsymbol{\Lambda}^{-1} + \mathbf{D}(\boldsymbol{q}))^{-1}$ is positive definite. Also, recall that $\hat{\mathbf{B}}(\boldsymbol{q})$ is positive semidefinite. Therefore, the inverse of $\nabla^2 R(\boldsymbol{q})$ is negative definite. □