Tractable Multi-product Pricing under Discrete Choice Models

by

Philipp Wilhelm Keller

B.Sc., McGill University (2005)
M.Sc., McGill University (2008)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author ....
Sloan School of Management
May 17, 2013

Certified by ....
Retsef Levi
J. Spencer Standish Professor of Management
Thesis Supervisor

Certified by ....
Georgia Perakis
William F. Pounds Professor of Management
Thesis Supervisor

Accepted by ....
Dimitris Bertsimas
Co-Director, Operations Research Center

Tractable Multi-product Pricing under Discrete Choice Models

by

Philipp Wilhelm Keller

Submitted to the Sloan School of Management
on May 17, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

We consider a retailer offering an assortment of differentiated substitutable products to price-sensitive customers. Prices are chosen to maximize profit, subject to inventory/capacity constraints, as well as more general constraints. The profit is not even a quasi-concave function of the prices under the basic *multinomial logit* (MNL) demand model. Linear constraints can induce a non-convex feasible region. Nevertheless, we show how to efficiently solve the pricing problem under three important, more general families of demand models.

Generalized attraction (GA) models broaden the range of nonlinear responses to changes in price. We propose a reformulation of the pricing problem over demands (instead of prices) which is convex. We show that the constrained problem under MNL models can be solved in a polynomial number of Newton iterations. In experiments, our reformulation is solved in seconds rather than days by commercial software.

For *nested-logit* (NL) demand models, we show that the profit is concave in the demands (market shares) when all the price-sensitivity parameters are sufficiently close. The closed-form expressions for the Hessian of the profit that we derive can be used with general-purpose nonlinear solvers. For the special (unconstrained) case already considered in the literature, we devise an algorithm that requires no assumptions on the problem parameters.

The class of *generalized extreme value* (GEV) models includes the NL as well as the *cross-nested logit* (CNL) model. There is generally no closed form expression for the profit in terms of the demands. We nevertheless how the gradient and Hessian can be computed for use with general-purpose solvers. We show that the objective of a transformed problem is nearly concave when all the price sensitivities are close. For the unconstrained case, we develop a simple and surprisingly efficient first-order method. Our experiments suggest that it always finds a global optimum, for any model parameters.

We apply the method to *mixed logit* (MMNL) models, by showing that they can be approximated with CNL models. With an appropriate sequence of parameter scalings, we conjecture that the solution found is also globally optimal.

Thesis Supervisor: Retsef Levi
Title: J. Spencer Standish Professor of Management

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management

Acknowledgments

I thank my advisors, Georgia and Retsef, for their support over the years and for their respective contributions to this work. Georgia first introduced me to the multi-product pricing problem and, although it took a while to find the right angle of attack, I'm amazed at how rich and interesting the topic turned out to be. Retsef has been a great help in motivating me to publish, and the second chapter would not be nearly as well written without the many suggestions he made while we prepared a journal submission. Of course, I'd also like to thank both my advisors for eventually reading the present thesis so carefully and providing helpful feedback.

I'd also like to thank Vivek Farias for serving on my committee, for his insightful comments, and for his positivity and encouragement. TA'ing for Cindy Barnhart, Amedeo Odoni and Carolina Osorio was a highlight of my time at MIT, and I thank them for it. I'd like to thank the ORC-affiliated faculty as a whole for the effort they've put into the program and for the many excellent classes I've been able to sit in. I thank Laura, Andrew and Paulette for the innumerable things they've done for me, and for always being so helpful, responsive and friendly.

I thank my colleagues in the ORC for their good company, and for the opportunity to work with many of them. It is a pleasure to know Shubham, Matthew, Allison, Claudio, Ton, Shreevatsa, Matthieu, Jonathan, Maxime, Cristian, Jason, Dan, Ruben, Thibault, Rajiv, Joline, Ross, Chaithanya, Mike, Ta, Adam, Alex, Vishal, Yehua, Andre, and just about everyone else who came through the program while I was there. I especially thank Gareth and Mallory for their friendship, and for all the good times.

I thank my family for their support and encouragement, and I thank my sisters for making me proud.

More than anyone, I thank Jackie for her love and support through it all.

Contents

Chapter 1

Introduction

The pricing problem is fundamental in revenue management, but, until recently, the effect of customer substitution on pricing has received little attention. In fact, most of the revenue management literature assumes that demands for different products are entirely independent [12, 26, 75]. There has been significant recent work on discrete-choice models in quantity-based revenue management, where prices are fixed and instead the product line offered to customers must be chosen. However, except for the basic *multinomial logit* (MNL) model and a limited class of *nested logit* (NL) models, the available algorithms are approximations or heuristics even in this setting [20, 33, 35, 52, 59, 60, 65, 76, 81].

When we began this work, little was known about the structure of the pricing problem in the face of customer choice. Even under the relatively simple MNL model, the profit is not a quasi-concave function of the prices. Nevertheless, heuristics had been devised to find good locally optimal solutions under the more complex *mixed logit* (MMNL) models [40], and under MNL models with certain types of constraints on allowable demands induced by the prices [31]. It was known that the globally optimal solution is unique for unconstrained MNL models. Nevertheless, concave reformulations of even the unconstrained MNL pricing problem appeared in the literature only concurrently with our own work [23, 36, 51, 53, 69, 73].

1.1 Discrete Choice Models

In this thesis, we propose efficient algorithms for optimally pricing a set of substitutable products when customer demand is represented by discrete choice models. Specifically, we first consider pricing under the ubiquitous MNL model with linear constraints on the resulting demands and certain types of constraints on allowable prices. We then study the pricing problem under three important generalizations of the MNL model:

(i) *Generalized attraction* (GA) demand models maintain the general form of MNL models, but allow for different nonlinear responses to changes in price. They include the MNL, *multiplicative competitive interaction* (MCI) or *Cobb-Douglas* models [78] and linear attraction demand models [34], as well as special-purpose models based on prior knowlege of the response to price changes and the semi-parametric models introduced by Hruschka [41].

(ii) *Generalized extreme value* (GEV) models allow for a richer dependence of the demand for a given product on the prices of the other altervatives, but restrict the class of functions than can model the responses to changes in price. This family includes the MNL, the common *nested logit* (NL) and *cross-nested logit* (CNL) models, as well as more complex models [77, 7].

(iii) *Mixed multinomial logit* (MMNL, or simply *mixed logit*), models define the demand as a weighted sum of MNL models. Each MNL model in the mixture may represent a customer segment choosing among the same set of products, or the models may result from sampling continuous mixtures of MNL models (sometimes called *logit kernel* models [58]).

All of the models in (i) - (iii) above define a probability distribution over n choices available to the consumer, including the choice of buying nothing[1]. The GA and GEV

[1]This "outside alternative" will be represented in the main body of this thesis by setting one of the prices to $x_i = 0$ and setting the corresponding function defined below to $f_i(\mathbf{x}) = 1$ for any values of the other prices.

models share the same basic form: the fraction of consumers, in expectation, opting to buy product i is given by

$$p_i = \frac{f_i(\mathbf{x})}{\sum_{j=1}^{n} f_j(\mathbf{x})},$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector containing the prices for each of the n products. The normalizing factor in the denominator ensures that the demands sum to one. The models differ in how the functions $f_1, \ldots, f_n$ are defined. For example, under the MNL model, they take the form

$$f_i(\mathbf{x}) = y_i \triangleq e^{d_i - b_i x_i}$$

for some parameters d_i and b_i. The *attraction* y_i of product i depends only on the price x_i of that product. Other scalar functions define different GA models with different attractions y_i. The GEV family of models is much richer because it allows each function f_i to depend on all of the prices, rather than only x_i. MMNL models, on the other hand, are defined as a mixture of MNL models with choice probabilities

$$p_i = \sum_{k=1}^{K} \gamma_k \left(\frac{f_i^k(\mathbf{x})}{\sum_j f_j^k(\mathbf{x})} \right),$$

where each set $\{f_1^k, \ldots, f_n^k\}$ defines an MNL model and $\gamma_k > 0$ is its weight in the mixture. In this thesis, we assume that the functions f_i^k are of the form

$$f_i^k(\mathbf{x}) = e^{d_i^k - b^k x_i},$$

where the parameters d_i^k may differ across products for each segment i, but where the price sensitivity parameter b^k is constant for each segment k. We refer the reader to the relevant chapters for more detailed definitions of the various models.

1.2 The Pricing Problem

The aim of the pricing problem, is to maximize the total expected profit subject to constraints limiting the set of demands that can be served:

$$\max_{\mathbf{p}\in\mathcal{P}} \left\{ \Pi(\mathbf{x}) = \sum_{i=1}^{n} p_i x_i \right\}.$$

Here $\mathcal{P}$ represents the feasible set of demand vectors $\mathbf{p} = [p_1, \ldots, p_n]^\top$, and $\mathbf{x} = [x_1, \ldots, x_n]^\top$ is the corresponding vector of prices defined above. Under discrete choice models, we may assume without loss of generality that the vector of demands $\mathbf{p}$ is a probability distribution over products. Scaling the demands by the total population size then yields the true demand.[2]

We begin by assuming that $\mathcal{P}$ is defined by linear constraints arising from, for example, inventory availability, capacity constraints or minimum market share targets. In fact, for the MNL model, we show that even orderings between prices $\mathbf{x}$ can be enforced as linear constraints on $\mathbf{p}$. It will become obvious that in some cases, general convex constraints in $\mathbf{p}$ can be incorporated as well. Such constraints arise naturally in the context of pricing problems, but have not been considered until recently in combination with customer substitution (see the literature review and model definition in Chapter 2). Unfortunately, when we consider the entire family of GEV models, we must drop the constraints and assume that $\mathcal{P}$ is equal to the entire probability simplex. This is because the techniques that we apply to solve these more complex models no longer preserve the linearity of the constraints.

Thus far, our model does not explicitly capture the fact that different marginal production costs c_i may apply to each product, or that profit margins $0 < a_i < 1$ may only represent a fraction of the price charged to the consumer, and may also change

[2]We include the "outside alternative" with a price of $x_i = 0$ in the set of products here, such that the entries of $\mathbf{p}$ capture every possible alternative and therefore sum to one.

across products. In practice, the actual profit usually has the form

$$\Pi'(\mathbf{x}) = \sum_{i=1}^{n} a_i p_i (x_i - c_i)$$

As we will show in Chapters 2 and 3, variable production costs c_i and profit margins a_i can naturally be incorporated into the parameters of the demand models that we consider, and we can usually consider the simpler function $\Pi(\mathbf{x})$ without loss of generality.

The difficulty of computing the global maximum of the pricing problem obviously depends on the specific model relating the vector of demands $\mathbf{p}$ and the vector of prices $\mathbf{x}$. After outlining the organization of the thesis, we give a brief overview of how we approach the problem for each of the demand models, and of they type of results obtained.

1.3 Contributions

Table 1.1 summarizes our results and the algorithms we propose. For GA models, we show that the pricing problem can be reformulated in terms of the demands $\mathbf{p}$ by inverting each scalar attraction function $f_i(x_i)$ individually, taking into account arbitrary linear constraints on the demands and certain constraints on the prices. We provide a sufficient condition on the attraction functions to ensure that the reformulation is a convex optimization problem. The condition is satisfied by all GA models commonly found in the literature. Moreover, we show that, for the MNL model, the reformulation can be solved in polynomial time using interior-point methods, and we demonstrate empirically that our reformulations are orders of magnitude faster to solve than naive formulations in terms of the prices. If the inverses $f_i^{-1}(y_i)$ can be computed efficiently, then the gradient and Hessian can also be obtained efficiently and a general-purpose (GP) convex optimization solver can be used. If this is not the case, we provide an alternative algorithm for solving the dual of the refomulation. For example, the approximated MMNL models discussed below, and the semi-parametric

Model	**Constraints**	**Result**	**Algorithm**	**Sec.**
MNL	Linear in $\mathbf{p}$, offsets and bounds in $\mathbf{x}$.	- Concave reformulation in $\mathbf{p}$ for MNL, MCI, Cobb-Douglas and linear attraction models. - Sufficient concavity condition for other GA models.	Polynomial-time solution via the self-concordant barrier method.	2.3
GA, w/ $f_i^{-1}(y_i)$	Linear in $\mathbf{p}$, bounds in $\mathbf{x}$.		Closed-form Hessian for GP nonlinear solver.	
GA, no $f_i^{-1}(y_i)$			Column-generation algorithm for dual of reformulation.	2.5, A.3
NL	Linear in $\mathbf{p}$.	Concave reformulation in $\mathbf{p}$ if price sensitivity parameter ratio < 2.	Closed-form Hessian for GP nonlinear solver.	3.7
	None	Bi-concave reformulation if no-purchase alternative alone in a nest.	Iteration with closed-form solutions at each step, converges to a stationary point.	3.8
CNL & other GEV models.		"Almost" concave reformulation if price sensitivity parameter ratio < 2.	Numerical Hessian for GP solver, via (potentially sparse) matrix inversion.	4.6
		FOCs often (generalized) diagonally dominant function.	Nonlinear Jacobi or Gauss-Seidel method for FOCs.	5.5
		Alternate FOCs defined by a diagonally dominant matrix.	Sub-stochastic iteration matrix to solve nonlinear FOCs.	5.6
MMNL	Linear in $\mathbf{p}$.	Approximate $\mathbf{p}$ via GA without $f_i^{-1}(y_i)$.	(see GA above)	2.4
	None	Approximate $\mathbf{p}$ *and* its Jacobian matrix via CNL.	Adapt GEV algorithm to solve MMNL directly.	5.7

Table 1.1: A summary of our novel results for various discrete choice models.

models of Hruschka [41] do not allow for easily computed inverses of the attraction functions.

Like for the MNL model, for the NL model with linear constraints on the demands, there exists a closed-form expression of the prices $\mathbf{x}$ in terms of the demands $\mathbf{p}$. By an appropriate series of chain rule applications, we are able to derive closed-form expressions for the gradient and Hessian of the profit $\Pi(\mathbf{p})$ in terms of the demands $\mathbf{p}$. **We then show that the NL profit function is concave with a unique local maximum as long as the price sensitivity parameters for the different products are all within a factor of 2 of each other.** The condition on the price sensitivity parameters appears mild, but it is often violated when fitting models to data.[3] To the best of our knownledge, this represents a significant improvement over the known results in the literature, even for the special case where they apply. For the case where the outside no-purchase alternative is in its own nest and there are no constraints, as is assumed in existing work [51, 36, 20], we develop a different *bi-concave* reformulation in terms of the *conditional* choice probabilities within each nest. Furthermore, we show that it can be solved iteratively by computing the solution of a MNL pricing problem in closed-form at each step. **Under the assumptions made in the existing literature, This algorithm converges to a stationary point for *any* price sensitivity parameters.** We conjecture that it in fact converges to the global maximum.

For GEV models, we first show that the demand model is in fact always invertible, even though no closed-form inverse demand function exists in general. Unfortunately, we cannot obtain a closed-form expression for the Hessian of the profit $\Pi(\mathbf{p})$, and our later results will lead us to believe that the profit is often *not* a concave function of $\mathbf{p}$. **Instead, we consider a reformulation of the GEV pricing problem in terms of the *unnormalized demands*. We are able to show that this reformulation is *almost* concave under the same condition on the price sensitivity parameters outlined above** (that the price-sensitivity parameters for the products

[3]Because the NL model cannot explicitly capture different price sensitivities for different customer segments (like the MMNL model), large variations in the per-product price sensitivity parameters may yield a better fit even if each individual has identical price-sensitivity across all products.

are all within a factor of 2 of each other). Specifically, the Hessian consists of a negative-definite matrix and a (usually small) error term. We provide expressions for the Hessian matrices of the reformulation and the original profit function $\Pi(\mathbf{p})$. They can be used to evaluate either quantity numerically when using a general-purpose nonlinear solver.

We also obtain a compact expression for the first-order optimality conditions (FOCs) of the GEV pricing problem in terms of the prices $\mathbf{x}$. We show that they are often defined by a *generalized diagonally dominant function* of $\mathbf{x}$. **This implies that the FOCs admit a unique solution, which can be computed using the nonlinear Jacobi or Gauss-Seidel method.** Unfortunately, as when using the gradient and Hessian mentioned above with a general-purpose solver, evaluating the quantities involved at each iteration is costly. Even computing the prices $\mathbf{x}$ corresponding to a given vector of demands $\mathbf{p}$ requires the solution of a nonlinear system. Evaluating the actual gradient and Hessian requires the inversion of a matrix that depends on $\mathbf{x}$ and that grows in the number of products. This computation is usually expensive (although it may be accelerated when the matrix in question is sparse, as for CNL models with limited cross-nesting).

An alternative expression that we obtain for the FOCs leads to the main algorithm of Chapter 5. **Specifically, we show that applying *linear* Jacobi iterations to the (re-stated) nonlinear FOCs always results in a sub-stochastic iteration matrix.** That is, the current solution is multiplied by a sub-stochastic matrix at each step. **The sequence of solutions obtained by our algorithm converges quickly and consistently in computational experiments, regardless of the starting point.** Computationally, this seems to be our most effective method for unconstrained pricing problems. We show that it converges to a local maximum, and conjecture that it in fact converges to a global maximum for all the problems that we have considered. The practically-minded reader may wish to skim the early chapters, and proceed directly to Chapter 5 (skipping Section 5.5).

Finally, we consider two approaches for solving the pricing problem under MMNL models. First, we show in Chapter 2 that the demands $\mathbf{p}$ under such models can be approximated locally by a certain model of the GA type. This allows constrained problems to be solved approximately. Second, we show in Chapter 5 that **the MMNL demands p *as well as* their partial derivatives with respect to the prices can be approximated locally by a CNL model.** Combining this observation with a known path-following algorithm for MMNL models [40] allows us to directly apply the linear Jacobi iterations for GEV models to MMNL models. **In computational experiments, convergence for MMNL models is as fast as for GEV models, and the solutions are as good as those found by the path-following algorithm in the literature.**

1.4 Thesis Organization

Chapters 2, 3 and 4 of this thesis consider the pricing problem under GA, NL and GEV models, respectively. For each class, we aim to provide sufficient conditions for the existence and uniqueness of the solution, as well as practical algorithms for computing it. In Chapter 5, we finish with a simple and efficient algorithm for solving the unconstrained pricing problem under any GEV model (including the MNL, NL and CNL models) as well as under MMNL models. Since we draw on different methods to handle each class of demand models, and because the existing work on pricing usually only applies to one of the demand models, we defer the literature review to Sections 2.1.1, 3.3, 4.3, 5.2 and 5.6.1 in the relevant chapters.

Chapter 2

Pricing under Attraction Demand Models

2.1 Introduction

In this chapter, we study a general modeling and optimization framework that captures fundamental multi-product pricing problems. We consider the basic setting of a retailer offering an assortment of differentiated substitutable products to a population of customers who are price sensitive. The retailer selects prices to maximize the total profit subject to constraints on sales arising from inventory levels, capacity availability, market share goals, sales targets, price bounds, joint constraints on allowable prices, and other considerations.

The profit functions and constraints we consider are captured in the following general nonlinear optimization problem. The decision variables x_i are the prices charged for each of the products indexed by $i = 1, \dots, n$.

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} a_i x_i d_i(\mathbf{x}) & & \\ \text{s.t.} \quad & \sum_{i=1}^{n} A_{ki} d_i(\mathbf{x}) \leq u_k & & k = 1, 2, \dots m \\ & \underline{x}_i \leq x_i \leq \overline{x}_i & & i = 1, 2, \dots n \end{aligned} \tag{P}$$

The input data to this model are the profit margins $a_i > 0$ for each product $i = 1, 2, \dots, n$, the matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and the vector $\mathbf{u} \in \mathbb{R}^m$ defining m constraints on the demand, and upper and lower bounds $\underline{\mathbf{x}}, \overline{\mathbf{x}} \in \mathbb{R}^n$ on allowable prices.

The price-demand functions $d_i(\mathbf{x})$, for $i = 1, \dots, n$, are central to the range of models that can be captured, as well as to the computational tractability of the resulting optimization problem. We represent customers' purchasing decisions through *attraction demand models*, which generalize the well-known *multinomial logit* (MNL) and *multiplicative competitive interaction* (MCI) demand models [57, 78]. This approach is common in the recent revenue management literature, as well as in marketing and economics. The function $d_i : \mathbb{R}^n \to (0,1)$ maps the prices of *all* the products to the observed customer demand for product i. We assume that attraction demand models have the following form:

$$d_0(\mathbf{x}) = \frac{1}{1 + \sum_j f_j(x_j)} \quad \text{and} \quad d_i(\mathbf{x}) = \frac{f_i(x_i)}{1 + \sum_j f_j(x_j)}, \quad i = 1, \dots, n. \tag{2.1}$$

The quantity $d_0(\mathbf{x})$ denotes the fraction of consumers opting not to purchase any product, and $d_i(\mathbf{x})$ is the demand for the i^{th} product when $i > 0$. The functions $f_i(x_i)$ are called *attraction functions* and are assumed to satisfy the following assumption:

Assumption 2.1. The attraction function $f_i : \mathbb{R} \to \mathbb{R}_{++}$ for each product $i = 1, 2, \dots, n$, satisfies:

(i) $f_i(\cdot)$ is strictly decreasing and twice differentiable on $\mathbb{R}$, and

(ii) $\lim_{x \to -\infty} f_i(x) = \infty$, and $\lim_{x \to \infty} x f_i(x) = 0 \quad$ (i.e., $f_i(x) \in o(\frac{1}{x})$).

Unlike much of the existing work in revenue management, these models allow the demands for each of the products to be interdependent functions of all the prices. However, under attraction demand models, the profit as a function of the prices set by the retailer is in general *not quasi-concave* (see [40] and Appendix A.2). Moreover, many realistic constraints involving the demand model, such as production/inventory capacity bounds, give rise to a *non-convex* region of feasible prices. Maximizing the profit thus presents a challenging optimization problem. Our experiments show that commercial software may take over a day to solve a pricing-based formulation, even for relatively small instances. Furthermore, we have no a priori guarantee that such an approach will converge to a globally optimal solution.

The contributions of this paper are multifold. First, we provide equivalent reformulations of the pricing problem that are provably tractable and can be solved efficiently by commercial software. Defining the *inverse* attraction functions as $g_i(y) = f_i^{-1}(y)$, $y > 0$, our reformulations take the following general form:

$$\begin{aligned} \max\ \Pi(\boldsymbol{\theta}) &= \sum_i a_i \theta_i g_i\left(\frac{\theta_i}{\theta_0}\right) & & \\ \text{s.t.} \sum_{i=1}^{n} A_{ki}\theta_i &\le u_k & k &= 1 \ldots m' \\ \sum_{i=0} \theta_i &= 1 & & \\ \theta_i &> 0 & i &= 0 \ldots n \end{aligned} \tag{COP}$$

The decision variables $\theta_1, \ldots, \theta_n$ represent the fraction of customers opting to purchase each product and the set of constraints defined by $(\mathbf{A}, \mathbf{u})$ has been extended with $2n$ additional linear constraints.

We establish a general and easily verifiable condition on the attraction demand model under which the reformulation gives rise to convex optimization problems. Specifically, our approach yields maximization problems of the form (COP) with *concave* objective functions $\Pi(\boldsymbol{\theta})$ and *linear* constraints. This is despite the nonlinear and non-separable nature of the demand model. Moreover, we prove that the logarithmic barrier method solves the pricing problem under MNL demand in a polynomial number of iterations of Newton's method. We confirm through extensive computational experiments that our formulations can be solved in seconds instead of days, compared to the naive formulations, and that they scale well to instances with thousands of products and constraints.

We then show how to apply our approach to obtain tractable approximations to the challenging pricing problem arising under *weak market segmentation*, where the pricing decisions affect the demand in multiple overlapping customer segments. Such models are exemplified by *mixed multinomial logit* models [13, 17]. Representing even a small number of distinct segments, such as, for example, business and leisure travelers on a flight, may significantly improve accuracy over a single-segment model. Moreover, *any* random utility maximization model can be approximated arbitrarily closely by a mixed MNL model (see Mcfadden and Train [58]).

We approximate such mixed demand functions with valid attraction demand models that yield convex optimization problems of the form (COP), and we bound the error with respect to the true multi-segment model. The attraction demand model in question is relatively complex, and the resulting objective function in our reformulation does not have a closed form. Nevertheless, we show how the objective function, its gradient and its Hessian can be computed efficiently, allowing standard optimization algorithms to be applied.

The remainder of this section reviews related work. Section 2.2 describes the price optimization problem, Section 2.3 presents our reformulation and Section 2.4 presents our approach for problems with multiple overlapping customer segments. Section 2.5 states the dual of our reformulation. Section 2.6 compares the different approaches we consider computationally. Further details about specific attraction demand models may be found in Appendices A.1 and A.2. The derivation of the dual of our reformulation and an algorithm for solving it are provided in Appendix A.3.

2.1.1 Literature Review

Pricing as a tool in revenue management usually arises in the context of perishable and nonrenewable inventory such as seats on a flight, hotel rooms, rental cars, internet service and electrical power supplies (see, e.g., the survey by Bitran and Caldentey [12]). Dynamic pricing policies are also adopted in retail and other industries where short-term supply is more flexible, and the interplay between inventory management and pricing may thus take on even greater importance. (See Elmaghraby and Keskinocak [26] for a survey of the literature on pricing with inventory considerations.)

The modeling framework studied in this paper generalizes a variety of more specialized pricing problems considered in the operations management literature. Much of the work focuses on *dynamic* pricing, under *stochastic* customer demand. The stochastic dynamic program arising under such models is generally intractable. Solution methods common in practice often rely on periodically re-solving single-period deterministic pricing problems, and this approach is known to be asymptotically optimal in some cases [32, 74]. Thus, the single-stage deterministic problems we consider play a central role. Our modeling framework relaxes two common but restrictive assumptions imposed in most existing multi-product pricing work. First, customers' substitution behavior can be modeled since demands are functions of *all* the prices. Secondly, a broad class of practical constraints can be enforced, well beyond just capacity bounds or inventory constraints allowed in existing models.

Specialized algorithms have been developed for solving certain single-period pricing problems. Hanson and Martin [40] devise a path following heuristic for the *unconstrained* pricing problem under mixtures of MNL models, and Gallego and Stefanescu [31] propose a column generation algorithm for a class of constrained problems with MNL demand (discussed in Section 2.2.2). Neither of these heuristics is guaranteed to find a globally optimal solution in finite time, nor can they be implemented directly with commercial software. They may also be computationally expensive in practice.

A number of recent papers have proposed multi-product pricing formulations using the *inverse* demand model to yield a concave objective function in terms of sales. Examples include Aydin and Porteus [5], Dong et al. [23] and Song and Xue [73]. In contrast to our framework, these papers focus on capturing inventory holding and replenishment costs in the objective function rather than considering explicit constraints on prices and sales. They also limit their attention to the MNL or other specific demand models. [69] proposes a formulation of the *product line design* (PLD) problem with continuous prices. The PLD problem is closely related to the pricing problems we consider, but it involves *discrete* decision variables and specific types of capacity constraints. The variable transformation which arises when inverting the demand in our pricing problem is similar to the generalized Charnes-Cooper transformation described by [66] for concave-convex fractional programs. However, the pricing problems studied in this paper are in general *not* concave-convex fractional programs.

Recent work on quantity-based revenue management also relaxes the assumption that the demands for different products are independent. In contrast to multi-product price-based revenue management, *network revenue management* (NRM) consists of choosing which subset of products to offer customers at each period from a menu with *fixed* prices, under inventory or capacity constraints. [33] and [52] consider customer substitution in this setting. [60] additionally consider overlapping customer segments, like in the weak market segmentation setting for which we provide an approximation. [75] provide an in-depth treatment of both price- and quantity-based revenue management and their relationship.

2.2 Modeling Framework

We first discuss the general pricing formulation (P), which has a non-concave (nor quasi-concave) objective and non-convex constraints in general. Then in Section 2.3 we describe the alternative formulation (COP), which is tractable.

A key question in pricing optimization is how to model the relationship between prices and the demands for each product. Assumption 2.1 is natural and captures, for example, the well known multinomial logit (MNL) model, with the attraction functions

$$f_i(x_i) = v_i e^{-x_i}, \tag{2.2}$$

and constant parameters $v_i > 0, i = 1, \ldots, n$. The technical requirement (ii) is mild, and in fact any attraction function can be modified to satisfy it without changing the objective values over the feasible region of (P). The demand $d_i(\mathbf{x})$ is equal to the attraction of product i normalized by the total attraction of all the customers' alternatives, including the option of not purchasing anything from the retailer in question. Without loss of generality, the attraction of the latter *outside alternative* is represented by the term 1 in the denominator. (Notice that the model is invariant to scaling of the attraction functions.) Further discussion of attraction demand models and explanations of how they can be adapted to satisfy Assumption 2.1 above may be found in Appendices A.1 and A.2, respectively. Due to the demand model, the objective of problem (P) is nonlinear and in general not quasi-concave. The problem appears challenging even without any constraints. Nevertheless, we will show that the broad class of *constrained* problems we consider is in fact tractable (specifically, unimodal).

Indeed, a wide variety of constraints can be represented by the formulation (P). Capacity and inventory bounds are very common and arise in revenue management problems. For example, in a flight reservation system each product may represent an itinerary with given travel restrictions, while seat availability on shared flight legs is represented by coupling resource constraints. To capture these constraints in our model, A_{ki} would be set to 1 if itinerary i uses leg k and zero otherwise. The upper bound u_k would be set to the number of seats available for leg k, divided by the total size of the population. In a retail setting, the identity matrix $\mathbf{A} = \mathbf{I}$ may be used and u_k may be set to the inventory available for item k. Minimum sales targets for a group of products may be represented with additional inequality constraints, with negative coefficients since they place a lower bound on the demands. In product-line design such operational constraints may be less important, but production capacities and minimum market-share targets take a similar form. Unfortunately, even though the constraints of (P) are linear in the demands for each of the products, they yield a *non-linear* and *non-convex* feasible region of prices in general. See Appendix A.2 for examples.

2.2.1 Marginal Costs, Joint Price Constraints and Other Extensions

We have defined the objective function of (P) in terms of relative profit margins, such as when a capacity reseller earns commissions. A per-unit production cost may be incorporated by using the objective

$$\sum_{i=1}^{n}(x_i - c_i)d_i(\mathbf{x}),$$

redefining the "prices" in our general formulation as the profit margins $\hat{x}_i \triangleq (x_i - c_i), i = 1, \ldots, n$, and shifting the attraction functions accordingly. This motivates why we allow negative prices in general, since the profit margin $\hat{x}_i$ may be negative even if the price x_i paid by the consumer is positive.

Moreover, any joint constraint on prices of the form

$$x_i \geq x_j + \delta_{ij}, \qquad \delta_{ij} \in \mathbb{R} \tag{2.3}$$

can be re-expressed as

$$\begin{aligned} f_i(x_i) \leq f_i(x_j + \delta_{ij}) \quad &\Leftrightarrow \quad f_i(x_i)d_0(\mathbf{x}) \leq f_i(x_j + \delta_{ij})d_0(\mathbf{x}) \\ &\Leftrightarrow \quad d_i(\mathbf{x}) \leq \frac{f_i(x_j + \delta_{ij})}{f_j(x_j)} d_j(\mathbf{x}), \end{aligned} \tag{2.4}$$

where we have used the monotonicity of $f_i(\cdot)$, the positivity of $d_0(\mathbf{x})$, and the fact that

$$f_i(x_i) = \frac{d_i(\mathbf{x})}{d_0(\mathbf{x})}. \tag{2.5}$$

Under mild assumptions on an MNL demand model, the ratio in the last inequality of (2.4) is a constant. The resulting linear constraint is captured by the formulation (P). A similar transformation is possible for linear attraction demand models. The details can be found in Appendix A.2.

We briefly mention two other straightforward extensions to our model. First, multiple customer segments may be represented by distinct, independent demand models. If it is possible to present different prices to each segment, or if disjoint subsets of products are offered to each segment, the pricing problem corresponds to multiple instances of (P) coupled only through linear constraints. Our approach in the next section carries through directly in such cases. (This is in contrast to the model discussed in Section 2.4 where all customers are offered all products at the same prices.) Secondly, since the equations describing inventory dynamics are *linear*, deterministic multi-stage pricing problems can be expressed in a similar form. Each period can be represented by a copy of (P). The copies are coupled only through the linear inventory dynamics constraints.

2.2.2 Special Cases with Convex Constraints

Proposition 2.2 below characterizes a class of constraints for which the feasible region of (P) *is* convex. Although the remainder of this paper considers the general formulation (P), the sub-class of problems considered in this sub-section has been implicitly studied in previous work. It arises naturally in specific revenue management problems, and encompasses the problems involving customer choice considered by Gallego and Stefanescu [31].

To our knowledge, the condition given here has never been made explicit. We provide an explanation of why versions of the pricing problem satisfying it have been found relatively easy to solve in Section 2.3.1.

Proposition 2.2. *If the attraction functions $f_1, f_2, \ldots, f_n$ are convex, and the constraints satisfy*

$$A_{ki} \geq u_k \geq 0, \qquad \textit{for each } k = 1, 2, \ldots, m, \textit{ and } i = 1, 2, \ldots, n, \tag{2.6}$$

then the feasible region of (P) *is a convex set.*

Proof. Clearly the bounds on the prices define a convex set. The k^{th} inequality constraint, $1 \leq k \leq m$, is convex since it may be expressed as a positive linear combination of convex functions:

$$\begin{aligned} \sum_{i=1}^{n} \frac{A_{ki} f_i(x_i)}{1 + \sum_{j=1}^{n} f_j(x_j)} \leq u_k \Leftrightarrow \sum_{i=1}^{n} A_{ki} f_i(x_i) &\leq \left(1 + \sum_{j=1}^{n} f_j(x_j)\right) u_k \\ \Leftrightarrow \sum_{i=1}^{n} (A_{ki} - u_k) f_i(x_i) &\leq u_k. \end{aligned}$$

□

The assumption of convex attraction functions implies that the marginal number of sales lost due to price increases is *decreasing*. This is in many cases a natural assumption. Moreover, it can be verified that the MNL, MCI and linear attraction demand models do satisfy it. Condition (2.6) is satisfied in certain revenue management problems where the columns of the matrix $\mathbf{A}$ represent the vectors of resources consumed in producing one unit of the respective products, and $\mathbf{u}$ is the vector of the inventories available from each resource. Suppose, for example, that each product represents a seat on the same flight but with different fare restrictions. Then, as long as there are more potential customers than seats on the flight, the condition is satisfied, because the parameters A_{ki} corresponding to the k^{th} capacity constraint are all equal to one, and $0 < u_k \leq 1$ (since the demands are normalized by the population size).

However, most problems of interest do not fall into this special class, thereby motivating our more general approach. For instance, if the customers choose between seats on different flights, some of the parameters A_{ki} will be set to 0, violating the condition. Thus, we consider the case of general input data in the remainder of this paper.

2.3 Market Share Reformulation

In this section, we transform problem (P) into the equivalent optimization problem (COP) over the space of market shares. This transformation eliminates the need to explicitly represent the nonlinear demand model in the constraints, while preserving the bounds on prices as linear constraints. We denote the fraction of lost sales and the market share of each product i in (2.1) by $\theta_0 = d_0(\mathbf{x})$ and $\theta_i = d_i(\mathbf{x})$, respectively. The attraction functions $f_1, \ldots, f_n$ are invertible since they are strictly decreasing by Assumption 2.1(i). Define the *inverse attraction function* [1] $g_i : \mathbb{R}_{++} \to \mathbb{R}$ as the inverse of f_i, for each product i. From (2.5), the prices corresponding to a given

[1]Although the inverse attraction functions always exist, they may not have a closed form for some complex demand models. In Section 2.4.2, we show that the objective function's derivatives can nevertheless be computed efficiently, allowing general purpose algorithms to be used.

vector of market shares $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_n)$ can thus be expressed as,

$$x_i = f_i^{-1}\left(\frac{d_i(\mathbf{x})}{d_0(\mathbf{x})}\right) = g_i\left(\frac{\theta_i}{\theta_0}\right), \qquad \text{for } i = 1, 2, \ldots, n. \tag{2.7}$$

Optimization problem (P) can be rewritten as (COP). The market shares θ_i play the role of decision variables, and the the prices x_i are represented as functions of $\boldsymbol{\theta}$. In addition to the original constraints in (P) there is a simplex constraint on the market shares, and strict positivity of the market shares is enforced. These constraints are implied in (P) since the fraction of lost sales and the market share of each product i in (2.1) naturally satisfy

$$\sum_{i=0}^{n} d_i(\mathbf{x}) = 1, \qquad \text{with} \qquad d_i(\mathbf{x}) > 0, \qquad \text{for } i = 0, 1, \ldots, n. \tag{2.8}$$

To solve (COP) in practice, we may relax the strict inequalities. As any of the market shares θ_i go to zero, some of the prices go to positive or negative infinity. The price bounds in (P) thus exclude such solutions. [2] These price bounds are captured in (COP) by extending the matrix $\mathbf{A}$ and vector $\mathbf{u}$. The new number of inequality constraints on the demands is $m' = m + 2n$, and the additional coefficients and right-hand sides are given by, $\forall i, j \in \{1, 2, \ldots, n\}$,

$$\begin{aligned} A_{m+2i-1,j} &= \begin{cases} f_i(\underline{x}_i) & \text{if } i \neq j \\ 1 + f_i(\underline{x}_i) & \text{if } i = j \end{cases}, \qquad & u_{m+2i-1} &= f_i(\underline{x}_i), \\ A_{m+2i,j} &= \begin{cases} -f_i(\overline{x}_i) & \text{if } i \neq j \\ -1 - f_i(\overline{x}_i) & \text{if } i = j \end{cases}, \qquad & u_{m+2i} &= -f_i(\overline{x}_i). \end{aligned} \tag{2.9}$$

The following lemma shows that (P) and (COP) are equivalent, in that there is a one-to-one correspondence between feasible points of the two problems which preserves the objective function value.

[2] Even in the absence of these price bounds, Assumption 2.1 ensures that an optimal solution to (COP) which is strictly positive in each component exists, when it satisfies the convexity condition of the next section. This follows from Proposition A.2 characterizing its dual in Appendix A.3.

Lemma 2.3. *Formulations* (P) *and* (COP) *are equivalent.*

Proof. Consider the problem

$$\begin{aligned} \max \ & \sum_i a_i x_i d_i(\mathbf{x}) \\ \text{s.t.} \ & \sum_i A_{ki} d_i(\mathbf{x}) \leq u_k \qquad k = 1 \ldots m', \end{aligned} \tag{P1}$$

where we have replaced the price bounds in (P) with the new constraints (2.9). Using the monotonicity of f_i and (2.5), the bounds on the price of the i^{th} product may by expressed in terms of the attractions as

$$\begin{aligned} \underline{x}_i \leq x_i \leq \overline{x}_i \ & \Leftrightarrow \ f_i(\underline{x}_i) \geq f_i(x_i) \geq f_i(\overline{x}_i) \\ & \Leftrightarrow \ f_i(\underline{x}_i) d_0(\mathbf{x}) \geq d_i(\mathbf{x}) \geq f_i(\overline{x}_i) d_0(\mathbf{x}). \end{aligned} \tag{2.10}$$

Using (2.8), the above can be written in terms of the market shares for all the products as

$$f_i(\underline{x}_i)\left(1 - \sum_{j=1}^{n} d_j(\mathbf{x})\right) \geq d_i(\mathbf{x}) \geq f_i(\overline{x}_i)\left(1 - \sum_{j=1}^{n} d_i(\mathbf{x})\right). \tag{2.11}$$

These are precisely the constraints described by the $(m+2i-1)^{\text{th}}$ and $(m+2i)^{\text{th}}$ rows of $\mathbf{A}$ and $\mathbf{u}$ in problem (P1). Consequently, the problems (P1) and (P) are equivalent.

We define the mapping $T : \mathcal{X} \to \Theta$ from the feasible region $\mathcal{X} \subseteq \mathbb{R}^n$ of (P) to the feasible region $\Theta \subset (0,1)^{n+1}$ of (COP) by $T(\mathbf{x}) = [d_0(\mathbf{x}), \ldots, d_n(\mathbf{x})]^\top$. The inverse mapping is given by $T^{-1}(\boldsymbol{\theta}) = \left[g_1\left(\frac{\theta_1}{\theta_0}\right), \ldots, g_n\left(\frac{\theta_n}{\theta_0}\right)\right]^\top$, as in (2.7). For any $\mathbf{x} \in \mathcal{X}$, $T(\mathbf{x})$ is feasible for (COP) because (i) the inequality constraints are equivalent to those of (P1) above, and (ii) the simplex and positivity constraints are satisfied by (2.8). Similarly, for any $\boldsymbol{\theta} \in \Theta$, $T^{-1}(\boldsymbol{\theta})$ is feasible for (P1) and (P). Finally, the objective value $\Pi(T(\mathbf{x}))$ of (COP) is equal to the objective value of (P). $\square$

2.3.1 Convexity condition

Theorem 2.4 below provides a condition on the attraction functions $f_1, f_2, \ldots, f_n$ under which (COP) is a convex problem, and can thus be solved with general purpose convex optimization algorithms. The condition is fairly general and requires only a property of the individual attraction functions. Corollary A.1 of Appendix A.2 verifies that it holds for MNL, MCI and linear attraction demand models.

In light of Lemma 2.3, Theorem 2.4 also provides insights into the structure of the original pricing problem (P). Specifically, it implies that there are no (strict) local maxima. In particular, note that together with Proposition 2.2, it implies that the pricing problems considered by [31] are in fact maximizations of a unimodal profit function over a convex feasible region. More generally, the theorem specifies a condition such that under Assumption 2.1 problem (P) does not have any strict local maxima, even when its feasible region is not convex.

Theorem 2.4. *If the attraction functions are such that, in the space of market shares,*

$$2g_i'(y) + yg_i''(y) \leq 0, \qquad \forall y > 0, \qquad i = 1, 2, \ldots, n, \tag{2.12}$$

or equivalently in the space of prices,

$$\frac{2\,|f_i'(x)|}{f_i(x)} \geq \frac{f_i''(x)}{|f_i'(x)|}, \qquad \forall x \in \mathbb{R}, \qquad i = 1, 2, \ldots, n, \tag{2.13}$$

then the objective $\Pi(\theta)$ *of* (COP) *is concave. Furthermore,*

(i) every local maximum of (COP) *is a global maximum, and*

(ii) every local maximum of (P) *is a global maximum.*

Proof. Since the objective function is $\Pi(\boldsymbol{\theta}) = \sum_{i=1}^{n} a_i \theta_i g_i(\frac{\theta_i}{\theta_0})$, with positive coefficients a_i, we need only show concavity of each term $\Pi_i(\theta_0, \theta_i) \triangleq \theta_i g_i(\frac{\theta_i}{\theta_0})$. The gradient and Hessian of $\Pi_i(\theta_0, \theta_i)$ are

$$\nabla \Pi_i = \begin{bmatrix} -\frac{\theta_i^2}{\theta_0^2} g_i'\left(\frac{\theta_i}{\theta_0}\right) \\ g_i\left(\frac{\theta_i}{\theta_0}\right) + \frac{\theta_i}{\theta_0} g_i'\left(\frac{\theta_i}{\theta_0}\right) \end{bmatrix} \qquad \text{and} \tag{2.14}$$

$$\nabla^2 \Pi_i = \left(2g_i'(\frac{\theta_i}{\theta_0}) + \frac{\theta_i}{\theta_0} g_i''(\frac{\theta_i}{\theta_0})\right) \begin{bmatrix} \frac{\theta_i^2}{\theta_0^3} & \frac{-\theta_i}{\theta_0^2} \\ \frac{-\theta_i}{\theta_0^2} & \frac{1}{\theta_0} \end{bmatrix}. \tag{2.15}$$

The first factor is non-positive by condition (2.12). Taking any vector $\mathbf{z} = [u, v]^{\mathsf{T}} \in \mathbb{R}^2$, we have that

$$\mathbf{z}^{\mathsf{T}} \begin{bmatrix} \frac{\theta_i^2}{\theta_0^3} & \frac{-\theta_i}{\theta_0^2} \\ \frac{-\theta_i}{\theta_0^2} & \frac{1}{\theta_0} \end{bmatrix} \mathbf{z} = \frac{1}{\theta_0}\left(u^2\left(\frac{\theta_i}{\theta_0}\right)^2 - 2uv\left(\frac{\theta_i}{\theta_0}\right) + v^2\right) = \frac{1}{\theta_0}\left(u\left(\frac{\theta_i}{\theta_0}\right) - v\right)^2 \geq 0,$$

so the Hessian of Π_i is negative semi-definite, and Π_i is concave.

Differentiating $x = g(y) \triangleq g_i(y)$, for some fixed i, with respect to y, we have

$$g'(y) = \left(f^{-1}(y)\right)' = \frac{1}{f'\left(f^{-1}(y)\right)} = \frac{1}{f'\left(g(y)\right)} = \frac{1}{f'\left(x\right)}, \tag{2.16}$$

and using the chain rule,

$$g''(y) = \frac{-1}{\left(f'\left(g(y)\right)\right)^2} f''\left(g(y)\right) g'(y) = \frac{-1}{\left(f'\left(x\right)\right)^2} f''\left(x\right) \frac{1}{f'(x)} = \frac{-f''(x)}{\left(f'\left(x\right)\right)^3}. \tag{2.17}$$

Substituting into (2.12) and multiplying by the strictly positive quantity $\left(f'(x)\right)^2$,

$$\frac{2}{f'(x)} + y\frac{-f''(x)}{\left(f'\left(x\right)\right)^3} \leq 0 \quad \Leftrightarrow \quad 2f'(x) - \frac{f(x)f''(x)}{f'(x)} \leq 0 \quad \Leftrightarrow \quad \frac{2f'(x)}{f(x)} \leq \frac{f''(x)}{f'(x)}.$$

We used the fact that $f'(x) < 0$ while $f(x) > 0$. The equivalence with (2.13) follows since $f'(x)$ is negative.

Then (i) follows directly from the concavity of the objective function and convexity of the feasible region in (COP). As shown in the proof of Lemma 2.3, there is a one-to-one, invertible, continuous mapping between feasible points of (COP) and (P), and the mapping preserves the value of the continuous objective function. Thus any local maximum of (P) would correspond to a local maximum of (COP). □

2.3.2 Self-Concordant Barrier Method for the MNL Demand Model

In this section, we restrict our attention to the MNL demand model, and show that problem (COP) may be solved in polynomial time using interior point methods. In particular, we show that applying the barrier method to (COP) gives rise to a *self-concordant* objective. (The latter concept is defined below.)

Note that a similar treatment could be applied to other attraction demand models, but for ease of exposition, we focus on the more common MNL demand model. Other optimization algorithms could also be applied. The barrier subproblem (2.18) below minimizes a twice-differentiable convex function over a simplex for *any* attraction functions satisfying the conditions of Theorem 2.4. [2] show that a first-order modified Frank-Wolfe algorithm exhibits linear convergence for such problems. In Section 2.6, we use a commercially available implementation of a primal-dual interior-point method to solve (CMNL).

Under the MNL model, the attraction functions and their inverses are defined for $i = 1, \ldots, n$ as $f_i(x_i) \triangleq v_i e^{-x_i}$, and $g_i(y_i) = -\log(y_i/v_i)$. Then problem (COP) is

$$\begin{aligned} \max \quad & \Pi(\boldsymbol{\theta}) = -\sum_{i=1}^{n} a_i \theta_i \log \frac{\theta_i}{v_i \theta_0} \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{\theta} \le u, \qquad \mathbf{e}^\top \boldsymbol{\theta} = 1, \qquad \boldsymbol{\theta} > 0 \end{aligned} \tag{CMNL}$$

where $\mathbf{e}$ denotes the vector of ones. We note that the objective has a form similar to the relative entropy, or Kullback-Leibler (KL) divergence,

$$\mathcal{K}(\boldsymbol{\pi}, \boldsymbol{\eta}) \triangleq \sum_{i=1}^{n} \pi_i \log \frac{\pi_i}{\eta_i}.$$

This is a measure of the distance between two probability distributions $\boldsymbol{\pi}, \boldsymbol{\eta} \in \mathbb{R}^n_+, \mathbf{e}^\top \boldsymbol{\pi} = \mathbf{e}^\top \boldsymbol{\eta} = 1$. For problems involving the KL-divergence where the denominators η_i are *constant*, the existence of a self-concordant barrier is known (see [16] and [21]). Unfortunately these results cannot be used in our setting, since the objective $\Pi(\boldsymbol{\theta})$ is *not* separable: each term also depends on the *decision variable* θ_0.

The barrier method solves (CMNL) by solving a series of problems parameterized by $t > 0$,

$$\begin{aligned} \min \quad & \Psi_t(\boldsymbol{\theta}) = -t\Pi(\boldsymbol{\theta}) + \Phi(\boldsymbol{\theta}) \\ \text{s.t.} \quad & \mathbf{e}^\top \boldsymbol{\theta} = 1 \end{aligned} \tag{2.18}$$

where the logarithmic barrier is defined by

$$\Phi(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log \theta_i - n \log \theta_0 - \sum_{k=1}^{m'} \log \left(u_k - \sum_{i=1}^{n} A_{ki}\theta_i \right). \tag{2.19}$$

We have changed the maximization problem to a minimization problem for consistency with the literature. Each inequality constraint in (CMNL) has been replaced by a term in the barrier which goes to infinity as the constraint becomes tight. The term for the positivity constraint on θ_0 is replicated n times for reasons that will become apparent. We let $\bar{m} = m' + 2n = m + 4n$ be the number of inequality constraints in a slightly modified version of problem (CMNL), including the $m' = m + 2n$ constraints represented by the pair $(\mathbf{A}, \mathbf{u})$, the n positivity constraints on the market share of each product, and n replications of the positivity constraint $\theta_0 > 0$, for which n logarithmic barrier terms were added to the barrier (2.19).

Denote the optimal solution of (2.18) for a given value of $t > 0$ by $\boldsymbol{\theta}^*(t)$. Given an appropriate, strictly feasible starting point $\boldsymbol{\theta}$, a positive initial value for t, a constant factor $\mu > 1$ and a tolerance $\epsilon > 0$, the barrier method consists of the following steps:

1. Solve (2.18) using equality-constrained Newton's method with starting point $\boldsymbol{\theta}$ to obtain $\boldsymbol{\theta}^*(t)$.

2. Update the starting point $\boldsymbol{\theta} := \boldsymbol{\theta}^*(t)$.

3. Stop if $\bar{m}/t \leq \epsilon$, otherwise update $t := \mu t$ and go to Step 1.

As the value of t becomes large, the solution $\boldsymbol{\theta}^*(t)$ tends towards the optimal solution of (CMNL). The termination condition in Step 3 guarantees that the objective value is sufficiently close to its optimal value. In practice, a phase I problem must be solved to find an appropriate initial point $\boldsymbol{\theta}$, and a redundant constraint is added to (CMNL) for technical reasons.

The computational complexity of Newton's method can be analyzed when the objective function is *self-concordant.*

Definition 2.5 (Self-concordance.). A convex scalar function $f : \mathcal{D} \to \mathbb{R}$ is said to be self-concordant when $|f'''(x)| \leq 2\,(f''(x))^{3/2}$ for every point $x \in \mathcal{D} \subseteq \mathbb{R}$ in the domain of f. A multivariate function $f : \mathcal{F} \to \mathbb{R}$ is self-concordant if it is self-concordant along every line in its domain $\mathcal{F} \subseteq \mathbb{R}^n$.

The class of self-concordant functions is closed under addition and composition with affine functions (see, for example, [14]). Our proof that the objective function of the problem (2.18) falls withing this class relies on Theorem 2.6 presented here.

Theorem 2.6. *The function*

$$f(x,y) = tx \log \frac{x}{\beta y} - \log xy \tag{2.20}$$

is strictly convex and self-concordant on $\mathbb{R}^2_{++}$ for $\beta > 0$ and $t \geq 0$.

Proof. We explicitly compute the derivatives of (2.20) and obtain

$$\nabla f(x,y) = \begin{bmatrix} t + t\log\frac{x}{y} - \frac{1}{x} - t\log\beta \\ -\frac{tx+1}{y} \end{bmatrix}, \quad \nabla^2 f(x,y) = \begin{bmatrix} \frac{tx+1}{x^2} & \frac{-t}{y} \\ \frac{-t}{y} & \frac{tx+1}{y^2} \end{bmatrix},$$

$$\text{and} \qquad \nabla^3 f(x,y) = \begin{bmatrix} -\frac{tx+2}{x^3} & 0 \\ 0 & \frac{t}{y^2} \end{bmatrix} \begin{bmatrix} 0 & \frac{t}{y^2} \\ \frac{t}{y^2} & \frac{-2tx-2}{y^3} \end{bmatrix}.$$

Then for an arbitrary direction $\mathbf{h} = [a, b]^\top \in \mathbb{R}^2$ and any $[x, y]^\top \in \mathbb{R}^2_{++}$

$$\nabla^2 f[\mathbf{h},\mathbf{h}] = \frac{a^2(tx+1)}{x^2} - \frac{2abt}{y} + \frac{b^2(tx+1)}{y^2}$$

For ease of notation, define $s \geq 0$ and $u, v \in \mathbb{R}$ such that $s = tx \geq 0$, $u = a/x$ and $v = b/y$. Then rewrite

$$\begin{aligned} \nabla^2 f[\mathbf{h},\mathbf{h}] &= u^2(s+1) - 2uvs + v^2(s+1) = s(u^2 - 2uv + v^2) + (u^2 + v^2) \\ &= s\underbrace{(u-v)^2}_{A} + \underbrace{(u^2+v^2)}_{B}. \end{aligned}$$

Both terms A and B are non-negative, and the second term B is positive unless

$u = v = 0$, i.e. $a = b = 0$. Thus the Hessian $\nabla^2 f$ is positive definite and f is strictly convex on $\mathbb{R}_{++}$. We also expand

$$\begin{aligned}
-\nabla^3 f[\mathbf{h}, \mathbf{h}, \mathbf{h}] &= \frac{a^3(tx+2)}{x^3} - \frac{3ab^2t}{y^2} + \frac{2b^3(tx+1)}{y^3} \\
&= u^3(s+2) - 3uv^2 s + 2v^3(s+1) \\
&= s(u^3 - 3uv^2 + 2v^3) + 2(u^3 + v^3) \\
&= s\underbrace{(u-v)^2(u+2v)}_{K} + \underbrace{2(u^3+v^3)}_{L}.
\end{aligned}$$

We now show that f is self-concordant, that is

$$|\nabla^3 f[\mathbf{h}, \mathbf{h}, \mathbf{h}]| \leq 2(\nabla^2 f[\mathbf{h}, \mathbf{h}])^{\frac{3}{2}} \quad \Leftrightarrow \quad (\nabla^3 f[\mathbf{h}, \mathbf{h}, \mathbf{h}])^2 \leq 4(\nabla^2 f[\mathbf{h}, \mathbf{h}])^3, \tag{2.21}$$

by appropriately factoring the difference of the two sides of the inequality, and showing that it is non-negative. That is, we verify the non-negativity of

$$\begin{aligned}
4(\nabla^2 f[\mathbf{h}, \mathbf{h}])^3 - (\nabla^3 f[\mathbf{h}, \mathbf{h}, \mathbf{h}])^2 &= 4(sA+B)^3 - (sK+L)^2 \\
&= (4s^3A^3 + 12s^2A^2B + 12sAB^2 + 4B^3) - (s^2K^2 + 2sKL + L^2) \\
&= 4s^3A^3 + s^2(12A^2B - K^2) + s(12AB^2 - 2KL) + (4B^3 - L^2). \qquad (2.22)
\end{aligned}$$

For the leading term that $A^3 = (u-v)^6 \geq 0$. Then for the second term

$$\begin{aligned}
12A^2B - K^2 &= 12(u-v)^4(u^2+v^2) - (u-v)^4(u+2v)^2 \\
&= (u-v)^4\left(12u^2 + 12v^2 - (u+2v)^2\right) \\
&= (u-v)^4(11u^2 - 4uv + 8v^2) \geq 0,
\end{aligned}$$

since the quadratic form can be written as $\left[u,v \right] \left[\begin{smallmatrix} 11 & -2 \\ -2 & 8 \end{smallmatrix} \right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix} \right] \geq 0$. For the term in s,

$$\begin{aligned} 12AB^2 - 2KL &= 12(u-v)^2(u^2+v^2)^2 - 4(u-v)^2(u+2v)(u^3+v^3) \\ &= 4(u-v)^2 \left(3(u^4+2u^2v^2+v^4) - (u^4+uv^3+2u^3v+2v^4)\right) \\ &= 4(u-v)^2 \left(2u^2(u^2-uv+v^2) + v^2(4u^2-uv+v^2)\right) \geq 0. \end{aligned}$$

Finally, the last term is $4B^3 - L^2 = 4(u^2+v^2)^6 - 4(u^3+y^3)^2 = 4u^2v^2(3u^2-2uv+3v^2) \geq 0$. Together, the preceding four inequalities with the fact that $s \geq 0$ show that (2.22) is non-negative, and that (2.21) holds. □

In order to prove Theorem 2.7, we shall use the result of Section 11.5.5 in [14]. The result applies to minimization problems, but the objective function of the maximization problem (CMNL) can be negated to obtain an equivalent minimization problem. [3] We first define some additional notation. The constant M is an *a priori* lower bound on the optimal value of (CMNL) (and thus an upper bound for the corresponding minimization problem). It is used in the phase I feasibility problem. The constant G is an upper bound on the norm of the gradient of the constraints. Since the positivity constraints have a gradient of norm 1, and the gradients of the inequality constraints are the rows of the matrix $\mathbf{A}$, which we denote here by $\mathbf{A}_{k,\cdot}$, we set

$$G = \max \left\{ 1, \max_{1 \leq k \leq m'} \|\mathbf{A}_{k,\cdot}\| \right\}.$$

We define R to be the radius of a ball centered at the origin containing the feasible set. Since any feasible vector $\boldsymbol{\theta}$ lies in the unit simplex, we may set $R = 1$. We define two constants depending on the parameters of the backtracking line search algorithm used in Newton's method. We let $\gamma = \frac{\alpha\beta(1-2\alpha)^2}{20-8\alpha}$ and $c = \log_2 \log_2 \frac{1}{\epsilon}$. Typical values are $\alpha \in [0.01, 0.3]$ and $\beta \in [0.1, 0.8]$. The constant c can reasonably be approximated by $c = 6$. Finally, we let $p^* > M$ be the optimal value of (CMNL), and we define $\bar{p}^*$ to be the optimal value of the phase I feasibility problem used to find a suitable

[3] Because of this negation, the values of M and p^* defined below are also the negation of the corresponding values in [14]. Therefore the phase I minimization problem is unchanged, but the objective function of the phase II minimization problem is the negation of the objective of (CMNL).

starting point (see Section 11.5.4 of [14]). The latter value is close to zero when a problem is nearly infeasible or nearly feasible, and is far from zero if the problem is clearly feasible or infeasible.

The bound of Theorem 2.7 depends on the number of constraints and the number of products through $\bar{m} = m + 4n$, and on two terms that depend on the problem data, $C_1 = \log_2 \frac{G}{|\bar{p}^*|}$ and $C_2 = \log_2 \frac{p^* - M}{\epsilon}$. The constant C_1 should be interpreted as measuring the difficulty of the phase I feasibility problem, while C_2 can be interpreted as measuring the difficulty of solving the phase II problem.

Theorem 2.7. *Problem* (CMNL) *may be solved to within a tolerance* $\epsilon > 0$ *in a polynomial number* $N = N_I + N_{II}$ *iterations of Newton's method, where*

$$N_I = \left\lceil \sqrt{\bar{m}+2} \log_2 \left(\frac{(\bar{m}+1)(\bar{m}+2)GR}{|\bar{p}^*|} \right) \right\rceil \left(\frac{1}{2\gamma} + c \right)$$

is the number of iterations required to solve the phase I problem, and

$$N_{II} = \left\lceil \sqrt{\bar{m}+1} \log_2 \left(\frac{(\bar{m}+1)(p^* - M)}{\epsilon} \right) \right\rceil \left(\frac{1}{2\gamma} + c \right)$$

is the number of iterations required to solve the phase II problem.

Proof. The objective of (2.18) is

$$\begin{aligned} \Psi_t(\boldsymbol{\theta}) &= -t\Pi(\theta) + \Phi(\theta) \\ &= t \sum_{i=1}^{n} a_i \theta_i \log(\theta_i/(\theta_0 v_i)) - \sum_{i=1}^{n} \log \theta_i - n \log \theta_0 - \sum_{k=1}^{m'} \log(u_k - \sum_{i=1}^{n} A_{ki}\theta_i) \\ &= \sum_{i=1}^{n} (a_i t \theta_i \log(\theta_i/(\theta_0 v_i)) - \log \theta_i \theta_0) - \sum_{k=1}^{m'} \log(u_k - \sum_{i=1}^{n} A_{ki}\theta_i). \end{aligned} \tag{2.23}$$

We show that Φ is a self-concordant barrier for (CMNL), that is, the function $\Psi_t(\boldsymbol{\theta})$ is self-concordant, convex and closed on the domain $\{\boldsymbol{\theta} \in \mathbb{R}^{n+1} : \boldsymbol{\theta} > 0, \mathbf{e}^\top \boldsymbol{\theta} = 1, \mathbf{A}\boldsymbol{\theta} < \mathbf{u}\}$. The terms of the first summation in (2.23) are self-concordant and convex by Theorem 2.6, since $a_i > 0, \forall i$. The function $-\log x$ is self-concordant and convex, and so are the terms of the second summation since both properties are preserved by composition with an affine function. Finally, Ψ_t is self-concordant and convex since both properties are preserved through addition. The function is closed on its domain since the barrier terms become infinite at the boundary, and all terms are bounded from below.

We observe that the level sets of (CMNL) are bounded, since the feasible set is bounded. We can now apply the result of Section 11.5.5 in [14] to show that no more than N_{I} Newton steps are required to solve a phase I problem yielding an initial strictly feasible point on the central path of an appropriate auxiliary phase II problem. The phase II problem may in turn be solved to within tolerance ϵ in at most N_2 Newton steps. The total number of Newton steps required to solve (CMNL) is thus at most $N_{\mathrm{I}} + N_{\mathrm{II}}$. □

2.4 Multiple Overlapping Customer Segments

In many cases, it is desirable to represent a number of customer segments, such as when business and leisure travelers are buying the same airline tickets. Suppose that different attraction demand models are available for each segment of the population. Regardless of her segment, a customer may purchase any of the products offered, but her choice probabilities depend on her particular segment. That is, we would like to be able to optimize over a demand model of the form

$$d_i^{\mathrm{MIX}}(\mathbf{x}) = \sum_{\ell=1}^{L} \Gamma_\ell d_i^\ell(\mathbf{x}) = \sum_{\ell=1}^{L} \Gamma_\ell \frac{f_i^\ell(x_i)}{1 + \sum_{j=1}^{n} f_j^\ell(x_j)}, \tag{2.24}$$

where the mixture coefficient Γ_ℓ represents the relative size of the ℓ^{th} market segment, whose demand is itself modeled by an attraction demand model. To continue representing demand as a fraction of the population, we assume that $\sum_{\ell=1}^{L} \Gamma_\ell = 1$, and $\Gamma_\ell > 0, \forall \ell$. We define the notation $d_0^\ell(\mathbf{x}), d_1^\ell(\mathbf{x}), \ldots, d_n^\ell(\mathbf{x})$ for the lost sales and demand functions of the ℓ^{th} segment, as in (2.1).

We point out that this model implicitly assumes that consumers from each segment may purchase *any* product. It is more general than the models from network revenue management that assume consumers only purchase products specific to their segment. Similarly, the work of [69] assumes that the retailer can set different prices for each of the segments. Both of these situations are better represented by standard attraction demand models, as discussed in Section 2.2.1.

On the other hand, the mixture of attraction demand models defined in (2.24) is *not* itself an attraction model. How to efficiently solve the pricing problem with multiple segments to optimality remains an open problem, and is beyond the scope of this paper. [40] have shown that the pricing objective may have multiple local maxima, and solution methods from network revenue management give rise to NP-hard sub-problems (see, e.g., [60], who solve them heuristically). Instead, we propose an approximation to the multi-segment demand functions $d_i^{\mathrm{MIX}}(\mathbf{x})$ by a valid attraction demand model.

2.4.1 Approximation by an Attraction Demand Model

Aydin and Porteus [5] suggest (for the specific case of MNL models) the following approximation, based on valid attraction functions

$$\bar{f}_i(x_i) = \sum_{\ell=1}^{L} \gamma_\ell f_i^\ell(x_i), \qquad i = 1, \dots, n, \tag{2.25}$$

where the coefficients $\gamma_1, \dots, \gamma_L \in \mathbb{R}_+$ are set equal to the segment sizes Γ_ℓ of (2.24). We also introduce the notation $\bar{d}_1(\mathbf{x}), \dots, \bar{d}_n(\mathbf{x})$ for the approximate demands when using the attraction functions (2.25). We define

$$\Pi^{\mathrm{MIX}}(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i d_i^{\mathrm{MIX}}(\mathbf{x}) \qquad \text{and} \qquad \bar{\Pi}(\mathbf{x}) = \sum_{i=1}^{n} a_i x_i \bar{d}_i(\mathbf{x})$$

as the exact and approximated profit functions, respectively.

In Theorem 2.8 we show that setting coefficients γ_ℓ as in (2.27) instead yields a local approximation to the desired multi-segment model (2.24) for prices near some reference point $\mathbf{x^0} \in \mathbb{R}^n$. In particular, our approximation is exact at the reference price $\mathbf{x} = \mathbf{x^0}$.

Theorem 2.8. *If the sets of attraction functions $\{f_i^\ell, i = 1, \ldots, n\}$ satisfy Assumption 2.1 for $\ell = 1, \ldots, L$, then so do the attraction functions $\bar{f}_1, \ldots, \bar{f}_n$ defined in* (2.25).

Furthermore, suppose that for some constant $B > 0$ and reference prices $\mathbf{x^0} \in \mathbb{R}^n$ the attraction functions satisfy the local Lipschitz conditions

$$|f_i^\ell(x_i) - f_i^\ell(x_i^0)| \leq B|x_i - x_i^0|, \quad \forall \mathbf{x} \in \mathcal{X}, \quad i = 1, \ldots, n, \quad \ell = 1, \ldots, L, \tag{2.26}$$

where $\mathcal{X} \subseteq \{\mathbf{x} : \|\mathbf{x} - \mathbf{x^0}\|_1 < 1/B\} \subset \mathbb{R}^n$ is a set around the reference prices $\mathbf{x^0}$. Let the coefficients of the approximation be

$$\gamma_\ell = \frac{\Gamma_\ell d_0^\ell(\mathbf{x^0})}{\sum_{\ell=1}^L \Gamma_\ell d_0^\ell(\mathbf{x^0})}, \qquad \ell = 1, \ldots, L. \tag{2.27}$$

Then the approximate demand functions $\bar{d}_1, \ldots, \bar{d}_n$ satisfy,

$$(1 - \epsilon_x) d_i^{MIX}(\mathbf{x}) \quad \leq \quad \bar{d}_i(\mathbf{x}) \quad \leq \quad (1 + \epsilon_x) d_i^{MIX}(\mathbf{x}), \qquad \forall x \in \mathcal{X},$$

where $\epsilon_x = \frac{2B\|\mathbf{x}-\mathbf{x^0}\|_1}{1-B\|\mathbf{x}-\mathbf{x^0}\|_1}$. Moreover, if the feasible prices are positive, i.e., $\mathcal{X} \subset \mathbb{R}^n_+$, the approximate profit function $\bar{\Pi}(\mathbf{x})$ satisfies

$$(1 - \epsilon_x)\Pi^{MIX}(\mathbf{x}) \quad \leq \quad \bar{\Pi}(\mathbf{x}) \quad \leq \quad (1 + \epsilon_x)\Pi^{MIX}(\mathbf{x}), \qquad \forall x \in \mathcal{X}.$$

Proof. Assumption 2.1 holds since (i) the sum of decreasing functions is decreasing and the sum of differentiable functions is differentiable, and (ii) the limit of a finite sum is the sum of the limits.

Since from the choice of coefficients $\sum_{\ell=1}^{\ell} \gamma_\ell = 1$, we rewrite

$$\begin{aligned}
\bar{d}_i(\mathbf{x}) &= \frac{\bar{f}_i(x_i)}{1 + \sum_{j=1}^n \bar{f}_j(x_j)} = \frac{\sum_{\ell=1}^L \gamma_\ell f_i^\ell(x_i)}{1 + \sum_{j=1}^n \sum_{\ell=1}^L \gamma_\ell f_j^\ell(x_j)} = \frac{\sum_{\ell=1}^L \gamma_\ell f_i^\ell(x_i)}{\sum_{\ell=1}^L \gamma_\ell \left(1 + \sum_{j=1}^n f_j^\ell(x_j)\right)} \\
&= \frac{\sum_{\ell=1}^L \Gamma_\ell d_0^\ell(\mathbf{x^0}) \left(d_i^\ell(\mathbf{x})/d_0^\ell(\mathbf{x})\right)}{\sum_{\ell=1}^L \Gamma_\ell d_0^\ell(\mathbf{x^0}) \left(1/d_0^\ell(\mathbf{x})\right)} = \sum_{\ell=1}^L \frac{\left(d_0^\ell(\mathbf{x^0})/d_0^\ell(\mathbf{x})\right)}{\sum_{\ell=1}^L \Gamma_\ell \left(d_0^\ell(\mathbf{x^0})/d_0^\ell(\mathbf{x})\right)} \Gamma_\ell d_i^\ell(\mathbf{x})
\end{aligned} \tag{2.28}$$

where we use fact (2.5) in the fourth equality. The ratios appearing in the last expression can be expressed as

$$\frac{d_0^\ell(\mathbf{x^0})}{d_0^\ell(\mathbf{x})} = \frac{1+\sum_{i=1}^n f_i^\ell(x_i)}{1+\sum_{i=1}^n f_i^\ell(\overline{x}_i)} = 1 + d_0^\ell(\mathbf{x^0}) \sum_{i=1}^n \left(f_i^\ell(x_i) - f_i^\ell(\overline{x}_i)\right)$$

where $d_0^\ell(\mathbf{x^0}) < 1$. Then, using assumption (2.26), we obtain

$$1 - B\|\mathbf{x}-\mathbf{x^0}\|_1 \leq \frac{d_0^\ell(\mathbf{x^0})}{d_0^\ell(\mathbf{x})} \leq 1 + B\|\mathbf{x}-\mathbf{x^0}\|_1.$$

Note that the lower bound is non-negative by the definition of $\mathcal{X}$. Since $\sum_{\ell=1}^\ell \Gamma_\ell = 1$, we obtain from (2.28)

$$\bar{d}_i(\mathbf{x}) \leq \frac{1+B\|\mathbf{x}-\mathbf{x^0}\|_1}{1-B\|\mathbf{x}-\mathbf{x^0}\|_1} \sum_{\ell=1}^\ell \Gamma_\ell d_i^\ell(\mathbf{x}) = \frac{1+B\|\mathbf{x}-\mathbf{x^0}\|_1}{1-B\|\mathbf{x}-\mathbf{x^0}\|_1} d_i^{\text{MIX}}(\mathbf{x})$$

$$\bar{d}_i(\mathbf{x}) \geq \frac{1-B\|\mathbf{x}-\mathbf{x^0}\|_1}{1+B\|\mathbf{x}-\mathbf{x^0}\|_1} \sum_{\ell=1}^\ell \Gamma_\ell d_i^\ell(\mathbf{x}) = \frac{1-B\|\mathbf{x}-\mathbf{x^0}\|_1}{1+B\|\mathbf{x}-\mathbf{x^0}\|_1} d_i^{\text{MIX}}(\mathbf{x})$$

Defining shorthand $a = B\|\mathbf{x}-\mathbf{x^0}\|_1 < 1$, notice that

$$\frac{1+a}{1-a} = 1 + \frac{2a}{1-a} = 1 + \epsilon_x, \qquad \text{and} \qquad \frac{1-a}{1+a} = 1 - \frac{2a}{1+a} \geq 1 - \frac{2a}{1-a} = 1 - \epsilon_x.$$

The statement regarding the profit function follows immediately if the prices are also positive, by bounding each term of $\bar{\Pi}(\mathbf{x})$ individually. □

An appropriate set $\mathcal{X}$ can be obtained by taking any set for which the Lipschitz condition (2.26) holds and restricting it to $\{\mathbf{x} : \|\mathbf{x}-\mathbf{x^0}\|_1 < 1/B\}$. Clearly, the accuracy of the approximation is highly dependent on the smoothness of the attraction functions. Such limitations are to be expected since the profit function $\Pi^{\text{MIX}}(\mathbf{x})$ may have multiple local maxima, while our approximation $\bar{\Pi}(\mathbf{x})$ cannot by Theorem 2.4 (for common demand models).

2.4.2 Solving the Approximated Pricing Problem

Although the approximation presented in the preceding section uses an attraction demand model, the corresponding (COP) formulations cannot be solved directly in practice even if Assumption 2.1 is satisfied. This is because, unlike for simpler attraction demand models, the attraction functions (2.25) do not have closed-form inverses. The same issue arises under other complex attraction demand models, such as the semi-parametric attraction models proposed by [41].

To solve (COP) using standard nonlinear optimization algorithms, we need to evaluate the gradient and the Hessian matrix of the objective function $\Pi(\boldsymbol{\theta})$ for any market shares $\boldsymbol{\theta}$. First, we note that the prices $\mathbf{x}$ corresponding to market shares $\boldsymbol{\theta}$ can be obtained efficiently. The equations (2.7) are equivalent to $f_i(x_i) = \theta_i/\theta_0, i = 1, \ldots, n$. Since the attraction functions f_i are decreasing by Assumption 2.1, they have a unique solution and may be solved by n one-dimensional line searches. The following proposition then shows how the partial derivatives of the objective may be recovered from the prices $x_1, \ldots, x_n$ corresponding to the market shares $\boldsymbol{\theta}$, and from the derivatives of the original attraction functions $f_1, \ldots, f_n$. In particular, the derivatives of each $\bar{f}_i$ in (2.25) are readily obtained from those of $f_i^1, \ldots, f_i^L$.

Proposition 2.9. *Let $x_1 \ldots x_n$ be the unique prices solving the equations (2.7) for a given market share vector $\boldsymbol{\theta} \in \mathbb{R}_+^{n+1}$. Then the elements of the gradient $\nabla\Pi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ of the objective function $\Pi(\boldsymbol{\theta})$ are*

$$\frac{\partial \Pi}{\partial \theta_0} = -\sum_{i=1}^{n} a_i \frac{(f_i(x_i))^2}{f_i'(x_i)}, \qquad \text{and} \qquad \frac{\partial \Pi}{\partial \theta_i} = a_i \left(x_i + \frac{f_i(x_i)}{f_i'(x_i)} \right), \qquad i = 1, \ldots n.$$

The elements of the Hessian $\nabla^2\Pi(\boldsymbol{\theta})$ are $\partial^2\Pi/\partial\theta_i\partial\theta_j = 0$, for $1 \le i < j \le n$,

$$\begin{aligned}
\frac{\partial^2 \Pi}{\partial \theta_0^2} &= \sum_{i=1}^{n} \frac{a_i}{\theta_0} \left(\frac{2(f_i(x_i))^2}{f_i'(x_i)} - \frac{(f_i(x_i))^3 f_i''(x_i)}{(f_i'(x_i))^3} \right), && \\
\frac{\partial^2 \Pi}{\partial \theta_i^2} &= \frac{a_i}{\theta_0} \left(\frac{1}{f_i'(x_i)} - \frac{f_i(x_i) f_i''(x_i)}{(f_i'(x_i))^3} \right), && \text{for } i = 1, \ldots, n, \\
\text{and} \quad \frac{\partial^2 \Pi}{\partial \theta_i \partial \theta_0} &= -\frac{a_i}{\theta_0} \left(\frac{f_i(x_i)}{f_i'(x_i)} - \frac{(f_i(x_i))^2 f_i''(x_i)}{(f_i'(x_i))^3} \right), && \text{for } i = 1, \ldots, n.
\end{aligned}$$

Proof. The quantities in the statement are obtained by summing

$$\nabla\Pi(\boldsymbol{\theta}) = \sum_{i=1}^{n} a_i \nabla\Pi_i(\boldsymbol{\theta}) \qquad \text{and} \qquad \nabla^2\Pi(\boldsymbol{\theta}) = \sum_{i=1}^{n} a_i \nabla^2\Pi_i(\boldsymbol{\theta}),$$

with the non-zero elements of the terms $\nabla\Pi_i(\boldsymbol{\theta})$ and $\nabla^2\Pi_i(\boldsymbol{\theta})$ given in (2.14) and (2.15). (By a slight abuse of notation, we now consider the terms $\Pi_i(\boldsymbol{\theta})$ to be functions of the entire market share vector $\boldsymbol{\theta}$ instead of only the variables θ_0 and θ_i on which they each depend.) Then, we substitute in $f_i(x_i)$ and its first and second derivatives using (2.16) and (2.17). □

2.5 The Dual Problem

The structure of the pricing problem (P) goes beyond concavity of the transformed objective. Notice that the reformulation (COP) is not *separable* over the market shares $\theta_1, \ldots, \theta_n$ only because of the occurrence of θ_0 in each term of the objective. Nevertheless, as is often the case with separable problems, its Lagrangian dual yields a tractable decomposition. The dual of (COP) is

$$\begin{aligned} \min \quad & \mu + \sum_{k=1}^{m'} \lambda_k u_k \\ \text{s.t.} \quad & \mu = \sum_{i=1}^{n} \max_{y_i > 0} \phi_i(y_i, \boldsymbol{\lambda}, \mu) \\ & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{DCOP}$$

where we define for $i = 1, 2, \ldots, n$

$$\phi_i(y, \boldsymbol{\lambda}, \mu) \triangleq y \left(a_i g_i(y) - \sum_{k=1}^{m'} \lambda_k A_{ki} - \mu \right), \qquad y > 0. \tag{2.29}$$

The dual (DCOP) is expressed in terms of one-dimensional maximization problems for each product. These subproblems are coupled through a *single* linear constraint.

From a practical point of view, the dual problem does not require working with the inverse attractions, or with their derivatives, directly. A column generation algorithm for solving the dual is provided in Appendix A.3, along with the derivation of the dual itself. The algorithm requires the solution of a linear program and n one-dimensional maximization problems involving the original attraction functions f_i at each iteration (as opposed to their inverses g_i). It can be used even when the convexity condition (2.12) is not satisfied, or when the derivatives of the attraction functions described in Proposition 2.9 are not readily available. Moreover, Proposition A.4 of Appendix A.3 provides an alternate condition on the attraction functions which guarantees that a unique primal solution corresponds to each dual solution, without requiring the convexity condition (2.12).

For the special case of the MNL demand model, the inner maximization problems can in fact be solved in closed form, yielding the following dual problem.

$$\begin{aligned} \min \quad & \mu + \sum_{k=1}^{m'} \lambda_k u_k \\ \text{s.t.} \quad & \mu \geq \sum_{i=1}^{n} a_i v_i \exp\left\{-1 - \frac{\sum_{k=1}^{m'} \lambda_k A_{ki} + \mu}{a_i}\right\} \\ & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{DMNL}$$

The minimization (DMNL) is a convex optimization problem. As a result, it can also be solved with general-purpose algorithms. However, our experiments in the next section suggest this is less efficient than solving the primal (COP) directly.

2.6 Computational Experiments

We have proposed three formulations of the pricing problem in (P), (COP) and (DCOP). First, we have shown that under certain conditions, any local maximum of the non-convex problem (P) in terms of the prices is also a global maximum. Second, under the same conditions, the equivalent problem (COP) is a convex optimization problem with linear constraints. Third, we can recover a solution from the dual problem (DCOP).

To compare the efficiency of the three formulations, we evaluate the solution times of instances with an MNL demand model using the commercial LOQO solver (see Vanderbei and Shanno [79], Shanno and Vanderbei [71]). This solver uses a primal-dual interior point algorithm for sequential quadratic programming. It was chosen because it is commercially available and is intended for both convex and non-convex problems. However, LOQO does *not* employ the barrier method analyzed in Section 2.3.2. The AMPL (see Fourer et al. [27]) modeling language provides automatic differentiation for all problems. All the experiments were run on computer with dual 2.83GHz Intel Xeon CPUs and 32GB of RAM.

The MNL demand model was chosen since it allows (DCOP) to be solved directly. The demand model parameters are sampled as described in Appendix A.1.1 to ensure that the aggregate demand is near 0 and 1 as the prices approach the bounds $\overline{x}_i$ and $\underline{x}_i = 0$, respectively, for each product i. Constraints are sampled uniformly from the tangents to the sphere of radius $\frac{1}{2} \cdot \frac{1}{n+1}$ centered at the uniform distribution $\theta_0 = \theta_1 = \ldots = \theta_n = \frac{1}{n+1}$. Specifically, the k^{th} constraint is defined by

$$\mathbf{z}_k^\top \left(\boldsymbol{\theta} - \frac{1}{n+1}\mathbf{e}\right) \leq \frac{1}{2} \cdot \frac{1}{n+1} \qquad \Leftrightarrow \qquad \mathbf{z}_k^\top \boldsymbol{\theta} \leq \frac{1}{n+1}\left(\frac{1}{2} + \mathbf{z}_k^\top \mathbf{e}\right), \tag{2.30}$$

where $\mathbf{z}_k$ is sampled uniformly from the unit sphere centered at the origin, and $\mathbf{e}$ is the vector of all ones. This choice ensures that we do not generate any redundant constraints, and that a number of constraints are likely to be active at optimality.

Table 2.1 shows the average number of iterations and the average solution times over 10 randomly generated instances of various size when solving each of the three formulations. We note that for the market share formulation (COP) and the dual formulation (DCOP), the price bounds on x_i are converted to linear constraints by replacing $d_i(\mathbf{x})$ with θ_i in equation (2.10) of Lemma 2.3, yielding a total of $m' = m+2n$ constraints. However, the additional $2n$ constraints are sparse and we do not expect them to be active at optimality. In contrast, using (2.11) would result in dense constraints for (COP). Default parameters are used for the LOQO solver except that the tolerance is reduced from 8 to 6 significant digits of agreement between the primal

Products (n)	Constraints (m)	Price Formulation (P) Iterations	Time	Market Share Form. (COP) Iterations	Time	Dual Formulation (DCOP) Iterations	Time
2	256	7	0.02	15	0.00	19	0.28
4	256	7	0.03	15	0.01	20	0.34
8	256	10	0.12	17	0.01	22	0.50
16	256	16	0.83	24	0.03	28	0.99
32	256	21	4.22	26	0.07	25	1.68
64	256	24	17.48	25	0.16	26	4.28
128	256	27	75.69	29	0.64	25	13.30
256	256	24	265.10	29	1.90	24	55.77
512	256	25	1,307.78	34	4.38	33	451.61
1,024	256	27	5,181.45	34	9.03	27	2,346.63
2,048	256	29	22,818.30	36	20.66	33	19,560.10
4,096	256	38	123,123.50	38	49.09	-	-
256	2	19	1.63	26	0.04	28	28.72
256	4	16	2.13	26	0.04	30	30.42
256	8	19	4.38	28	0.05	28	28.80
256	16	23	14.95	27	0.07	29	30.84
256	32	21	31.09	30	0.11	27	31.10
256	64	21	59.52	29	0.21	38	49.54
256	128	20	114.26	29	0.56	26	41.40
256	256	24	265.10	29	1.90	24	55.77
256	512	36	964.53	36	5.26	28	122.72
256	1,024	50	2,411.04	36	8.83	31	338.77
256	2,048	60	5,549.19	45	22.95	35	1,226.33
256	4,096	69	13,003.43	53	59.36	43	5,918.61

Table 2.1: Number of iterations and solution time in seconds as a function of the number of products and constraints for the three problem formulations. (Averages over 10 randomly generated instances.)

and dual solution, and a parameter governing the criteria used to declare problems infeasible is relaxed to prevent premature termination (we set `inftol2=100`). These adjustments are necessary since convergence is sometimes very slow in the first few iterations, and again after five or six digits or accuracy have been achieved.

As is generally the case with interior point methods, LOQO terminates in a moderate number of iterations for all the instances, but there is significant variability in the time per iteration. Examining first the results for the price formulation (P), we observe that the total solution time increases rapidly with both the number of products and the number of constraints. With 4,096 products and 256 constraints, approximately 34 hours of computation time are needed. The solution time scales somewhat better with the number of constraints, but 3.6 hours are still needed with 4,096 constraints and only 256 products. We cannot theoretically guarantee convergence when solving (P) in general. Nevertheless, the optimal solution is eventually found in all the instances we considered, with the chosen parameters.

In contrast, the market share formulation we introduced, even for the largest instances we considered, is solved in about one minute, and is about 2,500 times faster than the price formulation in the most extreme case. We believe this difference may be due to the sparsity of the Hessian matrix of the objective function and the linearity of the constraint in (COP). In the price formulation (P), the Hessian of the Lagrangian is dense since each term of both the constraints and the objective depends nonlinearly on *all* the variables. The non-convexity of the constraints and objective in (P) may also be the reason for the slow performance, though we would expect this effect to increase the number of iterations rather than the time per iteration, which does not appear to be the case.

Finally, the dual formulation (DCOP) is observed to be much slower for a large number of products than for a large number of constraints. In fact, the specific dual formulation (DMNL) has a *single* constraint involving a summation over all n products with nonlinear terms. The number of terms in the sum increases with the number of products, but adding constraints in the primal problem (i.e., increasing m) only increases the number of dual variables λ_i, which are zero unless the constraint in question is active. We note that we were not able to obtain a solution for the (DCOP) instances with $n = 4,096$ and $m = 256$ because AMPL required an excessive amount of memory. This may be due to the fact that the actual number of primal constraints (corresponding to the number of dual variables) is not $m = 256$, but $m' \triangleq m + 2n = 8,448$ when including the converted price bounds. Then the Hessian of the dual constraint is a dense matrix with $(m')^2$ entries. It seems likely that a more efficient solution approach is possible for the dual problem, since most of the dual variables are zero at optimality (i.e., most of the constraints are inactive). Indeed, an efficient algorithm tailored to the special case of Proposition 2.2 is possible, but we see limited interest in pursuing this approach for the MNL demand model since the primal problem (COP) can be solved efficiently.

2.6.1 Approximation to Multiple Segment Demand Models

To illustrate the performance of the algorithm for solving the multiple-segment approximation, Table 2.2 shows the number of iterations and the running time needed to solve instances of varying size. For each evaluation of the gradient and Hessian, the equations in (2.7) are solved with Brent's method (see Brent [15]) and the derivatives in Proposition 2.9 are computed.

We observe that the solution times for (COP) are comparable to those for the single-segment instances, since the computational cost is dominated by the optimization algorithm rather than by the function evaluations. Indeed, the number of distinct segments only impact the time needed to evaluate the objective, and we observe that increasing the number of segments does not increase the solution time significantly.

In contrast, solving the price formulation (P) with the exact demands $d_i^{\mathrm{MIX}}(\mathbf{x})$ of (2.24) takes significantly longer with multiple segments. For a given problem size, doubling the number of segments more than doubles the solution time. The largest successfully solved instance took over three days (298,770 seconds) to solve, compared to only 15.77 seconds for the (COP) formulation with the approximate demand model.

For our experiments, the reference price $\mathbf{x^0}$ and the coefficients γ_ℓ in (2.27) were chosen such that the demand model approximation was exact at the uniform demand

Segments (k)	Products (n)	Constraints (m)	(P) Iterations	(P) Time (sec.)	(COP) with demand model approximation Iterations	Time (sec.)	Error	Infeasibility
2	256	256	23	601.94	58	4.73	0.57%	1.2E-03
	1024	256	26	10,812.00	46	11.94	2.22%	2.3E-03
	4096	256	19	165,090.00	35	46.11	0.56%	3.4E-04
	256	256	23	601.94	58	4.73	0.57%	1.2E-03
	256	1024	39	4,066.70	67	16.46	0.05%	2.4E-04
	256	4096	45	18,741.00	50	65.31	0.03%	1.3E-04
4	256	256	27	1,400.30	48	3.96	0.51%	1.0E-03
	1024	256	27	25,549.00	73	18.51	2.27%	1.9E-03
	4096	256	-	-	33	44.67	-	3.5E-04
	256	256	27	1,400.30	48	3.96	0.51%	1.0E-03
	256	1024	47	9,705.20	52	13.05	0.02%	1.0E-04
	256	4096	53	43,809.00	59	74.96	0.03%	1.5E-04
8	256	256	36	3,787.10	37	3.12	0.55%	1.4E-03
	1024	256	37	77,454.00	68	17.53	2.02%	2.4E-03
	4096	256	-	-	38	49.79	-	2.6E-04
	256	256	36	3,787.10	37	3.12	0.55%	1.4E-03
	256	1024	58	24,259.00	56	13.98	0.01%	6.5E-05
	256	4096	-	-	57	72.80	-	5.5E-05
16	256	256	48	11,818.00	37	3.21	0.55%	1.4E-03
	1024	256	68	298,770.00	59	15.77	1.86%	1.9E-03
	4096	256	-	-	42	53.79	-	2.9E-04
	256	256	48	11,818.00	37	3.21	0.55%	1.4E-03
	256	1024	82	80,675.00	55	13.90	0.02%	5.9E-05
	256	4096	-	-	86	103.45	-	2.4E-05

Table 2.2: Number of iterations and solution time in seconds to solve the exact, non-convex multi-segment pricing problem in terms of prices, as well as the convex multi-segment approximation in terms of market shares.

distribution, $\theta_i = \bar{d}_i(\mathbf{x^0}) = d_i^{\mathrm{MIX}}(\mathbf{x^0}) = 1/(n+1)$, for $i = 0, \ldots, n$. Thus the demand model approximation is inexact at the computed optimum. The second-last column in Table 2.2 shows the error in the objective value relative to the solution obtained by solving (P) with the exact demand model. The solution to (COP) overestimates the true maximum by at most 2.27% in our experiments, and generally by much less. It is somewhat unsurprising that the maximum of the approximation (COP) exceeds the true maximum, because this occurs whenever the approximation exceeds the true maximum of (P) at *any* feasible point. The rightmost column of Table 2.2 shows the maximum absolute constraint violation for each instance. The constraint violations are small despite the approximation, since the right hand side of the constraints (2.30) is on the order of $1/(n+1)$. For reference, the tolerance used for the solver is 10^{-6}. We remark that a different choice of the reference point $\mathbf{x^0}$ may improve the accuracy at the optimum. Although it is unclear how best to select this parameter of the approximation *a priori*, the accuracy of the approximation and the constraint violations at the computed optimal value can easily be checked empirically once the solution has been obtained.

2.7 Conclusions

We have developed an optimization framework for solving a large class of constrained pricing problems under the important class of attraction demand models. Our formulations incorporate a variety of constraints which naturally occur in numerous problems studied in the literature. They provide increased representation power for typical revenue management settings such as airline, hotel and other booking systems. Moreover, they capture problems such as product line pricing, or joint inventory and pricing problems where capacity constraints alone may not be sufficient.

We provided a condition on the demand model guaranteeing that our formulations are convex optimization problems with linear constraints. It is satisfied by MNL, MCI and linear attraction demand models, in particular. We further proved that the pricing problem can be solved in polynomial time under MNL demand models using interior point methods. Our computational experiments show that our new formulations can be solved orders of magnitude faster than naive formulations, using commercially available software. The efficiency of the solutions suggests our models may be effectively adapted for use in multi-period stochastic pricing problems, where they promise to increase modeling power at reasonable computational cost.

Furthermore, we proposed an approximation to the demand encountered when there are multiple overlapping market segments. Such scenarios are an active research topic in the closely related area of network revenue management. We provided a bound on the approximation error, and showed that the resulting pricing problems can be solved using standard nonlinear optimization algorithms despite the lack of a closed-form objective function. Our approximation represents a new way to approach pricing in the presence of multiple overlapping market segments, and provides an efficient way to solve certain instances.

Chapter 3

Pricing under the Nested Logit Demand Model

3.1 Introduction

In this chapter, we consider the problem of selecting prices $x_1, x_2, \ldots, x_n \in \mathbb{R}$ of n differentiated substitutable products offered to price sensitive customers whose purchasing behavior is modeled by a *nested logit* (NL) dicrete choice model. We aim to maximize the profit,

$$\max_{\mathbf{x} \in \mathcal{X}} \left\{ \Pi(\mathbf{x}) = \sum_{i=1}^{n} a_i (x_i - c_i) p_i(\mathbf{x}) \right\}, \tag{3.1}$$

where the constants $a_i > 0$ and $c_i \in \mathbb{R}$ represent a profit margin and marginal production cost specific to each product i. The function $p_i(\mathbf{x})$ denotes the demand for product i as a function of all the prices. We denote the fraction of customers opting not to make a purchase by $p_{n+1}(\mathbf{x}) = 1 - \sum_{i=1}^{n} p_i(\mathbf{x})$. Because the choice probabilities arising from the NL model are non-zero, the vector obtained by appending to no-purchase probability lies on the interior of the $(n+1)$-dimensional simplex:

$$\bar{\mathbf{p}} = \begin{bmatrix} \mathbf{p} \\ p_{n+1} \end{bmatrix} \in \Delta_{n+1} \triangleq \left\{ \mathbf{u} \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} u_i = 1, \text{and } u_i > 0, \forall i \right\}, \tag{3.2}$$

Because the last coordinate of $\bar{\mathbf{p}}$ is redundant, we also write $\mathbf{p} \in \Delta_{n+1}$ for simplicity.[1]

We assume that the non-empty feasible set of price vectors, $\mathcal{X} \subset \mathbb{R}^n$, is defined by constraints of the form

$$g_j(p(\mathbf{x})) \leq 0, \qquad j = 1, \ldots, J, \qquad \text{and} \qquad h_k(p(\mathbf{x})) = 0, \qquad k = 1, \ldots, K,$$

where each $g_j(\cdot)$ is convex and each $h_k(\cdot)$ is linear terms of the vector of *demands* $p(\mathbf{x})$ the seller will serve. For example, inventory constraints, production capacity limits and minimum sales targets are often *linear* contraints in terms of the demand. Unfortunately, the profit $\Pi(\mathbf{x})$ is generally not a concave (nor quasi-concave) function of $\mathbf{x}$, and $\mathcal{X}$ is generally not a convex set, even if we limit our attention to the special case of the *multinomial logit* (MNL) model. (See Chapter 2 and, in particular, Appendix A.2.)

We prefer to solve the equivalent, more natural, optimization problem over the vector of choice probabilities $\mathbf{p}$,

$$\max_{\mathbf{p} \in \mathcal{P}} \left\{ \Pi(\mathbf{p}) = \sum_{i=1}^{n} a_i(x_i - c_i)p_i \right\}, \tag{3.3}$$

where, by a slight abuse of notation, $\Pi(\mathbf{p})$ denotes the profit as a function of the market shares. Because we have assumed that $\mathcal{X}$ is defined by convex constraints in terms of $\mathbf{p}$, the feasible region $\mathcal{P} \subseteq \Delta_{n+1}$ is a nonempty convex set. We will show that the mapping between $\mathbf{x}$ and $\mathbf{p}$ is invertible. Specifically, under NL models, the function $\bar{p} : \mathbb{R}^n \to \Delta_{n+1}$ mapping $\mathbf{x}$ to $\bar{\mathbf{p}}$ is one-to-one and onto.

In Section 3.7 we derive a sufficient condition on the parameters of the NL model that guarantees strict concavity of $\Pi(\mathbf{p})$ over its entire domain. In particular, if this condition is satisfied, the optimal choice probabilities $\mathbf{p}^*$, and the corresponding prices $\mathbf{x}^*$, are unique. Concave maximization problems over convex feasible regions can be efficiently solved in practice with general-purpose algorithms, provided that the gradient and Hessian of the objective function can be computed efficiently.[2]

[1] Strictly speaking, $\mathbf{p} \in \{\mathbf{v} \in \mathbb{R}^n : \sum_{i=1}^n v_i < 1, \text{and } v_i > 0, \forall i\}$. The bijection with Δ_{n+1} is clear.
[2] One can safely optimize over the closure of Δ_{n+1} to avoid having to represent the strict positivity

More generally, $\Pi(\mathbf{p})$ may not be concave. In Section 3.8 we consider an important subclass of unconstrained ($\mathcal{X} = \mathbb{R}^n$ and $\mathcal{P} = \Delta_{n+1}$) profit maximization problems considered in the recent revenue management literature [51, 36, 20]. We propose an intuitive decomposition for this special case. Our algorithm converges to a stationary point of $\Pi(\mathbf{p})$, which is globally optimal if the sufficient condition is satisfied. Even if the condition is violated, our experience and the results in Chapter 5 of this thesis lead us to believe that the stationary point is in fact unique, and therefore a global maximum. Our method solves a set of convex optimization problems via simple line searches on each iteration, so no general-purpose solver is needed.

We present the NL model in Section 3.2, review related results in Section 3.3, and verify the invertibility of the NL demand model in Section 3.5. In Section 3.6 we derive the Jacobian matrix of the prices and state a useful property.

3.2 The Nested Logit (NL) Model

In the NL model, products are partitioned into M *nests*, each represented by a separate MNL model. An arriving customer first chooses a nest (with probabilities depending on the prices), and then chooses according the the MNL model in that nest. For example, a traveler booking a flight may first choose a destination airport, and then a flight time.

The *attraction* of each product is

$$y_i = \exp\{d_i - b_i x_i\}, \qquad i = 1, \ldots, n,$$

where the quality parameter $d_i \in \mathbb{R}$ represents the inherent desirability of product i, and $b_i > 0$ governs how sensitive customers are to changes in its price. For the example of an airline booking, a mid-day flight might be less popular, but may tend to draw more price-sensitive travelers. It would have smaller d_i and larger b_i than the flight with the earlier arrival time. The attraction of the $(n+1)^{th}$ choice representing

constraints. See the proof of Corollary 3.6.

the no-purchase alternative is set to $y_{n+1} = e^0 = 1$ without loss of generality, because all the y_i can be scaled without affecting the choice probabilities.[3].

The choice probabilities for nest $m \in \{1, 2, \dots, M\}$, and the conditional choice probabilities for each product $i \in \{1, 2, \dots, n+1\}$ once a nest has been chosen are, respectively,

$$Q_m \triangleq \frac{\left(\sum_{j=1}^{n+1} \alpha_{jm} y_j^{\mu_m}\right)^{\frac{1}{\mu_m}}}{G(\mathbf{y})}, \qquad \text{and} \qquad p_{i|m} \triangleq \frac{\alpha_{im} y_i^{\mu_m}}{\sum_{j=1}^{n+1} \alpha_{jm} y_j^{\mu_m}}.$$

The indicator variable $\alpha_{im} \in \{0, 1\}$ is equal to one if and only if product i is in nest m. Within each nest m, the scale parameter $\mu_m \geq 1$ determines how sharply the demand shifts from one product to the other as prices vary. If all the scale parameters are $\mu_m = 1$, it is as if there were only a single nest. (The NL reduces to the MNL model.) If the scale parameters are large, a small change in the attraction y_i can lead to a drastic change in the demand $p_{i|m}$. To see this, notice that the parameter μ_m scales the price sensitivity parameters b_i (as well as the quality parameters d_i) in the numerator of $p_{i|m}$,

$$y_i^{\mu_m} = \exp\left\{\mu_m(d_i - b_i x_i)\right\}.$$

The denominator in the expression for Q_m is the normalizing factor

$$G(\mathbf{y}) = \sum_{m=1}^{M} \left(\sum_{j=1}^{n+1} \alpha_{jm} y_j^{\mu_m}\right)^{\frac{1}{\mu_m}}. \tag{3.4}$$

We denote the unique nest containing product i by m_i, such that $m_i = m \Leftrightarrow \alpha_{im} = 1 \Leftrightarrow p_{i|m} > 0$. The demand for product i is

$$p_i = \sum_{m=1}^{M} p_{i|m} Q_m = p_{i|m_i} Q_{m_i}.$$

[3]This fact is easily verified for the NL model. It follows from *homogeneity* of the choice probabilities in terms of the vector containing all the y_i values. We further discuss this property in Chapter 4 of this thesis, when we consider the class of GEV models.

The nest probabilities in terms of $\bar{\mathbf{p}}$ are

$$Q_m = Q_m \left(\sum_{i=1}^{n+1} p_{i|m} \right) = Q_m \left(\sum_{i=1}^{n+1} \alpha_{im} p_{i|m} \right) = \sum_{i=1}^{n+1} \alpha_{im} p_{i|m} Q_m = \sum_{i=1}^{n+1} \alpha_{im} p_i. \quad (3.5)$$

For clarity, denote the nest containing the outside alternative by $m^* = m_{n+1}$. Unlike some authors, we allow the nest m^* to contain any number of products in addition to the no-purchase option $(n+1)$. Therefore our model is more general than those from the existing work on pricing under NL models.[4] It captures any NL model satisfying the classical definition from statistics, without imposing additional assumptions on the nesting structure. Clearly, the values of $y_1, \ldots, y_{n+1}$, the nest probabilities Q_m, and the choice probabilities p_i are all strictly positive.

The NL model belongs to class of *generalized extreme value* (GEV) discrete choice models. The choice probabilities above can also be defined in terms of the partial derivatives of the function $G(\mathbf{y})$ defined in (3.4), which is termed the *GEV generating function* for the NL model. We employ this definition in the proof of Lemma 3.1. Other GEV models include the *cross-nested logit* (CNL) discussed in Chapter 5, where the parameters α_{im} are allowed to take on fractional values, and the *network GEV* model proposed by Daly and Bierlaire [19] for more complex nesting structures.

We defer some of the formal proofs that apply to all GEV models to Chapter 4. Restricting our attention to NL models in this chapter allows for a simpler development and stronger results in this special case. In fact, the recent work on revenue management under customer choice only considers an even more restricted class of NL models.

3.3 Literature Review: Pricing under the NL Model

Li and Huh [51] showed that if the outside alternative is contained in a nest of its own, and if the price sensitivity parameters b_i are all equal for products within each nest, then the profit is concave in $\mathbf{p}$. While the price sensitivity of a given consumer may

[4]See the literature review in Section 3.3. We present novel results under the common but restrictive assumption that next m^* contains *only* the no-purchase option in Section 3.8.

be constant across products, the price sensitivity usually varies across a population. Meanwhile, all consumers are free to purchase products in any nest. Therefore, when fitting NL models to data, it is natural to allow the parameters $b_1, \ldots, b_n$ to vary across products without regard for the nesting structure.[5] Their values can vary significantly in practice.

Gallego and Wang [36] show that concavity of the profit persists if the price sensitivity parameters satisfy

$$\max_{\{i,j|m_i=m_j=m\}} \frac{b_i}{b_j} \leq \frac{\mu_m}{\mu_m - 1}, \tag{3.6}$$

for the products within each nest m. Therefore, when $\mu_m \geq 1$ is close to 1, the sensitivity parameter of products in nest m may differ significantly. Of course, as all the μ_m approach 1, the NL reduces to the MNL model, for which we have already established concavity under any b_i values in Chapter 2. Even for moderate values of the μ_m, say $\mu_m = 5$, the condition (3.6) restricts the ratio between the b_i to be less than $\frac{5}{4}$. Therefore this result represents only a slight generalization for practical problem instances.

In constrast to both results, we impose no restriction on the nesting structure and show that the profit remains concave when the ratios b_i/b_j are less than 2 *regardless* of the scale parameters $\mu_1, \ldots, \mu_M$. Therefore, we allow for drastic changes in demand as the prices vary, and we represent substitution behavior that is substantially different than under MNL models, by allowing for large values of the μ_m parameters.

Li and Huh [51] also suggest a bisection root-finding algorithm to find the (unique and optimal, under their assumptions) stationary point of the profit. Gallego and Wang [36] decompose the pricing problem into individual non-concave pricing problems for each nest. They observe that the subproblems may have multiple local

[5]This observation is even more applicable to MNL models because they do not represent any nesting structure at all. It is somewhat less applicable to CNL models, which have more parameters and explicitly allow for the products to belong to multiple nests. For MMNL models, it is indeed customary to chose a fixed price-sensitivity for *all* products, for each component of the mixture (see Chapter 5). However, because NL models have many fewer parameters than MMNL models, assuming a fixed price-sensitivity within each nest is restrictive.

maxima when their concavity condition is violated. For the model they consider, we use a different decomposition in terms of the conditional choice probabilities $p_{i|m}$. Our method computes the unique solutions of concave maximization problems via line searches at each step. When appropriately expressed, the profit is *biconcave* for *any* price sensitivity parameters b_i and *any* scale parameters $\mu_m \geq 1$. We apply results from the survey of Gorski et al. [39] on the optimization of such functions. Because the sub-problems are always concave, our algorithm is simpler, more reliable and more efficient in practice than that of Gallego and Wang [36]. It always converges to a stationary point of the profit. This is the unique global maximum when the ratios b_i/b_j are less than 2. In practice, our algorithm for this special case appears to converge to a global maximum for any price sensitivity parameters b_i (under the restrictive assumption on the nesting structure).

There is also some work on *assortment optimization* under NL models, where the seller selects the most profitable subset to offer customers from a (usually large) set of potential products with *fixed* prices. Davis et al. [20] show that this problem is polynomially solvable when the outside alternative is in its own nest, but becomes NP-hard under a variant of the NL with a separate no-purchase option within each nest.[6] In principle, one could create multiple copies of each product in the pricing problem to represent discretized prices, and then solve the assortment optimization problem with constraints to ensure that a single price is chosen for each product. Gallego and Topaloglu [35] show that the assortment optimization problem remains tractable under per-nest cardinality and space constraints (via an approximation algorithm in the latter case). However, they suggest a hybrid pricing-assortment problem rather than pursue a discretization of prices.

[6]The general NL model with the no-purchase option in any single nest seems somewhat less relevant in this setting, but it is not clear to us whether assortment optimization remains polynomially solvable.

3.4 Outline of the Proofs

In Section 3.5, Lemma 3.1 shows that the NL demand function is invertible. We also show how the demand model can be adjusted to take into account varying profit margins and production costs for different products. The profit can be expressed as $\Pi(\mathbf{p}) = \mathbf{p}^\top \mathbf{z}$, where $\mathbf{z}$ is the vector of adjusted prices.

In Section 3.6, we characterize the Jacobian matrix of the prices $\mathbf{z}$ with respect to the demands $\mathbf{p}$ as a *negative inverse M-matrix* (Proposition 3.3). We then obtain a closed-form expression for the Jacobian (Lemma 3.4).

Section 3.7 is devoted to showing concavity of the profit $\Pi(\mathbf{p})$. Having computed the Jacobian of the prices, we obtain a closed-form expression for the Hessian matrix of $\Pi(\mathbf{p})$ (Proposition B.1, Lemma B.2 and Lemma B.3 in the appendix). In Theorem 3.5 and Corollary 3.6, we show that the Hessian is negative definite and the profit is strictly concave when the ratio between the price sensitivity parameters is bounded by 2. Under this condition, the constrained pricing problem can be solved to optimality by using a general purpose nonlinear solver, along with our closed-form expressions for the gradient and Hessian of the profit.

In Section 3.8, we consider the special case where the no-purchase alternative is in a nest by itself (Assumption 3.7). We show that stating the pricing problem in terms of the *conditional choice probabilities* within each nest results in a *biconcave* profit function $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_\mathbf{m})$ (Lemma 3.9). Lemmas 3.11 and 3.12 define the steps of the *alternating convex search* (ACS) algorithm to maximize such functions. Theorem 3.13 proves that the algorithm converges to a stationary point, and Corollary 3.14 states that this stationary point is the global maximum, when the ratio between the price sensitivity parameters is bounded by 2. We expect this algorithm to be simpler to implement and more effective in practice than using a general purpose nonlinear solver, when it is applicable, because it quickly solves concave maximization subproblems in nearly closed-form at each iteration.

3.5 Inverting the NL Demand Function

We begin by showing that the demand function under NL models is in fact invertible. The proof is substantially more complicated than for the GA models of Chapter 2, and has not appeared in the literature, to our knowledge.[7] We will extend the result to the entire family of GEV models in Chapter 4, but even Lemma 3.1 below relies on the GEV generating function $G(\mathbf{y})$ for NL models defined in (3.4).

What distinguishes NL models from CNL and more general GEV models is that we can recover a closed-form expression for each y_i in terms of the choice probabilities $\mathbf{p}$. Because y_i is a strictly decreasing function of the price x_i for each product i, we can also uniquely recover the prices for any given choice probabilities in closed form.

Lemma 3.1. *Under a NL model, for $i = 1, \ldots, n+1$,*

$$y_i = \frac{p_{i|m_i}^{\frac{1}{\mu_{m_i}}} Q_{m_i}}{p_{n+1|m^*}^{\frac{1}{\mu_{m^*}}} Q_{m^*}} = \frac{p_i^{\frac{1}{\mu_{m_i}}} Q_{m_i}^{1-\frac{1}{\mu_{m_i}}}}{p_{n+1}^{\frac{1}{\mu_{m^*}}} Q_{m^*}^{1-\frac{1}{\mu_{m^*}}}}.$$

Therefore the demand function $p : \mathbb{R}^n \to \Delta_{n+1}$ is invertible (one-to-one and onto). Moreover, both $p(\cdot)$ and $p^{-1}(\cdot)$ are differentiable.

Proof. See Appendix B.1. □

Recall that $m^* = m_{n+1}$ is simply shorthand for the nest containing the no-purchase option. The two expressions are equivalent because $p_i = p_{i|m_i} Q_{m_i}, \forall i$, under the NL model. The first one allows us to decompose the profit maximization problem in terms of the conditional choice probabilities. The second one is useful when directly optimizing over $\mathbf{p}$. Both can be used to obtain closed-form expressions for the price x_i. But first, we adjust the parameters of the demand model in order to lighten the notation.

[7] Gallego and Wang [36] do re-express the prices in terms of the demands, but they rely on their more restrictive assumption on the nesting structure and their condition (3.6). They make no claims about invertibility in general.

3.5.1 Accounting for Profit Margins and Production Costs

We define the adjusted price vector $\mathbf{z} \in \mathbb{R}^n$ by

$$z_i \triangleq a_i(x_i - c_i), \qquad i = 1, \ldots, n.$$

The profit simplifies to

$$\Pi(\mathbf{p}) = \mathbf{p}^\top \mathbf{z} = \sum_{i=1}^{n} p_i z_i.$$

If, for every product i, $a_i = 1$ and $c_i = 0$, then $\mathbf{z} = \mathbf{x}$. In general, the NL model can be equivalently expressed in terms of the modified price vector $\mathbf{z}$. The choice probabilities depend on $\mathbf{x}$ only through the variables, for $i = 1, \ldots, n$,

$$y_i = e^{d_i - b_i x_i} = \exp\left\{d_i - b_i\left(\frac{z_i}{a_i} + c_i\right)\right\} = \exp\left\{(d_i - b_i c_i) - \frac{b_i}{a_i} z_i\right\} = e^{d_i' - b_i' z_i}, \quad (3.7)$$

where we have substituted $x_i = \frac{1}{a_i} z_i + c_i$, and defined

$$d_i' \triangleq d_i - b_i c_i, \qquad \text{and} \qquad b_i' \triangleq \frac{b_i}{a_i}.$$

Our results do not depend on the cost parameters c_i, but some care will be required when the profit margins differ from 1. It is convenient to define the matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ with the values of b_i' on the diagonal. That is, we let

$$\mathbf{B} \triangleq \operatorname{diag}\left(\left[\frac{b_1}{a_1}, \frac{b_2}{a_2}, \ldots, \frac{b_n}{a_n}\right]\right) = \begin{bmatrix} \frac{b_1}{a_1} & 0 & \cdots & 0 \\ 0 & \frac{b_2}{a_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{b_n}{a_n} \end{bmatrix}$$

denote the matrix with the b_i' on the diagonal.

In the case where the profit margins a_i and the price sensitivity parameters b_i are the same for all the products $i = 1, \ldots, n$, the matrix $\mathbf{B}$ is simply a multiple of the identity matrix. In fact, since the units of the prices $\mathbf{x}$ are arbitrary, we may assume without loss of generality that $\mathbf{B} = \mathbf{I}$. Our later results will depend on how much the actual problem parameters differ from this "ideal" situation. We remark that it may be realistic to require equal profit margins and price sensitivity across products in some settings.

3.5.2 Recovering the Prices

From Equation (3.7) and Lemma 3.1, the adjusted price of product $i \leq n$ is

$$
\begin{aligned}
z_i &= \frac{1}{b'_i}\left(d'_i - \log y_i\right) \\
&= \frac{1}{b'_i}\left(d'_i - \frac{1}{\mu_{m_i}} \log p_{i|m_i} + \frac{1}{\mu_{m^*}} \log p_{n+1|m^*} - \log \frac{Q_{m_i}}{Q_{m^*}}\right) \qquad (3.8) \\
&= \frac{1}{b'_i}\left(d'_i - \frac{1}{\mu_{m_i}} \log p_i - \left(1 - \frac{1}{\mu_{m_i}}\right) \log Q_{m_i}\right. \\
&\qquad \left. + \frac{1}{\mu_{m^*}} \log p_{n+1} + \left(1 - \frac{1}{\mu_{m^*}}\right) \log Q_{m^*}\right). \qquad (3.9)
\end{aligned}
$$

The original price x_i is obtained straightforwardly from the definition of z_i.

3.6 The Jacobian of the Prices

We define the Jacobian[8] matrix of (adjusted) prices with respect to $\mathbf{p}$,

$$
\mathbf{J_z} = \left[\frac{\partial z_j}{\partial p_i}\right]_{ij} = \begin{bmatrix} \frac{\partial z_1}{\partial p_1} & \frac{\partial z_2}{\partial p_1} & \cdots & \frac{\partial z_n}{\partial p_1} \\ \frac{\partial z_1}{\partial p_2} & \frac{\partial z_2}{\partial p_2} & \cdots & \frac{\partial z_n}{\partial p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial p_n} & \frac{\partial z_2}{\partial p_n} & \cdots & \frac{\partial z_n}{\partial p_n} \end{bmatrix}
$$

[8] Throughout this thesis, we use the *denominator layout* notation convention, also known as the *Hessian formulation* or simply *gradient* notation. Gradients are expressed as column vectors, and each column of a Jacobian matrix is a gradient. For example, the first column of $\mathbf{J_z}$ is the gradient $\frac{\partial z_1}{\partial \mathbf{p}}$ of the scalar z_1 with respect to the vector $\mathbf{p}$. It is layed out according to the denominator, which is the column vector $\mathbf{p}$ in this case.

Applying the product rule of vector calculus, the gradient of the profit $\Pi(\mathbf{p}) = \mathbf{p}^\top \mathbf{z}$ is the column vector

$$\frac{\partial \Pi}{\partial \mathbf{p}} = \left[\frac{\partial \Pi}{\partial p_i}\right]_i = \begin{bmatrix} \frac{\partial \Pi}{\partial p_1} \\ \frac{\partial \Pi}{\partial p_2} \\ \vdots \\ \frac{\partial \Pi}{\partial p_n} \end{bmatrix} = \mathbf{z} + \mathbf{J_z}\mathbf{p}.$$

We will use an explicit expression for $\mathbf{J_z}$ and a property of its inverse $\mathbf{J_z}^{-1}$ to show that the profit function is often concave.

3.6.1 Substitutable products and M-matrices

The Jacobian of the *demands* with respect to the prices is $\mathbf{J_z}^{-1}$ by the inverse function theorem of multivariate calculus. Without computing the derivatives explicitly[9], it is easy to see that the matrix must have sign pattern

$$\mathbf{J_z}^{-1} = \left[\frac{\partial p_j}{\partial z_i}\right] = \begin{bmatrix} \frac{\partial p_1}{\partial z_1} & \frac{\partial p_2}{\partial z_1} & \frac{\partial p_3}{\partial z_1} & \cdots & \frac{\partial p_n}{\partial z_1} \\ \frac{\partial p_1}{\partial z_2} & \frac{\partial p_2}{\partial z_2} & \frac{\partial p_3}{\partial z_2} & \cdots & \frac{\partial p_n}{\partial z_2} \\ \frac{\partial p_1}{\partial z_3} & \frac{\partial p_2}{\partial z_3} & \frac{\partial p_3}{\partial z_3} & \cdots & \frac{\partial p_n}{\partial z_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_1}{\partial z_n} & \frac{\partial p_2}{\partial z_n} & \frac{\partial p_3}{\partial z_n} & \cdots & \frac{\partial p_n}{\partial z_n} \end{bmatrix} = \begin{bmatrix} - & + & + & \cdots & + \\ + & - & + & \cdots & + \\ + & + & - & \cdots & + \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ + & + & + & \cdots & - \end{bmatrix}.$$

That is, for each row (product) i, increasing the price z_i leads to a decrease in p_i and an increase in $p_j, j \neq i$. The no-purchase probabilitiy, p_{n+1}, also increases. On the other hand, the sum of the choice probabilities remains one, and therefore the sum of their changes is zero. Then each row of $\mathbf{J_z}^{-1}$ must sum to

$$\sum_{j=1}^{n} \frac{\partial p_j}{\partial z_i} < 0,$$

because we have omitted the positive term $\partial p_{n+1}/\partial z_i$. But then, $\mathbf{J_z}^{-1}$ must be strictly diagonally dominant. This in turn implies that every principal submatrix is also

[9] We do so in Section 4.7 of Chapter 4 of this thesis, for general GEV models.

diagonally dominant, and that every principal minor is negative. A matrix with this last property and the sign pattern of $\mathbf{J}_\mathbf{z}^{-1}$ is the negation of an *M-matrix.*

Definition 3.2. An *M-matrix* is a square nonsingular matrix with non-positive off-diagonal entries and all principal minors positive. In particular, a strictly diagonally-dominant matrix with positive diagonal entries and non-positive off-diagonal entries is an M-matrix. We also define *inverse M-*, *negative M-* and *negative inverse M-*matrices as the classes of inverses, negations and negated inverses of M-matrices.

We will use some well-known facts about M-matrices and their inverses:

- The transpose of an M-matrix is also an M matrix.
- Scaling an M-matrix by a positive diagonal matrix yields an M-matrix.
- An inverse M-matrix is positive-definite.
- All the entries of an inverse M-matrix are non-negative.

We refer the interested reader to the recent survey by Johnson and Smith [43] for proofs. The following proposition characterizing $\mathbf{J}_\mathbf{z}$ will suffice to prove our main result in Theorem 3.5.

Proposition 3.3. *The Jacobian $\mathbf{J}_\mathbf{z}$ is a negative inverse M-matrix. Consequently, $\mathbf{J}_\mathbf{z}$ is negative-definite and has non-positive entries. The same holds for the negative inverse M-matrices $\mathbf{J}_\mathbf{z}\mathbf{E}$ and $\mathbf{E}\mathbf{J}_\mathbf{z}^\top$, where $\mathbf{E}$ is any strictly positive diagonal matrix.*

Proof. A formal proof that $\mathbf{J}_\mathbf{z}^{-1}$ is the negation of a diagonally-dominant M-matrix under the entire class of GEV models is given in Proposition 4.17. The class of inverse M-matrices is closed under transposition and positive diagonal scaling [43, Theorems 1.2.2 and 1.2.3]. It is well-known that inverse M-matrices are positive definite and have non-negative entries [43, Theorem 1.1]. □

3.6.2 Deriving the Jacobian

The second expression in Lemma 3.1 is a simple function of p_i, p_{n+1}, Q_{m_i} and Q_{m^*}. These are in turn all affine functions of the vector $\mathbf{p}$ by their definitions. Then the partial derivatives $\partial y_j/\partial p_i$ can be obtained using the chain rule from vector calculus. Because z_j is differentiable with respect to y_j, the matrix $\mathbf{J_z}$ can be obtained by again applying the chain rule. We obtain a simple expression for $\mathbf{J_z}$, but we must first define the intermediate Jacobian matrices.

Recall that the vector $\bar{\mathbf{p}}$ defined in (3.2) is simply the vector $\mathbf{p}$ with the no-purchase probability p_{n+1} in the last coordinate. That is, writing $\mathbf{e}_{n+1}$ for the vector of all zeros except for a 1 in the last coordinate, and writing $\mathbf{e}$ for the vector of all ones,

$$\bar{\mathbf{p}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{e}^\top \end{bmatrix} \mathbf{p} = \mathbf{e}_{n+1} + \mathbf{M}^\top \mathbf{p}$$

where $\mathbf{M}$ the Jacobian matrix of $\bar{\mathbf{p}}$ with respect to $\mathbf{p}$, and can be written as

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{I} & -\mathbf{e} \end{bmatrix} \in \{-1, 0, 1\}^{n \times (n+1)},$$

We define the vector $\mathbf{Q_m} \triangleq \begin{bmatrix} Q_1, \dots, Q_M \end{bmatrix}^\top$ of nest probabilities. Summing up the probabilities within each nest, and from (3.5), this vector may be written as

$$\mathbf{Q_m} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n+1} \alpha_{i1} p_i \\ \sum_{i=1}^{n+1} \alpha_{i1} p_i \\ \vdots \\ \sum_{i=1}^{n+1} \alpha_{iM} p_i \end{bmatrix} = \mathbf{N}^\top \bar{\mathbf{p}},$$

where the incidence matrix of products to nests,

$$\mathbf{N} \triangleq \begin{bmatrix} \alpha_{im} \end{bmatrix}_{im} \in \{0, 1\}^{n \times M},$$

is also the Jacobian of the vector of nest probabilities $\mathbf{Q_m}$ with respect to $\bar{\mathbf{p}}$.

As an example, if there are $n = 4$ products in $M = 2$ nests, with the $(n+1)^{\text{th}}$

no-purchase option belonging to the second nest, we may have

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \in \mathbb{R}^{4\times 5} \qquad \text{and} \qquad \mathbf{N} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{5\times 2}.$$

Lemma 3.4. *Let $\bar{\mathbf{P}} \in \mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{Q} \in \mathbb{R}^{m\times m}$ be the positive diagonal matrices with the probabilities $\bar{\mathbf{p}}$ and $Q_1, \dots, Q_m$ on their respective diagonals. Define the diagonal matrices of parameters $\bar{\mathbf{U}} \in \mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{V} \in \mathbb{R}^{M\times M}$ with entries*

$$\bar{U}_{ii} = \frac{1}{\mu_{m_i}} > 0 \qquad \textit{and} \qquad V_{mm} = \left(1 - \frac{1}{\mu_m}\right) \geq 0.$$

Then the partial derivatives of $\mathbf{z}$ with respect to $\mathbf{p}$ are

$$\mathbf{J_z} = \left[\frac{\partial z_j}{\partial p_i}\right]_{ij} = -\mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} + \mathbf{N}\mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^{\top}\right)\mathbf{M}^{\top}\mathbf{B}^{-1}.$$

Proof. See Appendix B.2. □

The inner factors $\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}$ and $\mathbf{V}\mathbf{Q}^{-1}$ are non-negative diagonal matrices. If all the (adjusted) price sensitivities are equal to one, $\mathbf{B} = \mathbf{I}$, and it is easy to see that the Jacobian matrix is symmetric negative definite. By Proposition 3.3, it remains negative-definite regardless of the profit margins and price sensitivity parameters. That is, the Jacobian matrix remains negative-definite despite the arbitrary positive diagonal scaling by $\mathbf{B}^{-1}$ because it is a negative inverse M-matrix.

3.7 Concavity of the Profit Function

Differentiating the gradient of $\Pi(\mathbf{p})$ again, we obtain an expression for the Hessian matrix of the profit. The derivation involves several steps, and is relegated to Appendix B.3.4

It is convenient to extend the matrix of price-sensitivity parameters with an $(n+1)^{\text{th}}$ entry corresponding to the no-purchase alternative. We will primarily be dealing with the matrix $\mathbf{B}^{-1}$ that appears in the expression for $\mathbf{J_z}$. We let

$$\bar{\mathbf{B}}^{-1} \triangleq \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & b_{n+1}^{-1} \end{bmatrix},$$

with the *negative* last entry set to

$$b_{n+1}^{-1} \triangleq -\frac{1}{p_{n+1}} \sum_{i=1}^{n} \frac{1}{b_i'} p_i = \frac{-\mathbf{e}^\top \mathbf{B}^{-1} \mathbf{p}}{p_{n+1}}.$$

This can be seen as a type of average of the inverse price-sensitivity parameters, though it is scaled by $-p_{n+1}^{-1}$. Note that b_{n+1} changes with $\mathbf{p}$ unlike the true price sensitivity parameters $b_1, \ldots, b_n$. We can now write the Hessian of $\Pi(\mathbf{p})$, and provide a sufficent condition ensuring that it is negative definite.

Theorem 3.5. *Under NL models, the Hessian of* $\Pi(\mathbf{p})$ *is*

$$\mathbf{H} = \mathbf{J_z} + \mathbf{J_z}^\top - \tilde{\mathbf{J}}_\mathbf{z}$$

where

$$\tilde{\mathbf{J}}_\mathbf{z} = -\mathbf{M} \left(\bar{\mathbf{U}} \bar{\mathbf{P}}^{-1} \bar{\mathbf{B}}^{-1} + \mathbf{N} \mathbf{Q}^{-1} \mathbf{W} \mathbf{N}^\top \right) \mathbf{M}^\top.$$

and the matrix $\mathbf{W}$ *is related to* $\mathbf{V}$ *by*

$$\mathbf{W} = \mathbf{V} \mathbf{N}^\top \bar{\mathbf{B}}^{-1} \bar{\mathbf{P}} \mathbf{N} \mathbf{Q}^{-1}.$$

If $\frac{1}{2} \leq b_1', b_2', \ldots, b_n' \leq 1$, *then* $\Pi(\mathbf{p})$ *is strictly concave.*

Proof. See Appendix B.3.4. $\square$

The symmetric matrix $\tilde{\mathbf{J}}_\mathbf{z}$ is similar to the negative-definite matrix $\mathbf{J_z}$, except for the presence of $\bar{\mathbf{B}}^{-1}$ with its negative last entry. If $\mathbf{B}^{-1}$ were absent, the two terms in the brackets would reduce to those of $\mathbf{J_z}$ because $\mathbf{N}^\top \bar{\mathbf{P}} \mathbf{N} = \mathbf{Q}$. Under the stated condition, we have that $\frac{1}{2}\mathbf{I} \leq \mathbf{B} \leq \mathbf{I}$ and $\tilde{\mathbf{J}}_{\mathbf{z}}$ is appropriately bounded by $\mathbf{J}_{\mathbf{z}}$ to ensure that the sum $\mathbf{H}$ remains negative-definite.

More generally, if the ratio between the adjusted price sensitivity parameters is at most two, then the maximum of the pricing problem is unique.

Corollary 3.6. *If the price sensitivity parameters and the profit margins satisfy*

$$\max_{1 \leq i,j \leq n} \frac{a_i b_j}{b_i a_j} < 2,$$

then the solution $\mathbf{p}^* = p(\mathbf{x}^*)$ *of problem* (3.3)*, and the corresponding solution* $\mathbf{x}^*$ *of problem* (3.1) *exist and are unique.*

Moreover, if the problem is unconstrained, that is, if $\mathcal{P} = \Delta_{n+1}$*, then* $\mathbf{p}^*$ *is the unique stationary point of the profit.*

Proof. We may consider the equivalent problem with the scaled price vector $\gamma\mathbf{z}$, and the adjusted price sensitivity parameters $\gamma^{-1}b_i'$, for the scalar $\gamma > 0$ such that the assumption of Theorem 3.5 is satisfied. This has no impact on the choice model because the prices affect it only through the y_i variables. The profit $\mathbf{p}^\top\mathbf{z}$ is simply scaled by γ along with the prices. A strictly concave function has at most one local maximum, which is also global.

Showing that the maximum exists is slightly complicated because the domain of $\Pi(\mathbf{p})$ is open, and the price z_i may become infinite if the choice probability p_i goes to zero. Suppose first that there are no constraints and that $\mathcal{P} = \Delta_{n+1}$. Consider the continuous extension of $\Pi(\mathbf{p})$ over the closure of Δ_{n+1}. Then a maximizer of the continuous function exists over the closed and bounded domain. We need only show that the maximum does not lie on the boundary.

For any product $i = 1, \ldots n$, as $p_i \to 0$, we also have that $y_i \to 0$ by the second equality of Lemma 3.1. But then $z_i \to \infty$ from equation (3.9), while the other prices remain finite. Thus once the price becomes sufficiently large, it will be profitable to shift some demand from other products to product i. But this precisely implies that $p_i > 0$. Since this holds for all products, none of the choice probabilities can be zero at optimality.

In the constrained case, $\mathcal{P}$ is non-empty and convex by assumption. Because the constraints in terms of $\mathbf{p}$ define a closed set, taking the closure of $\mathcal{P}$ only adds points from the boundary of the simplex. The argument therefore carries through, except that the optimal solution may no longer be a stationary point of the unconstrained profit if any of the constraints in terms of $\mathbf{p}$ are active at optimality. $\square$

In practice, we note that the solution of the optimal solution of the pricing problem appears to remain unique even when the condition of Corollary 3.6 is substantially violated. This is to be expected, because we have used the negative M-matrix property to show that the Hessian is negative-definite. This a sufficient but certainly not necessary condition. Moreover, concavity is also a sufficient but not necessary condition for the uniqueness of local maxima. It remains an open question to determine exactly under what circumstances the profit $\Pi(\mathbf{p})$ is concave, and under what circumstatnces it has a unique local maximizer.

3.8 A Separate Nest for the Outside Alternative

In this section we develop an intuitive decomposition of the NL pricing problem for the special case captured by the following assumption (without constraining the value of the a_i, b_i and μ_i parameters). We refer to the existing literature for its practical justification [51, 36, 20, 35].

Assumption 3.7. The no-purchase option with index $i = (n+1)$ is the unique choice in nest $m^* = m_{n+1} = M$. That is, for any product $i = 1, \ldots, n$, the indicator variable $\alpha_{im^*} = 0$ is zero and $m_i \neq m^*$.

This assumption could be relaxed somewhat: if the scale parameter for the nest m^* is minimal, such that $\mu_m \geq \mu_{m^*}, \forall m$, then the two subproblems (3.13) and (3.14) described below remain concave. However, the objective of (3.14) is no longer separable over nests, and the development would be more complicated.

3.8.1 The Profit as a Biconcave Function

Under Assumption 3.7, we have that $p_{n+1|m^*} = 1$. Then the price of product i in Equation (3.8) of Section 3.5.2 simplifies to

$$z_i = \frac{1}{b_i'}\left(d_i' - \frac{1}{\mu_{m_i}}\log p_{i|m_i} - \log\frac{Q_{m_i}}{Q_{m^*}}\right).$$

The profit in terms of the vector of conditional choice probabilities $\mathbf{p}_{|\mathbf{m}} \in \mathbb{R}^n$ and the vector of nest choice probabilities $\mathbf{Q_m} = \left[Q_1, Q_2, \ldots, Q_M\right]^\mathsf{T}$ is then

$$\begin{aligned}\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m}) &= \sum_{i=1}^{n} p_i z_i = \sum_{i=1}^{n} p_{i|m_i} Q_{m_i} z_i = \sum_{m=1}^{M} Q_m \sum_{i:m_i=m} p_{i|m_i} z_i \\ &= \sum_{m=1}^{M} Q_m \sum_{i:m_i=m} \frac{1}{b_i'} p_{i|m}\left(d_i' - \frac{1}{\mu_m}\log p_{i|m} - \log\frac{Q_m}{Q_{m^*}}\right).\end{aligned}$$

The profit is a *biconcave* function:

Definition 3.8 (Adapted from [39]). A function $f : S \times T \to \mathbb{R}$ on the cartesian product of convex sets S and T is called *biconcave* if $f_s(\cdot) \triangleq f(s, \cdot) : T \to \mathbb{R}$ and $f_t(\cdot) \triangleq f(\cdot, t) : S \to \mathbb{R}$ are both concave functions on their domain.

The concept can be generalized to functions defined on *biconvex sets*, which need not be cartesian products. The original vector pair $(\mathbf{p}, \mathbf{q})$ belongs to such a set.[10] We prefer the expression of $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m})$ because the theoretical results for optimization over general biconvex sets are weaker than for optimization over cartesian products. We verify our claim:

Lemma 3.9. *Under Assumption 3.7, the profit*

$$\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m}) : (S \times T) \to \mathbb{R}$$

[10] Fixing the value of $\mathbf{q} = \mathbf{q}^*$ is equivalent to the convex constraint $\mathbf{N}^\mathsf{T}\bar{\mathbf{p}} = \mathbf{q}^*$. Fixing the value of $\mathbf{p}$ constrains $\mathbf{q}$ to a single point. Therefore the feasible region remains convex when one of the two quantities is fixed. This corresponds to the definition of a biconvex set.

is a biconcave function on the cartesian product of the convex sets

$$S = \Delta_{n_1} \times \Delta_{n_2} \times \cdots \times \Delta_{n_M} \qquad \textit{and} \qquad T = \Delta_M,$$

where $n_m = |\{i : m_i = m\}|$ *is the number of products in nest* m.

Proof. Fixing either set of variables yields an optimization problem with terms of the form

$$f(x, y) = x \log \frac{x}{y}.$$

The gradient and Hessian matrix are

$$\nabla f(x,y) = \begin{bmatrix} 1 + \log \frac{x}{y} \\ -\frac{x}{y} \end{bmatrix} \qquad \text{and} \qquad \nabla^2 f(x,y) = \begin{bmatrix} \frac{1}{x} & \frac{-1}{y} \\ \frac{-1}{y} & \frac{x}{y^2} \end{bmatrix}. \tag{3.10}$$

For $x, y > 0$, the Hessian is a symmetric matrix with positive trace and a determinant of zero. Therefore, one eigenvalue is zero and the other is real and positive. Then the Hessian is positive-semidefinite, and $f(x,y)$ is convex on the strictly positive orthant. Its negation, $-f(x,y)$ is concave.

We may rearrange the terms of the profit as

$$\begin{aligned} \Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_\mathbf{m}) &= \sum_m Q_m \left(\sum_{i:m_i=m} \frac{1}{b_i'} p_{i|m} \left(d_i' - \log \frac{Q_m}{Q_{m^*}} \right) - \frac{1}{\mu_m} \sum_{i:m_i=m} \frac{1}{b_i'} p_{i|m} \log p_{i|m} \right) \\ &= \sum_m \left(\sum_{i:m_i=m} p_{i|m} \delta_{im} - \sum_{i:m_i=m} \frac{1}{\beta_{im}} p_{i|m} \log p_{i|m} \right), \end{aligned} \tag{3.11}$$

with appropriately defined variables δ_{im} and β_{im}. If the value of $\mathbf{Q}_\mathbf{m}$ is held fixed, $\Pi(\cdot, \mathbf{Q}_\mathbf{m})$ is a sum of linear terms and terms of the form $-f(p_{i|m_i}, 1)$. Therefore it is concave. On the other hand, we can also rearrange the terms of the profit as

$$\begin{aligned}\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m}) &= \sum_m Q_m \sum_{i:m_i=m} \frac{1}{b'_i} p_{i|m} \left(d'_i - \frac{1}{\mu_m} \log p_{i|m} \right) \\ &\qquad - \sum_m \left(\sum_{i:m_i=m} \frac{1}{b'_i} p_{i|m} \right) Q_m \log \frac{Q_m}{Q_{m^*}} \\ &= \sum_m Q_m \delta_m - \sum_m \frac{1}{\beta_m} Q_m \log \frac{Q_m}{Q_{m^*}}, \end{aligned} \tag{3.12}$$

with appropriately defined variables δ_m and β_m. If the value of $\mathbf{p}_{|\mathbf{m}}$ is held constant, $\Pi(\mathbf{p}_{|\mathbf{m}}, \cdot)$ is a sum of linear terms and terms of the form $-f(Q_m, Q_{m^*})$. Therefore it is concave. □

3.8.2 The Alternating Convex Search Algorithm (ACS)

Because $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m})$ is also differentiable, its stationary points correspond exactly with its *partial optima* [39, Corollary 4.3], where both sub-functions are maximized.

Definition 3.10. Let $f : S \times T \to \mathbb{R}$ be a given function. Then for $\mathbf{s}^* \in S$, $\mathbf{t}^* \in T$, $(\mathbf{s}^*, \mathbf{t}^*)$ is called a *partial optimum* of f on $S \times T$ if

$$f(\mathbf{s}^*, \mathbf{t}^*) \geq f(\mathbf{s}, \mathbf{t}^*), \forall \mathbf{s} \in S \qquad \text{and} \qquad f(\mathbf{s}^*, \mathbf{t}^*) \geq f(\mathbf{s}^*, \mathbf{t}), \forall \mathbf{t} \in T.$$

The following *alternating convex search* (ACS) algorithm suggests itself:

1. Set $k = 0$ and choose a starting point $(\mathbf{p}^0_{|\mathbf{m}}, \mathbf{Q_m}^0) \in S \times \Delta_M$.

2. Let

$$\mathbf{Q_m}^{k+1} = \arg\max_{\mathbf{Q_m} \in \Delta_M} \Pi(\mathbf{p}^{k+1}_{|\mathbf{m}}, \mathbf{Q_m}) \tag{3.13}$$

4. Let

$$\mathbf{p}^{k+1}_{|\mathbf{m}} = \arg\max_{\mathbf{p}_{|\mathbf{m}} \in S} \Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q_m}^k). \tag{3.14}$$

4. Stop if a convergence criterion is met, or increment k and go to step 2.

The order of steps 2 and 3 is arbitrary. In light of Theorem 3.13 below, the algorithm could be terminated when the change in the solutions from one iteration to the next becomes small. We now show that both problems admit easily computed unique solutions, and that the algorithm converges to the set of stationary points.

3.8.3 Solving the Convex Subproblems

On each iteration, we need only find the roots of $(M+1)$ strictly decreasing functions of one variable. They correspond to each of the M nest, and to the choice over nests. As might be expected, the subproblems are essentially profit maximization problems for a (single-nest) MNL model.

It is natural to express the solutions in terms of "prices" for each nest $Z_1, \ldots Z_M$, and "prices" for each product $z_{1,m_1}, z_{2,m_2}, \ldots, z_{n,m_n}$. (The latter variables are subscripted with the nest of each product because z_{i,m_i} arises from the m_i^{th} subproblem.) At optimality, within the subproblems, all the Z_m (or $z_{i|m}$) are determined by one variable λ (or λ_m) representing the price markup. Such an interpretation is known for the unconstrained MNL model [31] and a similar interpretation of the actual prices $\mathbf{x}$ is has been pursued for the NL model [51, 36].

The next two lemmas give the unique optimal solutions for the subproblems. Their proofs are closely related, but differ because the optimization over $\mathbf{p}_{|\mathbf{m}}$ does not involve a denominator inside the logarithms.

Lemma 3.11. *The solution of* (3.13) *is given by*

$$Q_m^{k+1} = \frac{e^{-\beta_m Z_m}}{1 + \sum_{\ell=1}^{M} e^{-\beta_\ell Z_\ell}}, \qquad \forall m > 0, \qquad \textit{and} \qquad Q_{m^*}^{k+1} = \frac{1}{1 + \sum_{\ell=1}^{M} e^{-\beta_k Z_k}},$$

where we define the nest prices $Z_1, \ldots, Z_M$ *by*

$$Z_m = \lambda + \frac{1}{\beta_m} - \delta_m,$$

and where λ is the uniqe root of the strictly decreasing function

$$F_Q(\lambda) = \sum_{m>0} \frac{1}{\beta_m} e^{-\beta_m Z_m}.$$

Proof. Consider the objective function of (3.13), given explicitly in (3.12). A maximum exists because we may consider the continuous extension of $\Pi(\mathbf{p}_{|\mathbf{m}}, \cdot)$ on the compact closure of Δ_M [see 14, Appendix A.3.3]. Moreover, $\mathbf{Q_m} > \mathbf{0}$ at the maximum because a term of the gradient $\nabla f(x, y)$ in (3.10) goes to negative infinity as either of the variables approaches zero. Therefore we can effectively ignore the positivity constraints at the optimum.

Suppose without loss of generality that $m^* = M$ is the last nest. Assigning dual variable λ to the simplex constraint

$$\sum_{m=1}^{M} Q_m = 1$$

and referring to (3.10), the necessary KKT optimality conditions obtained by differentiating with respect to each of $Q_1, \ldots, Q_M = Q_{m^*}$ are

$$\delta_m - \frac{1}{\beta_m}\left(1 + \log\frac{Q_m}{Q_{m^*}}\right) - \lambda = 0, \qquad m = 1, \ldots, M-1$$
$$\sum_{m=1}^{M} \frac{1}{\beta_m}\frac{Q_m}{Q_{m^*}} - \lambda = 0$$

Solving the first set of equations gives $Q_m/Q_{m^*} = e^{-\beta_m Z_m}, m = 1, \ldots, M-1$. Combining these equations with the simplex constraint, and substituting them into the last KKT condition yields the result. $F_Q(\lambda)$ is strictly decreasing in λ because each term is. $\square$

Lemma 3.12. *The solution of* (3.14) *is given by, for each $i = 1, \ldots, n$ and $m = m_i$,*

$$p_{i|m_i}^{k+1} = e^{-\beta_{im} z_{im}},$$

where we define the conditional prices for each product by

$$z_{im} = \lambda_m + \frac{1}{\beta_{im}} - \delta_{im}, \qquad i = 1, \ldots, n, \qquad m = m_i$$

and where $\lambda_m, m = 1, \ldots, M$ are the unique solutions of the strictly decreasing functions

$$F_{p,m}(\lambda_m) = \sum_{m>0} e^{-\beta_{im} z_{im}} - 1.$$

Proof. The objective function (3.11) is separable over nests $m = 1, \ldots, M$. The problem for each nest is similar to the one we have already solved in Lemma 3.11. We consider a slightly modified version of that problem.

Fix the value of $Q_{m^*} = Q_M = 1$ and replace the simplex constraint with

$$\sum_{m=1}^{M-1} Q_m = 1.$$

We recover the same form as the terms of the objective function (3.11). The KKT conditions from Lemma 3.11 simplify to

$$\delta_m - \frac{1}{\beta_m}\left(1 + \log \frac{Q_m}{Q_M}\right) - \lambda = 0, \qquad m = 1, \ldots, M-1,$$

where $Q_M = 1$ is constant. Solving, for Q_m, we obtain

$$Q_m = \exp\left\{\beta_m\left(\delta_m - \frac{1}{\beta_m} - \lambda\right)\right\} = e^{-\beta_m Z_m}.$$

From the new simplex constraint, we obtain

$$F(\lambda) = \sum_{m>0} e^{-\beta_m Z_m} - 1,$$

with each term strictly decreasing in λ.

Now, consider the m^{th} term of (3.11). The conclusion follows by respectively equating the quantities $n_m, p_{i|m}, \beta_{im}, \delta_{im}, \lambda_m, z_{im}$ and $F_{p,m}(\cdot)$ indexed by i in the statement, with the quantities $M, Q_m, \beta_m, \delta_m, \lambda, Z_m$ and $F(\cdot)$ indexed by m in the proof. $\square$

3.8.4 Convergence of the ACS Algorithm

The ACS algorithm is known to converge to stationary points of biconcave functions under mild conditions. If the stationary point is unique and corresponds to the global maximum, then the ACS algorithm necessarily converges to the global maximum.

When the condition of Theorem 3.5 is satisfied, the profit function $\Pi(\mathbf{p})$ is strictly concave in $\mathbf{p}$ and its unique stationary point is the global maximum. We show that the biconcave profit function $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_{\mathbf{m}})$ then also has a unique stationary point corresponding to the global maximum. This is despite the fact that it is not generally jointly concave in $\mathbf{p}_{|\mathbf{m}}$ and $\mathbf{Q}_{\mathbf{m}}$.

Theorem 3.13. *The sequence of points generated by the ACS algorithm is a Cauchy sequence, and its accumulation points form a connected, compact set of stationary points of the profit* $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_{\mathbf{m}})$.

Proof. Because the unique optimal (3.13) and (3.14) are interior, we can equivalently optimize over the respective closures of S and Δ_M. Moreover, any partial optimum is interior, and therefore a stationary point of the profit [39, Corollary 4.3]. Then by [39, Theorem 4.9], the set of accumulation points (1) is nonempty, (2) consists of partial optima with the same value of $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_{\mathbf{m}})$, and (3) is a connected, compact set. Moreover, by the same theorem, the sequence is Cauchy. $\square$

Corollary 3.14. *If the price sensitivity parameters and the profit margins satisfy the condition of Corollary 3.6, namely*

$$\max_{1 \leq i,j \leq n} \frac{a_i b_j}{b_i a_j} < 2,$$

then the sequence of points generated by the ACS algorithm converges to the unique solution of the pricing problem (3.1) *with* $\mathcal{P} = \Delta_{n+1}$.

Proof. Any improvement direction from a point $\mathbf{p} \in \Delta_n$ maps to an improvement direction from the corresponding point $(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_\mathbf{m}) \in S \times \Delta_M$ and vice-versa. Therefore every stationary point of $\Pi(\mathbf{p}_{|\mathbf{m}}, \mathbf{Q}_\mathbf{m})$ corresponds to a stationary point of $\Pi(\mathbf{p})$.

By Corollary 3.6, the sationary point of the profit $\Pi(\mathbf{p})$ in (3.3) is unique, and it is the global maximum. Then by Theorem 3.13, the sequence of points converges to the unique solution because the sequence is Cauchy and all of its accumulation points corresponds to the global maximum. □

It remains an open question to characterize the speed of convergence of our algorithm. In practice, the line searches at each step can be accomplished quickly, and the overall convergence is much faster than when using a general-purpose solver to maximize Π over values of $\mathbf{p}$. The ACS algorithm should be preffered when there are no constraints to be enforced and when Assumption 3.7 is satisfied.

3.9 Conclusions

We showed that, under the NL model, the profit is a concave function of the market shares if the ratios between the price sensitivity parameters are all less than two, regardless of the scale parameters for each nest. We have derived simple expressions for the gradient and the Hessian matrix of the profit as a function of the market shares. They can be used with general purpose nonlinear optimization algorithms to solve the constrained pricing problem. Our practical observations lead us to believe that the profit remains concave even when our sufficient condition is significantly violated. In Chapters 4 and 5 we explore different but related sufficient conditions for the uniqueness of the optimal solution for the pricing problem under generalizations of the NL model.

For the special case of the NL pricing problem where the outside alternative is in its own nest and there are no constraints, we showed that iteratively solving the concave MNL profit maximization sub-problems for each nest converges to a set of stationary points of the profit. The subproblems can be solved by simple line searches for any values of the price sensitivity parameters. If our concavity condition is satisfied, our algorithm converges to the global optimum.

Chapter 4

Pricing under GEV Demand Models

4.1 Introduction

In this chapter, we consider the problem of selecting prices $z_1, z_2, \ldots, z_n \in \mathbb{R}$ of n substitutable products offered to customers whose purchasing behavior is modeled by a discrete choice model from the family of *generalized extreme value* (GEV) models. We have already discussed the pricing problem under the most common members of this family, the *multinomial logit* (MNL) and *nested logit* (NL) model, in Chapters 2 and 3. However, our results thus far do not extend to the *cross-nested logit* (CNL) or to more complex members of the family such as the *network GEV* model proposed by Daly and Bierlaire [19]. We again consider the more convenient formulation in terms of the probabilities $p_1, p_2, \ldots, p_n$ that a customer purchases each product,

$$\max_{\mathbf{p} \in \Delta_{n+1}} \left\{ \Pi(\mathbf{p}) = \sum_{i=1}^{n} p_i z_i(\mathbf{p}) \right\}, \tag{4.1}$$

The prices $z_i(\mathbf{p})$ are expressed as a function of the vector of choice probabilities $\mathbf{p}$ induced by the GEV model.

We contrast formulation (4.1) with formulation (3.3) for the special case of the NL model. First, profit margins a_i and marginal costs c_i for each product $i = 1, \ldots, n$

can be incorporated by letting $z_i(\mathbf{p}) = a_i(x_i(\mathbf{p}) - c_i)$, where $x_1, \ldots, x_n$ are the true prices, and the parameters of the GEV model have been suitably adjusted. This is described in Section 3.5.1 and generalizes in a straightforward manner. We work only with the adjusted prices $\mathbf{z}$ in this chapter for clarity.

Second, we now consider only the *unconstrained* optimization problem over the entire probability simplex Δ_{n+1}. This is because the profit is generally not a concave function of $\mathbf{p}$. Instead, we will propose a transformation that may not, in general, preserve the convexity of the feasible region $\mathcal{P}$ considered in Chapter 3.

Because there is generally no closed-form expression for the prices z_i in terms of the vector $\mathbf{p}$, we cannot generalize the approach taken for the NL model. Fortunately, the choice probabilities under a GEV model *are* expressed in closed form. We will apply a transformation to problem (4.1) that exploits their properties. Our main contribution in this Chapter is to show that the objective function of the transformed problem is *almost* concave when the sensitivities of customers to the prices of the different products are sufficently close. However, the first issue we must address is whether or not the demands are an invertible (one-to-one *and* onto) function of the prices, so that we may restate the pricing problem in terms of the demands as we have done for the special case of NL models.

First, we present the family of GEV models in Section 4.2. We review the relevant literature on the models and the methods that we employ in Section 4.3. We verify that the demand is indeed an invertible function of the prices for all GEV models in Section 4.5. After presenting our main result in Section 4.6, we further discuss the properties of the original profit function $\Pi(\mathbf{p})$ in Section 4.7.

4.2 The Family of GEV Discrete Choice Models

We first define the model in general over n choices offered to customers, and then adjust our notation to account for the presence of the $(n+1)^{\text{th}}$ choice of not purchasing anything. A GEV discrete choice model is characterized by a homogeneous *GEV generating function* $G : \mathbb{R}^n_+ \to \mathbb{R}_+$. The class of such functions is defined below.

Denote the partial derivative of $G(\mathbf{y})$ with respect to to each entry y_i of $\mathbf{y}$ by $G_i(\cdot)$. The probability that each alternative i is chosen is given by

$$p_i(\mathbf{z}) = \frac{y_i G_i(\mathbf{y})}{\mu G(\mathbf{y})}, \qquad i = 1, \ldots, n. \tag{4.2}$$

The constant μ is the degree of homogeneity of G, and is generally chosen to be $\mu = 1$. The vector $\mathbf{y}$ represents the attraction of each product. We define

$$y_i = \exp\{d'_i - b'_i z_i\}, \qquad i = 1, \ldots, n, \tag{4.3}$$

where the quality parameter $d'_i \in \mathbb{R}$ represents the inherent desirability of product i, and $b'_i > 0$ determines how sensitive customers are to changes in its price. (These parameters may be adjusted to account for varying profit margins and production costs across products, as described in Section 3.5.1.) The formal definition of G and Euler's theorem, also stated below, ensure that the probabilities are positive and sum to one.

Definition 4.1. A function $G : \mathbb{R}^n_+ \to \mathbb{R}_+$ is *homogeneous* of degree μ if for all $\mathbf{y} \in \mathbb{R}^n_+$ and for all $\lambda > 0$, $G(\lambda \mathbf{y}) = \lambda^\mu G(\mathbf{y})$.

Definition 4.2. The function $G : \mathbb{R}^n_+ \to \mathbb{R}_+$ is a μ-GEV generating function if

(i) G is homogeneous of degree $\mu > 0$,

(ii) $G(\mathbf{y}) > 0, \forall \mathbf{y} \in \mathbb{R}^n_{++}$,

(iii) $\lim_{y_i \to \infty} G(\mathbf{y}) = \infty, \forall i = 1, \ldots, n$

(iv) the mixed partial derivatives of G exist and are continuous. Moreover, the k^{th} partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even.

The last part of the definition states that $G_i(\mathbf{y})$ is non-negative, but that its derivative with respect to $y_j, j \neq 1$ is non-positive. Roughly speaking, this implies that the choice probability p_i is non-negative and that it *increases* as the price z_j

increases (and y_j decreases). We make the following additional technical assumption on the demand model to ensure that all the demands are strictly positive for any prices, like in the MNL, NL and CNL models [1]. Recall that we have defined the vector $\mathbf{y} > \mathbf{0}$ such that it is always strictly positive in equation (4.3). We assume that the partial derivatives G_i are also strictly positive.

Assumption 4.3. For each $i = 1, \ldots, n$, we have that $y_i > 0 \Rightarrow G_i(\mathbf{y}) > 0$.

The definition of G already requires that $G_i(\mathbf{y}) \geq 0$ and that $\lim_{y_i \to \infty} G(\mathbf{y}) = \infty$. This assumption eliminates the possibility that demand for a product i remains zero based on some condition on the prices of the *other* products. It is a mild assumption since it is satisfied even if the demand is positive but very small.

For concreteness, we state the GEV generating function for the models considered in the other chapters of this thesis:

Example 4.4. Under the *cross-nested logt* (CNL) model,

$$G(\mathbf{y}) = \sum_{m=1}^{M} \left(\sum_{j=1}^{n} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}},$$

with constant parameters $\alpha_{im} \in [0, 1]$ and $\mu_m > \mu$ such that $\sum_m \alpha_{im} = 1$, for each i. The *nested logit* (NL) model requires that each α_{im} is either 0 or 1 . The *multinomial logit* (MNL) model additionally requires that $\mu_m = \mu, \forall m$, effectively reducing the function to the sum $G(\mathbf{y}) = \sum_i y_i^{\mu}$ of the attractions. The parameter μ is redundant for the MNL model, since it simply scales the parameters d_i' and b_i' defining each y_i.

When $\mu = 1$, the function in Example 4.4 is a sum of weighted p-norms, which are convex. However, we do not assume that G is a convex function in general. Some authors [36, 20] have considered variants of the NL model that allow for "synergistic" products through nest scale parameters $\mu_m < \mu = 1$. These violate part (*iv*) of the definition of G for the second partial derivatives[2], so our results do not apply. Roughly speaking, they may allow a choice probability p_i to increase even when the

[1] This is obvious from Example 4.4 below

[2] Computed in the proof of Lemma 5.9.

price z_j of a different product decreases. For simplicity, we do not pursue such an extension[3]. If the NL model is deemed inappropriate, we would generally suggest that a more powerful GEV model such as the CNL be used instead of attempting to devise NL variants outside of the GEV class. GEV models allow for rich correlation (including a certain "synergy") between products, without violating the assumption that they are substitutes [77].

4.2.1 Euler's Theorem and Corollaries

We introduce additional notation for the partial derivatives of G. Let

$$G_i(\mathbf{y}) = \left.\frac{\partial G}{\partial y_i}\right|_{\mathbf{y}}, \qquad G_{ij}(\mathbf{y}) = \left.\frac{\partial^2 G}{\partial y_i \partial y_j}\right|_{\mathbf{y}}, \qquad \text{and} \qquad G_{ijk}(\mathbf{y}) = \left.\frac{\partial^3 G}{\partial y_i \partial y_j \partial y_k}\right|_{\mathbf{y}}.$$

Also define the gradient, the Hessian, and the tensor of the third derivatives of G, respectively, by,

$$\mathbf{g} = [G_i]_i, \qquad \mathbf{G} = [G_{ij}]_{ij}, \qquad \text{and} \qquad \mathcal{G} = [G_{ijk}]_{ijk}.$$

The main tool used in our proofs is Euler's Theorem [56, Appendix B] and its well-known Corollary 4.6. We apply it to the quantities we have just defined in Corollary 4.7. In particular, this ensures that the choice probabilities sum to one.

Theorem 4.5 (Euler's Theorem [56]). *Let $G : \mathbb{R}^n_+ \to \mathbb{R}_+$ be continuous and differentiable on $\mathbb{R}_{++}$. Then G is homogeneous of degree μ if and only if for all $\mathbf{y} \in \mathbb{R}^n_{++}$,*

$$\mu G(\mathbf{y}) = \sum_{i=1}^{n} y_i G_i(\mathbf{y}),$$

where $G_i(\mathbf{y}) = \left.\frac{\partial G}{\partial y_i}\right|_{\mathbf{y}}$ is the i^{th} partial derivative evaluated at $\mathbf{y}$.

[3] We use that $G_{ij} \leq 0, i \neq j$ to show that the matrix $\mathbf{G}$ defined below is diagonally dominant. This remains true if the off-diagonal G_{ij} are positive but sufficiently small relative to the diagonal elements G_{ii}. We lose the M-matrix property of $\mathbf{L}^{-1}$ but not necessarily of $\mathbf{J}_{\mathbf{z}}^{-1}$. Both matrices still have all-positive principal minors (they are *P-matrices.*). Our results could potentially be extended.

Corollary 4.6. *Let $G : \mathbb{R}^n_+ \to \mathbb{R}_+$ be continuous and twice differentiable on $\mathbb{R}_{++}$. If G is homogeneous of degree μ, then $G_i(\mathbf{y}) = \frac{\partial G}{\partial y_i}\Big|_{\mathbf{y}}$ is homogeneous of degree $(\mu - 1)$.*

Proof. Differentiating both sides of the equality of Theorem 4.5 with respect to y_j

$$\mu G_j(\mathbf{y}) = G_j(\mathbf{y}) + \sum_{i=1}^{n} y_j G_{ij}(\mathbf{y}) \qquad \Leftrightarrow \qquad (\mu - 1)G_j(\mathbf{y}) = \sum_{i=1}^{n} y_j G_{ij}(\mathbf{y}).$$

Then by the second part of Theorem 4.5, $G_j(\mathbf{y})$ is $(\mu - 1)$ homogeneous. □

Corollary 4.7. *If G is a μ-GEV generating function, then $\mathbf{g}$, $\mathbf{G}$, and $\mathcal{G}$ are homogeneous of degree $(\mu - 1)$, $(\mu - 2)$ and $(\mu - 3)$, componentwise, respectively, and*

(i) $\mathbf{y}^\top \mathbf{g} = \sum_j y_j G_j(\mathbf{y}) = \mu G(\mathbf{y})$,

(ii) $\mathbf{G}\mathbf{y}(\mathbf{y}^\top \mathbf{G})^\top = (\mu - 1)\mathbf{g}$,

(iii) $\mathbf{y} \cdot \mathcal{G} = (\mu - 2)\mathbf{G}$,

where $(\mathbf{y} \cdot \mathcal{G}) \in \mathbb{R}^{n \times n}$ denotes the tensor product with (i, j) component

$$(\mu - 2)G_{ij} = \sum_k y_k \mathcal{G}_{ijk}.$$

Proof. Follows from Corollary 4.6 and from Theorem 4.5. □

4.2.2 Notation: Adding an Outside Alternative

Although we have defined the GEV model with n choices, we actually use a model with $(n + 1)$ alternatives for pricing, where the last option represents the possibility that a customer decides not to purchase any of the products. We set $y_{n+1} = 1$ without loss of generality: by homogeneity, scaling all of the y_i variables by $\lambda > 0$ is the same as scaling both the numerator and denominator in (4.2) by λ^μ, and has no impact on the probabilities. Most of our proofs rely on the homogeneity of $G(\mathbf{y})$. The major difficulty arises because we have fixed the last coordinate in this manner, and the resulting function $G(y_1, \ldots, y_n)$ is *no longer homogeneous* in terms of the remaining free variables $y_1, \ldots, y_n$.

Throughout this chapter, we consider quantities related both to the vector $\mathbf{p} \in \mathbb{R}^n$ of choice probabilities, and the vector $\bar{\mathbf{p}} \in \mathbb{R}^{n+1}$ with the last component p_{n+1}. We adopt the convention that quantities of dimension $(n+1)$ are marked with a bar. We define the vectors $\mathbf{y} \in \mathbb{R}^n_+$ and $\bar{\mathbf{y}} \in \mathbb{R}^{n+1}_+$, and the corresponding matrices $\mathbf{Y}$ and $\bar{\mathbf{Y}}$ with the respective vectors on the diagonal. By a slight abuse of notation we continue to write $G(\mathbf{y})$ when it is clear that y_{n+1} is fixed. The notation defined above for the partial derivatives of $G(\mathbf{y})$ becomes $\bar{\mathbf{g}}, \bar{\mathbf{G}}$ and $\bar{\mathcal{G}}$. We re-define quantities $\mathbf{g}, \mathbf{G}$ and $\mathcal{G}$ as the corresponding n-, n^2- and n^3-dimensional quantities with $(n+1)^{th}$ components in each dimension removed, respectively. Care must be taken when applying Corollary 4.7 and similar results, since they apply only to the former $(n+1)$-, $(n+1)^2$- and $(n+1)^3$-dimensional objects $\bar{\mathbf{g}}, \bar{\mathbf{G}}$ and $\bar{\mathcal{G}}$ that include all the components.

Without regard for their dimension, we write $\mathbf{I}$ for the identity matrix, $\mathbf{e}$ for the vector of all ones, and $\mathbf{e}_j$ for the vector of all zeros with a 1 in the j^{th} position. We write $\operatorname{diag}(\mathbf{x})$ to denote the diagonal matrix with the vector $\mathbf{x}$ on the diagonal. That is, for example, $\mathbf{Y} = \operatorname{diag}(\mathbf{y})$ and $\bar{\mathbf{Y}} = \operatorname{diag}(\bar{\mathbf{y}})$

4.3 Literature Review: Pricing under GEV Demand Models

There has been no work, to our knowlege, on revenue management under GEV models in general. We refer the reader to Sections 2.1.1 and 3.3 for a discussion of work on pricing under MNL and NL models, which belong to the GEV family. In Section 5.2 we discuss work on pricing under *mixed logit* (MMNL) models. They further generalize GEV models, in a sense, but they also give rise to profit functions with potentially many local maxima.

Although our results in this section do not assume any particular GEV model, the *cross-nested logit* (CNL) is of interest because it is a powerful but straightforward generalization of the NL model. CNL models are similar to the seemingly less tractable MMNL models, in that they represent a set of nests, each containing *all* of the products. On the other hand, the nest probabilities in a CNL model depend on the prices like for more restrictive NL models, and CNL models remain within the GEV family unlike MMNL models. Both Gallego and Wang [36] and Davis et al. [20] consider certain variants of the NL with a no-purchase option in each nest (for pricing and assortment optimization, respectively). These variants may be thought of as CNL instances with very limited cross-nesting. Unfortunately, the pricing problem under these models requires solving equations for each nest with potentially non-unique roots. The assortment problem, which consists of selecting a subset of products to offer from a mene with *fixed* prices is NP-hard. Daly and Bierlaire [19] further generalize CNL models by allowing for a multi-level nesting structure, but focus only on estimation. By exploiting the defining properties of the GEV family, we aim to solve the pricing problem under all of these discrete choice models.

In Section 4.7, we will show that the demand function under GEV models satisfies the *univalence*[4] condition of Gale and Nikaido [30], so it is one-to-one (injective). This result has been substantially generalized by Mas-Colell [54] and, recently, by Berry et al. [8] for non-differentiable demand functions. Other conditions for univalence are given by Fujisawa and Kuh [29], More [62], El Baz [24] and Frommer [28]. The latter authors provide algorithms to evaluate the inverse demand function computationally, so that optimization over the market-shares is in fact practically possible. Any of these injectivity results could potentially be applied to demand models outside the GEV family to yield a similar approach to ours. For the case of the GEV model, we show that the demand is also *surjective* onto the interior of the probability simplex, so it is not necessary to impose any additional constraints on the domain when optimizing over market shares (demands). Mas-Colell [55, Proposition 2] gives a general sufficient condition for the invertibility of *linearly*-homogeneous functions, but imposes restrictions on their Jacobian at the boundary of the positive orthant, which are violated in our case. We instead show surjectivity based on the necessary optimality conditions of a certain auxiliary minimization problem.

[4]This term is synonymous with injectivity, but is traditionally used in economics to refer to this specific result.

Our transformation of the pricing maximization problem relies on an intermediate optimization problem defined by homogeneous functions. The duality result we apply is a special case of the far more general family of such problems considered by Lasserre and Hiriart-Urruty [47]. They use this approach to show that certain nonconvex quadratic optimization problems can be reduced to convex minimization problems. Our transformed problem has the nonlinear fractional programming form considered by Dinkelbach [22], who shows that such problems have quasiconcave objectives when the numerator and denominator are concave and convex, respectively. A concise survey of theoretical results for nonlinear fractional programs is provided by Schaible and Shi [68]. Unfortunately, the numerator and denominator do not satisfy the requirements for quasi-concavity in our case.

4.4 Outline of the Proofs

Like for the special cases of the MNL and NL models discussed in Chapters 2 and 3, we proceed by first showing that the GEV demand function is invertible. We then appropriately reformulate the pricing problem by using the inverse demand function, and exploit the structure of the reformulation. However, the profit Π is generally *not* a concave function of the demand vector $\mathbf{p}$, even for the CNL model. Instead, we pursue a reformulation in terms of the *unnormalized demands* $\mathbf{q}$, which are different from but closely related to the demands $\mathbf{p}$. As the name suggests, $\mathbf{p}$ is simply the vector $\mathbf{q}$ normalized to yield a probability distribution (taking into account the no-purchase probability p_{n+1}).

Section 4.5 is concerned with the invertibility of the demand function. We characterize the Jacobian matrix $\mathbf{L}^{-1}$ of $\mathbf{q}$ with respect to $\mathbf{y}$ as an M-matrix by expressing it in terms of the partial derivatives of the generating function G in (Lemma 4.8).[5] This characterization is sufficient to show surjectivity of the demand function in The-

[5]In Chapter 3, we instead worked with the Jacobian matrix $\mathbf{J}_{\mathbf{z}}$ of $\mathbf{z}$ with respect to $\mathbf{p}$ and obtained a closed form for the NL model. This is not possible in general, so we instead work with the analogue of its inverse, the Jacobian matrix $\mathbf{J}_{\mathbf{z}}^{-1}$ of $\mathbf{p}$ with respect to $\mathbf{z}$. We also use the unnormalized demands $\mathbf{q}$ instead of $\mathbf{p}$. The vectors $\mathbf{y}$ and $\mathbf{z}$ are essentially interchangeable, since their relationship is straighforward and there is a clear bijection between the two.

orem 4.9, where the necessary optimality conditions of an auxiliary minimization problem are used to also show injectivity. This completes the proof that the demand is invertible in that there is a bijection between $\mathbf{z}$ and $\mathbf{p}$, and equivalently between $\mathbf{y}$ and $\mathbf{q}$. The characterization of $\mathbf{L}^{-1}$ also implies that the demand can be inverted computationally using classical algorithms for solving nonlinear systems of equations.

We describe our reformulation in Section 4.6. The pricing problem expressed in terms of $\mathbf{q}$ is a *nonlinear fractional program* with objective function

$$\Pi(\mathbf{q}) = \frac{\Psi(\mathbf{q})}{\mu G(y(\mathbf{q}))},$$

where $\Psi(\mathbf{q}) = \mathbf{q}^\top z(\mathbf{q})$ is simply the unnormalized profit. Unfortunately, the pricing problem is not a *concave-convex fractional program*, for which solution methods are known. Nevertheless, in the proof of Theorem 4.10, we restate the pricing problem as a *homogeneous optimization problem* and recover a parametric programming formulation of the same form as is sometimes used to solve concave-convex fractional programs. The objective function is parameterized by $\lambda > 0$:

$$\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda\mu G(y(\mathbf{q})).$$

The function Γ is decreasing in the parameter λ. For the optimal value of $\lambda = \lambda^*$, $\max_{\mathbf{q}} \{\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda^*\mu G(y(\mathbf{q}))\} = 0$, and the solution corresponds to a solution of the GEV pricing problem (4.1). Then being able to maximize Γ is sufficient to find the optimal value of λ via a line search. The proof relies on a duality result for homogeneous problems, which we defer to Lemma 4.11 in order to keep the discussion separate.

Using the expression for $\mathbf{L}^{-1}$ discussed above, we then derive an expression for the Hessian matrix of the parametric programming objective function Γ with respect to $\mathbf{q}$ in the sequence of Lemmas 4.12 through 4.14. A certain sub-stochastic matrix that we denote by $\mathbf{S}$ plays an important role in the expression for the Hessian (as well as the gradient) of the objective function. It is analyzed in Lemmas 4.15 and C.1 and used in the proof of the main result stated in Theorem 4.16. Specifically, we show that, if the ratio of the price-sensitivity parameters is bounded by two[6], then the Hessian matrix of the parametric programming objective function consists of a negative-definite matrix plus two correction terms. In computational experiments, these terms are usually small, and therefore the objective function is usually concave.

Finally, in order to allow general-purpose nonlinear optimization algorithms to be applied directly to GEV pricing problem, we derive the Jacobian matrix $\mathbf{J}^{-1}$ of the normalized demands $\mathbf{p}$ with respect to $\mathbf{y}$ in Proposition 4.17 of Section 4.7. We then obtain expressions for the Jacobian matrix $\mathbf{J}_{\mathbf{z}}$ of the prices $\mathbf{z}$ with respect to the demands $\mathbf{p}$, and for the Hessian matrix of the profit $\Pi(\mathbf{p})$. Unlike for the special case of the NL model, there is no closed form for the profit in terms of $\mathbf{p}$, and a matrix inversion is required to compute the partial derivatives of the profit, in general.

4.5 Invertibility of GEV Demand Models

It will become clear that general results from economics [30, 54, 8] imply that the demand is univalent (injective, one-to-one) after we are finally able to derive the Jacobian matrix of the demands $\mathbf{p}$ with respect to the prices $\mathbf{z}$ in Section 4.7. However, we must also show that there is a vector of prices $\mathbf{z}$ which results in an arbitrary vector of choice probabilities $\mathbf{p} \in \Delta_{n+1}$. If this were not the case, we would have to enforce constraints on the feasible region when optimizing over $\mathbf{p}$ that might make the problem intractable. Fortunately, the demand under GEV models is also surjective onto the interior of the probability simplex. This fact is unsurprising since these statistical models are intended to fit actual data, but it requires proof in the pricing context where the model parameters other than the prices have already been fixed.

Rather than rely on the results mentioned above, we prove invertibility by working with the *unnormalized* demands defined in the next section. This allows use to significantly lighten the notation, prove surjectivity as well as injectivity, and introduce some notation that will be used in later proofs.

[6]This is essentially the same condition required for concavity of the profit under NL demand models in Lemma B.2 in Chapter 3

4.5.1 The Jacobian of the Unnormalized Demand Function

Working directly with the demand function $p(\mathbf{z})$ is somewhat cumbersome. Instead, in most of this chapter, we consider the μ-homogeneous function $q : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ defined by

$$q_i(\mathbf{y}) = y_i G_i(\mathbf{y}), \qquad i = 1, \ldots, n+1.$$

We also define the vectors $\mathbf{q}$ and $\bar{\mathbf{q}}$ analogously to $\mathbf{p}$ and $\bar{\mathbf{p}}$. Obviously, $\bar{\mathbf{q}} = \bar{\mathbf{p}} \cdot \mu G(\mathbf{y})$. The function $q(\mathbf{y})$ is then simply the demand function before it has been normalized to yield a probability vector.

The Jacobian of $q(\cdot)$ is readily derived [7] as

$$\bar{\mathbf{L}}^{-1} \triangleq \left[\frac{\partial q_j}{\partial y_i}\right]_{ij} = \mathrm{diag}\,(\bar{\mathbf{g}}) + \bar{\mathbf{G}}\bar{\mathbf{Y}}.$$

Denote by $\mathbf{L}^{-1}$ the submatrix of $\bar{\mathbf{L}}^{-1}$ with the last row and column removed. Then

$$\mathbf{L}^{-1} \triangleq \left[\frac{\partial q_j}{\partial y_i}\right]_{ij} = \mathrm{diag}\,(\mathbf{g}) + \mathbf{G}\mathbf{Y}$$

is the Jacobian of $\mathbf{q} = q(\mathbf{y}) = \mathbf{Y}\mathbf{g}$. We use the inverse notation for consistency with later notation. Both $\bar{\mathbf{L}}^{-1}$ and $\mathbf{L}^{-1}$ are *M-matrices*, and as such they are invertible. (See Definition 3.2 in Section 3.6.)

Lemma 4.8. *For $\mathbf{y} > \mathbf{0}$, the matrix $\bar{\mathbf{L}}^{-1}$ is a strictly row diagonally dominant M-matrix. The same holds for $\mathbf{L}^{-1}$ and any other principal submatrix obtained by removing the rows and columns indexed by any set $N \subset \{1, 2, \ldots, n+1\}$.*

Proof. From Corollary 4.7, we have that the row-sums of $\bar{\mathbf{L}}^{-1}$ are

$$\bar{\mathbf{L}}^{-1}\mathbf{e} = \bar{\mathbf{g}} + \bar{\mathbf{G}}\mathbf{y} = \bar{\mathbf{g}} + (\mu - 1)\bar{\mathbf{g}} = \mu\bar{\mathbf{g}} > 0, \tag{4.4}$$

where positivity in each coordinate follows from Assumption 4.3. On the other hand, from the definition of $G(\cdot)$, the off-diagonal elements of $\bar{\mathbf{L}}^{-1}$ are non-positive. Therefore, each diagonal element must be positive and larger in magnitude than the sum

[7] For an explicit derivation, see the proof of Lemma 4.12 that states the second partial derivatives.

of the off-diagonal elements in its row. It follows that $\bar{\mathbf{L}}^{-1}$ is strictly diagonally dominant, and an M-matrix. This property remains unchanged by deleting rows along with their corresponding columns, because we are simply removing non-positive terms from the sum above. $\square$

4.5.2 Proving Invertibility of the Demand Function

Before proceeding, recall that the relationship between $\mathbf{y}$ and $\mathbf{z}$ defined in (4.3) is clearly invertible. Then the vector function $p(\mathbf{z})$ defined by (4.2) can equivalently be written as a function $p(\mathbf{y}) = p(z(\mathbf{y}))$ of $\mathbf{y}$. Temporarily relaxing the assumption made in Section 4.2.2, that $y_{n+1} = 1$ in the vector $\bar{\mathbf{y}}$, allows us to exploit homogeneity of $G(\bar{\mathbf{y}})$ in the following theorem.

Theorem 4.9. *For GEV models, the mapping $p : \mathbb{R}^n \to \Delta_{n+1}$ from vectors of prices $\mathbf{z}$ to choice probabilities $\mathbf{p} > \mathbf{0}$ is invertible (one-to-one and onto). Moreover, both $p(\mathbf{z})$ and the inverse function $z(\mathbf{p}) = p^{-1}(\mathbf{z})$ are differentiable.*

Proof. To show that $p(\cdot)$ is surjective (onto), we fix a vector $\mathbf{p} \in \Delta_{n+1}$ and recover the corresponding vector of prices $\mathbf{z} \in \mathbb{R}^n$. The function $p(\cdot)$ is homogeneous of degree 0 in $\bar{\mathbf{y}}$ because it is invariant to scaling of $\bar{\mathbf{y}}$, as mentioned above when we fixed $y_{n+1} = 1$. We will find a suitable element of $\bar{Y} = \left\{\bar{\mathbf{y}} \in \mathbb{R}^{n+1}_{++} : G(\bar{\mathbf{y}}) = \frac{1}{\mu}\right\}$. Afterwards, we can scale $\bar{\mathbf{y}}$ such that $y_{n+1} = 1$ to recover $\mathbf{y}$ and the corresponding $\mathbf{z}$. Over the set $\bar{Y}$, $\mathbf{p} = \mathbf{q}$ by definition, so we need only show that the system of equations

$$q_i(\bar{\mathbf{y}}) = y_i G_i(\bar{\mathbf{y}}) = q_i, \qquad i = 1, \ldots, n+1$$

has a solution for any fixed vector $\bar{\mathbf{q}} > 0$. Dividing each equation by $y_i > 0$, we have the equivalent system

$$G_i(\bar{\mathbf{y}}) = \frac{q_i}{y_i}, \qquad i = 1, \ldots, n+1.$$

These are exactly the necessary first-order optimality conditions of the unconstrained minimization problem

$$\min_{\bar{\mathbf{y}}>0}\left\{Z(\bar{\mathbf{y}}) = G(\bar{\mathbf{y}}) - \sum_{i=1}^{n+1} q_i \log y_i.\right\}$$

The objective function $Z(\cdot)$ is coercive on its domain, that is, $\lim_{t\to\infty} Z(\bar{\mathbf{y}}^t) = \infty$ for any sequence $\{\bar{\mathbf{y}}^t\}$ converging to a point on the boundary of the positive orthant. This follows from the definitions of G and of the logarithm. Specifically, keeping the other coordinates fixed,

$$\lim_{y_i\to\infty} Z(\mathbf{y}) = \lim_{y_i\to 0} Z(\mathbf{y}) = \infty \qquad \text{for each } y_i.$$

The limit as y_i tends to zero is immediate from homogeneity of G. The limit as y_i becomes large follows because G grows exponentially along a ray by homogeneity, whereas the negative term grows only logarithmically. Then $Z(\cdot)$ achieves a minimum where the first-order optimality conditions are satisfied (if we consider the closure of the domain [see, for example, 9, Proposition A.8]). Recovering the prices $\mathbf{z}$ from $\mathbf{y}$, after scaling such that $y_{n+1} = 1$, is trivial. Differentiability follows from the definition of G and the inverse function theorem [8], since the Jacobian $\bar{\mathbf{L}}^{-1}$ of $q(\cdot)$ is invertible by Lemma 4.8.

Rather than attempt to directly show that the maximum of $Z(\mathbf{y})$ is also unique, we apply a result from the literature. To show that $p(\cdot)$ is injective (one-to-one), it is sufficent to show that $q(\cdot)$ is an injective function of $\bar{\mathbf{y}}$, because the scaling of $\bar{\mathbf{y}}$ such that $y_{n+1} = 1$ is clearly unique. Lemma 4.8 states that the Jacobian $\bar{\mathbf{L}}^{-1}$ of $q(\cdot)$ is strictly diagonally dominant on the convex domain $\mathbf{y} > \mathbf{0}$. Then $q(\cdot)$ is a *strictly diagonally dominant function* [62, Theorem 2.5]. As such it is injective (one-to-one) [62, Theorem 3.3]. □

[8]The inverse function theorem states that if the Jacobian matrix of a function is non-singular at a point, then the function is invertible in a neighborhood, and is also continuously differentiable, and the Jacobian matrix of the inverse function is the inverse of the original Jacobian matrix. This implies that $q(\cdot)$ is invertible. Because $G(\mathbf{y}) > 0$ for $y > 0$, then $q(\cdot)$ is also invertible.

4.5.3 The Jacobian of the Prices

By the inverse function theorem, the matrix

$$\mathbf{L} = \left[\frac{\partial y_j}{\partial q_i}\right]_{ij}$$

is the Jacobian matrix of the vector $\mathbf{y}$ with respect to the vector of unnormalized demands $\mathbf{q}$. We define the diagonal Jacobian matrix

$$\mathbf{D}^{-1} \triangleq \left[\frac{\partial y_j}{\partial z_i}\right]_{ij} = \operatorname{diag}\left(-b'_1 y_1, -b'_2 y_2, \ldots, -b'_n y_n\right) = -\mathbf{BY}.$$

containing the derivatives of $y_j = \exp\{d'_j - b'_j z_j\}$ with respect to each z_i. Immediately,

$$\mathbf{D} \triangleq \left[\frac{\partial z_j}{\partial y_i}\right]_{ij} = \operatorname{diag}\left(\left[\tfrac{-1}{b'_1 y_1}, \ldots, \tfrac{-1}{b'_n y_n}\right]\right) = -\mathbf{Y}^{-1}\mathbf{B}^{-1}.$$

Then, by the vector chain rule, the Jacobian matrix of prices with respect to $\mathbf{q}$ is

$$\mathbf{LD} = \left[\frac{\partial z_j}{\partial q_i}\right]_{ij} = -\mathbf{LY}^{-1}\mathbf{B}^{-1}.$$

This matrix is analogous to the Jacobian matrix $\mathbf{J_z}$ of the prices with respect to the normalized demand vector $\mathbf{p}$ that we will define later in Section 4.7 for GEV models. We have already expressed $\mathbf{J_z}$ in closed form for the special case of the NL model in Section 3.6. The subsequent development is greatly simplified by working with $\mathbf{LD}$ and its inverse instead.

4.5.4 Inverting the Demand Function Computationally

In the proof of Theorem 4.9, we showed that $q(\mathbf{y})$ belongs to the class of *stricly diagonally-dominant functions*. It is also an *M-function* [62, Theorem 4.5]. Such functions can be inverted using the nonlinear Jacobi, Gauss-Seidel and related asynchronous methods [62, 24]. Essentially, these methods proceed by independently solving each equation in a system of nonlinear equations for one variable while keeping the other variables fixed. They differ in the order and frequency of variables updates. The resulting nonlinear operators can be shown to converge when the system of equations is defined by an M-function. In the linear case, these algorithms correspond with the linear Jacobi and Gauss-Seidel iterations.

A linear damped Jacobi iteration similar to that presented in Chapter 5 for *solving* the pricing problems has proved effective in our computational experiments, without requiring the solution of any nonlinear equations as in the Jacobi and Gauss-Seidel methods. We expect that any general purpose algorithm for solving nonlinear systems would also perform well.

4.6 A Reformulation of the Pricing Problem

We restate the pricing problem in terms of the unnormalized demands $\mathbf{q}$ introduced in the last section. Recall that $q_i(\mathbf{y}) = y_i G_i(\mathbf{y}) = p_i \cdot \mu G(\mathbf{y})$. We have shown that the mapping from $\mathbf{z}$ (and $\mathbf{y}$) to the demands $\mathbf{p}$ (and $\mathbf{q}$) is invertible. The objective function of the problem (4.1), when expressed as a function of $\mathbf{y}$, is

$$\Pi(\mathbf{y}) = \mathbf{p}^\top \mathbf{z} = \frac{\sum_{i=1}^n y_i G_i(\mathbf{y}) z_i(\mathbf{y})}{\mu G(\mathbf{y})}.$$

This is a nonlinear fractional program [22], but the numerator is not concave in $\mathbf{y}$ in general and the denominator may not be convex. The numerator is also *not* homogeneous because each z_i varies logarithmically in $\mathbf{y}$. When expressed in terms of $\mathbf{q}$, the profit is

$$\Pi(\mathbf{q}) = \frac{\mathbf{q}^\top \mathbf{z}}{\mu G(y(\mathbf{q}))} = \frac{\sum_{i=1}^n q_i z_i(\mathbf{q})}{\mu G(y(\mathbf{q}))}.$$

The denominator may no longer be convex in $\mathbf{q}$, even if G is a convex function of $\mathbf{y}$. On the other hand, we will show that the numerator is *often* concave, and the denominator is *nearly* convex over the region of interest. If they were concave and convex, respectively, then $\Pi(\mathbf{q})$ would be a quasi-concave function by well-known results from fractional programming [22, 68].

In the next theorem, we express the pricing problem as an optimization problem over $\mathbf{y}$ with a single constraint. The problem is defined entirely by homogeneous functions. By duality, we recover an optimization problem parameterized by the scalar λ. The optimal value λ^* corresponds to the optimal profit $\Pi^* = \Pi(\mathbf{p}^*)$, and the optimal solutions of the parametric program with $\lambda = \lambda^*$ correspond to the optimal solutions of the pricing problem. Such a transformation is well-known for fractional programs, and the duality result is a special case of a known result for minimization problems defined by homogeneous functions [47]. We define the function Ψ to denote the numerator

$$\Psi(\mathbf{y}) = \mathbf{q}^\top \mathbf{z} = \sum_{i=1}^{n} y_i G_i(\mathbf{y}) z_i(y_i),$$

where it is understood that $y_{n+1} = 1$ remains fixed, and the vectors $\mathbf{q}$ and $\mathbf{z}$ are functions of $\mathbf{y}$. By invertibility, we can equivalently write $\Psi(\mathbf{q})$, as we did above in the expression for $\Pi(\mathbf{q})$. Notice that each price z_i is a scalar function of y_i, but that it depends on the entire vector $\mathbf{q}$ through y_i. We point out that q_{n+1} (and p_{n+1}) is not fixed in general, but it is nevertheless determined by the n entries of $\mathbf{q}$ (and $\mathbf{p}$).

4.6.1 The Parametric Programming Formulation

Theorem 4.10. *The optimal objective value λ^* of the pricing problem* (4.1) *is the unique value of λ such that*

$$\max_{\mathbf{y}} \{\Gamma(\mathbf{y}) = \Psi(\mathbf{y}) - \lambda\mu G(\mathbf{y})\} = 0.$$

Moreover, the maximizers of $\Pi(\mathbf{y})$ are exactly the maximizers of $\Gamma(\mathbf{y})$ when $\lambda = \lambda^*$.

Proof. Because $y_{n+1} = 1$, then $y_i = y_i/y_{n+1}$ for each i. Because $z_i = \frac{1}{b_i'}(d_i - \log y_i)$

by definition of y_i, we can equivalently write

$$\begin{aligned}
\Pi(\bar{\mathbf{y}}) &= \mathbf{p}^{\top}\mathbf{z} \\
&= \sum_{i=1}^{n} \frac{1}{b_i'} p_i \left(d_i' - \log y_i\right) \\
&= \sum_{i=1}^{n} \frac{1}{b_i'} \frac{y_i G_i(\bar{\mathbf{y}})}{\mu G(\bar{\mathbf{y}})} \left(d_i' - \log \frac{y_i}{y_{n+1}}\right) \\
&= \frac{H(\bar{\mathbf{y}})}{\mu G(\bar{\mathbf{y}})}
\end{aligned}$$

where both $G(\cdot)$ and

$$H(\bar{\mathbf{y}}) \triangleq \sum_{i=1}^{n} \frac{1}{b_i'} y_i G_i(\bar{\mathbf{y}}) \left(d_i' - \log \frac{y_i}{y_n}\right),$$

are homogeneous functions of degree μ. The function $H(\bar{\mathbf{y}})$ is μ-homogeneous since $G_i(\bar{\mathbf{y}})$ is $(\mu-1)$-homogeneous by Corollary 4.7, since each y_i is clearly 1-homogeneous, and since the quantity in parentheses is also clearly 0-homogeneous. But then,

$$\max_{\bar{\mathbf{y}}>\mathbf{0}, y_{n+1}=1} \frac{H(\bar{\mathbf{y}})}{\mu G(\bar{\mathbf{y}})} = \max_{\bar{\mathbf{y}}>\mathbf{0}} \frac{H(\bar{\mathbf{y}})}{\mu G(\bar{\mathbf{y}})} = \max_{\bar{\mathbf{y}}>\mathbf{0}, G(\bar{\mathbf{y}})=\frac{1}{\mu}} \frac{H(\bar{\mathbf{y}})}{\mu G(\bar{\mathbf{y}})} = \max_{\bar{\mathbf{y}}>\mathbf{0}, G(\bar{\mathbf{y}})=\frac{1}{\mu}} H(\bar{\mathbf{y}})$$

The equality in the last problem can be relaxed to yield

$$\begin{aligned}
\max \quad & H(\bar{\mathbf{y}}) \\
\text{s.t.} \quad & G(\bar{\mathbf{y}}) \leq \frac{1}{\mu} \\
& \bar{\mathbf{y}} > \mathbf{0}
\end{aligned} \tag{P}$$

because both $H(\bar{\mathbf{y}})$ and $G(\bar{\mathbf{y}})$ are positive at optimality, as well as homogeneous. Therefore, if the constraint were not tight, the objective could be increased by scaling up $\bar{\mathbf{y}}$. To see that the profit is positive at optimality, observe that for sufficiently large values of y_{n+1} (or sufficiently small values of $\mathbf{y}$), all the prices $\mathbf{z}$ are positive.

We defer the discussion of the duality result that we use until the next section for clarity. By Lemma 4.11, proved below, the dual of (P) is

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t} \quad & 0 \geq \max_{\bar{\mathbf{y}}>\mathbf{0}} H(\bar{\mathbf{y}}) - \lambda\mu G(\bar{\mathbf{y}}) \\ & \lambda \geq 0 \end{aligned} \tag{D}$$

The value of the constraint in (D) is strictly decreasing in λ, uniformly over all values of $\bar{\mathbf{y}}$. Therefore all feasible λ are contained in the half line $[\lambda^*, \infty)$, and the dual optimal solution is some λ^* for which the constraint holds with equality. We can scale $\bar{\mathbf{y}}$ such that $\mu G(\bar{\mathbf{y}}) = 1$. The same scaling is applied to $H(\bar{\mathbf{y}})$ by homogeneity. Then the optimal objective value of the primal (P) is also λ^*, by weak duality. Clearly, the solutions of the two problems coincide, up to scaling. But we fix the scaling in the statement by setting $y_{n+1} = 1$.

By a similar reasoning as above, the constraint can be rewritten as

$$0 \geq \max_{\bar{\mathbf{y}}>\mathbf{0}} H(\bar{\mathbf{y}}) - \lambda\mu G(\bar{\mathbf{y}}) \Leftrightarrow 0 \geq \max_{\bar{\mathbf{y}}>\mathbf{0}, y_{n+1}=1} H(\bar{\mathbf{y}}) - \lambda\mu G(\bar{\mathbf{y}}),$$

because the sign of the right hand side does not change under scaling of $\bar{\mathbf{y}}$. But when $y_{n+1} = 1$,

$$H(\bar{\mathbf{y}}) = \sum_{i=1}^{n} y_i G_i(\mathbf{y}) z_i = \sum_{i=1}^{n} q_i z_i = \mathbf{q}^\top \mathbf{z} = \Psi(\mathbf{y}).$$

This yields the desired statement. $\square$

4.6.2 Duality for Unconstrained Homogeneous Problems

Lemma 4.11. *Suppose G and H are non-zero homogeneous of degree μ and that the set Y is a cone. Then the Lagrangian dual of* (P) *is* (D).

Proof. The Lagrangean dual of (P) is

$$\begin{aligned}
&\inf_{\lambda \geq 0} \sup_{\mathbf{y} \in Y} H(\mathbf{y}) - \lambda\left(\mu G(\mathbf{y}) - 1\right) \\
&= \inf_{\lambda \geq 0} \lambda + \sup_{\mathbf{y} \in Y} H(\mathbf{y}) - \lambda \mu G(\mathbf{y})
\end{aligned}$$

The function $f(\mathbf{y}) = H(\mathbf{y}) - \lambda \mu G(\mathbf{y})$ is homogeneous of degree μ. If there exists $\mathbf{y} \in Y$ such that $f(\mathbf{y}) > 0$, then $f(\mathbf{y})$ can be made arbitrarily large by scaling up $\mathbf{y}$. If not, it can be made arbitrarily close to zero by scaling down $\mathbf{y}$. Thus, for a given value of λ, the value of inner maximization above is either zero or infinite. If it is infinite for all values of λ, the optimal dual solution is $+\infty$ and the primal is unbounded because the numerator $H(\cdot)$ can be made arbitrarily larger than $G(\cdot)$. Therefore, the dual can be written as

$$\begin{aligned}
\inf \quad & \lambda \\
\text{s.t} \quad & 0 \geq \sup_{\mathbf{y} \in Y} H(\mathbf{y}) - \lambda G(\mathbf{y}) \\
& \lambda \geq 0.
\end{aligned}$$

□

We have stated Lemma 4.11 in terms of a general cone, rather than only over the set $\{\mathbf{y} > \mathbf{0}\}$. This suggests that the result can be generalized to allow a conic feasible region in $\mathbf{y}$. We omit such an extension for simplicity, and limit our goal to showing that the unconstrained reformulation is tractable.

As mentioned when defining the demand model, the reformulation is made difficult because we have fixed the value of y_{n+1}. One might consider solving the homogeneous maximization problem in terms of $H(\bar{\mathbf{y}})$ instead. But then the Jacobian matrix $\mathbf{D}$ of vectors $\mathbf{y}$ with respect to $\mathbf{z}$ would no longer be square, nor diagonal. It would gain a dense last row, because each price z_j would depend on the value of y_{n+1}. This is closer to the approach of Chapter 3, where we computed the Hessian of the profit with respect to the full vector $\bar{\mathbf{p}}$ for the special case of the NL model.

In Chapter 2, we applied the *generalized Charnes-Cooper* variable transformation, another well-known technique for fractional programming [66]. The dual pricing problem under generalized attraction (GA) models in Section 2.5 is similar to the dual in Theorem 4.10. There, we considered a constrained problem, and the "attractions" $q_i = y_i G_i(\mathbf{y})$ were simply denoted by y_i. Unlike for GEV models, they were each functions of only the price of product i.

4.6.3 The Hessian of the Reformulation

It is straighforward to derive the Hessian matrices of Ψ and G with respect to $\mathbf{y}$. Using a chain rule, we then obtain the Hessians with respect to $\mathbf{q}$. To this end, we require the Hessians matrices of each q_k with respect to the vector $\mathbf{y}$:

Lemma 4.12. *The Hessian matrix of q_k with respect to $\mathbf{y}$, for $k = 1, \ldots, n+1$, is*

$$\bar{\mathcal{K}}^k_{ij} = y_k \bar{\mathcal{G}}^k + \bar{\mathbf{G}}^k \mathbf{e}_k^\top + \mathbf{e}_k (\bar{\mathbf{G}}^k)^\top,$$

where $\bar{\mathbf{G}}^k = \bar{\mathbf{G}}\mathbf{e}_k$ is the k^{th} column (or row, transposed) of the symmetric matrix $\bar{\mathbf{G}}$. Removing the last row and column, the Hessian with respect to $\mathbf{y}$ is

$$\mathcal{K}^k_{ij} = y_k \mathcal{G}^k + \mathbf{G}^k \mathbf{e}_k^\top + \mathbf{e}_k (\mathbf{G}^k)^\top.$$

Moreover, these Hessian matrices $\bar{\mathcal{K}}^k$ are related to the Hessian matrix $\bar{\mathbf{G}}$ of $G(\cdot)$ by

$$\mathbf{e} \cdot \bar{\mathcal{K}} = \sum_{k=1}^{n+1} \bar{\mathcal{K}}^k = \mu \bar{\mathbf{G}}.$$

Proof. See Appendix C.1.1. □

Lemma 4.13. *The Hessian of $G(\mathbf{y})$ with respect to $\mathbf{q}$ is*

$$\frac{\partial^2 G}{\partial \mathbf{q} \partial \mathbf{q}} = \mathbf{L}\left(\mathbf{G} - \mathbf{L}\mathbf{g} \cdot \mathcal{K}\right)\mathbf{L}^\top$$

Proof. This follows from the chain rule for Hessians of Proposition B.1. □

Lemma 4.14. *The gradient of* $\Psi(\mathbf{y})$ *with respect to* $\mathbf{q}$ *is*

$$\frac{\partial \Psi}{\partial \mathbf{q}} = \mathbf{z} + \mathbf{LDq}$$

and its Hessian is

$$\frac{\partial^2 \Psi}{\partial \mathbf{q} \partial \mathbf{q}} = \mathbf{L}\left(\mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} - \mathbf{LDq} \cdot \mathcal{K} - \mathbf{D}\,\mathrm{diag}\left(\mathbf{g}\right)\right)\mathbf{L}^{\top}.$$

Proof. See Appendix C.1.2. □

4.6.4 The Sub-stochastic Matrix S

The expressions for the Hessians of $\Psi(\cdot)$ and $G(\cdot)$ involve the terms $\mathbf{LDq} \cdot \mathcal{K}$ and $\mathbf{Lg} \cdot \mathcal{K}$, respectively. To analyze them, we define the scaled sub-stochastic matrix

$$\mathbf{S} \triangleq \mathbf{L}\,\mathrm{diag}\left(\mathbf{g}\right) \in \mathbb{R}^{n \times n}.$$

We re-express the vectors in the two tensor products above as

$$\mathbf{Lg} = \mathbf{Se} \qquad \text{and} \qquad -\mathbf{LDq} = \mathbf{LB}^{-1}\mathbf{Y}^{-1}\mathbf{q} = \mathbf{SB}^{-1}\mathbf{e}.$$

The analogous matrix $\bar{\mathbf{S}} \triangleq \bar{\mathbf{L}}\,\mathrm{diag}\left(\bar{\mathbf{g}}\right) \in \mathbb{R}^{(n+1)\times(n+1)}$ is actually stochastic when scaled by μ. Then $\mu\bar{\mathbf{S}}\mathbf{e} = \mathbf{e}$. We showed in Lemma 4.12 that

$$\mu\bar{\mathbf{S}}\mathbf{e} \cdot \bar{\mathcal{K}} = \mathbf{e} \cdot \bar{\mathcal{K}} = \sum_{k=1}^{n+1} \bar{\mathcal{K}}^k = \mu\bar{\mathbf{G}}.$$

Unfortunately, when working with $\mathcal{K}$ instead of $\bar{\mathcal{K}}$ we have removed the last term of this summation. This gives rise to the second error term $\tilde{\mathbf{G}}$ below. The first error term is due to the fact that $\mathbf{B}^{-1} \neq \mathbf{I}$, in general.

Lemma 4.15. *The inverse M-matrix* $\mu\mathbf{S}$ *is sub-stochastic. That is, its entries are* *non-negative and its rows sum to less than one. Moreover,*

$$\begin{aligned}\mathbf{S}\mathbf{B}^{-1}\mathbf{e}\cdot\mathcal{K} = \mathbf{L}\mathbf{B}^{-1}\mathbf{g}\cdot\mathcal{K} &= -\mathbf{L}\mathbf{D}\mathbf{q}\cdot\mathcal{K} \\ &= \mathbf{G} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}},\end{aligned}$$

where $\tilde{\mathbf{G}}$ is the upper-left $n \times n$ sub-matrix of the tensor product

$$\frac{1}{\mu}\left(\frac{1}{\bar{S}_{n+1,n+1}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right)\cdot\bar{\mathcal{K}}.$$

Proof. See Appendix C.1.3, following the analogous result for $\bar{\mathbf{S}}$. □

To interpret the error term $\tilde{\mathbf{G}}$, recall that Lemma 4.12 implies that $\frac{1}{\mu}\mathbf{e}\cdot\bar{\mathcal{K}} = \frac{1}{\mu}\sum_{k=1}^{n+1}\bar{\mathcal{K}}^k = \mathbf{G}$. Here, the vector of all ones $\mathbf{e}$ is replaced by the positive vector consisting of the last column of the stochastic matrix $\mu\bar{\mathbf{S}}$, normalized by its last entry. Unlike for $\bar{\mathbf{S}}^{-1}$, we are unable to establish a relationship bounding the off-diagonal elements of $\mathbf{S}$ in terms of the diagonal elements, so the error term $\tilde{\mathbf{G}}$ can potentially have a large impact relative to $\mathbf{G}$.[9] In practice, it is usually small.

4.6.5 Almost-Concavity of the Reformulation

From the preceding results, we obtain the main theorem of this chapter. It re-expresses the Hessian matrix of $\Gamma(\mathbf{q})$ derived in Lemmas 4.13 and 4.14 as the sum of a negative-definite matrix $\mathbf{A}_{concave}$ and two correction terms. The condition on the price-sensitivity parameters is nearly identical to that of Theorem 3.5 and Corollary 3.6 ensuring concavity of the profit function in the special case of the NL model. It requires that the ratio of the price sensitivity parameters is bounded by two. For technical reasons, we require the bound to be strict here. In practice, the conclusion

[9] It is interesting to note that for the special case of the NL model where the no-purchase alternative is in a nest by itself, all the entries $S_{1,n+1},\ldots,S_{n,n+1}$ in the last column of $\mathbf{S}$ except for $S_{n+1,n+1}$ are zero. Then the error term reduces to $\frac{1}{\mu}\bar{\mathcal{K}}^{n+1}$, which also has special structure in this case. We considered such models in Section 3.8, and almost all of the existing results from the literature on pricing under models in the GEV family consider models of this form. Unfortunately this property is lost when multiple products are in the nest containing the outside alternative. The stochastic matrix $\mu\bar{\mathbf{S}}$ is dense for CNL models.

often holds even when the condition is violated, as was the case for the results of Chapter 3. Our computational experiments in Chapter 5 for the CNL model bear this out.

Theorem 4.16. *The Hessian of* $\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda\mu G(\mathbf{q})$ *is*

$$\frac{\partial^2 \Gamma}{\partial \mathbf{q} \partial \mathbf{q}} = \mathbf{L}\left(\mathbf{A}_{concave} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \mathcal{K} - \tilde{\mathbf{A}}\right)\mathbf{L}^{\top}.$$

If the price sensitivity parameters $b'_1, b'_2, \ldots, b'_n$ *belong to the range* $(\frac{1}{2}, 1)$, *then the matrix* $\mathbf{A}_{concave}$ *is negative definite. The matrix* $\tilde{\mathbf{A}}$ *is the upper-left* $n \times n$ *submatrix of the tensor product*

$$\frac{1}{\mu}\left(\mathbf{e} + (1 + \lambda\mu)\left(\frac{1}{S_{n+1,n+1}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right)\right) \cdot \bar{\mathcal{K}},$$

where the vector $\bar{\mathbf{S}}\mathbf{e}_{n+1}$ *denotes the last column of the matrix* $\bar{\mathbf{S}}$ *and* $S_{n+1,n+1}$ *is its last entry.*

Proof. See Appendix C.1.4. □

Concavity of the profit is equivalent to the matrix in the parentheses being negative definite, because $\mathbf{L}$ has full rank. Unfortunately, Theorem 4.16 cannot show global concavity of the objective function. The matrix $\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}$ is non-zero if the price sensitivity parameters are not all equal. (They can trivially be scaled if they are equal but different from one, as shown in Corollary 3.6 for the NL model.) The matrix $\tilde{\mathbf{A}}$ may have some (usually small) negative eigenvalues, and its impact may become large if λ is large. Like the matrix $\tilde{\mathbf{G}}$ in Lemma 4.15, it arises because we have fixed the price z_{n+1} of the outside alternative, and would vanish if this were not the case.

On the other hand, $-\frac{1}{\mu}\mathbf{e} \cdot \bar{\mathcal{K}} = -\bar{\mathbf{G}}$ is negative semi-definite when $\mu \geq 1$.[10] In the matrix $\tilde{\mathbf{A}}$ we have added an additional term to $\mathbf{e}$. As was the case for the error term $\tilde{\mathbf{G}}$ in Lemma 4.15, the new vector is the last column of stochastic $\mu\bar{\mathbf{S}}$, normalized by

[10]See Lemma 4.12 and definition 4.2. The matrix $\bar{\mathbf{G}}$ is clearly a diagonally dominant M-matrix when $\mu > 1$. It remains positive semi-definite and weakly diagonally dominant when $\mu = 1$. To see this, multiply by $\mathbf{Y}$ and apply Corollary 4.7 to show the row-sums are non-negative, as was done in Lemma 4.8 for the matrix $\mathbf{L}^{-1}$.

its last entry. This vector is alway positive, but we are unable to bound its entries in general.

In practice, the Hessian is almost always negative-definite when $\mathbf{B}$ satisfies the assumption. The entries of $\tilde{\mathbf{A}}$ are usually small in magnitude relative to those of $\mathbf{A}_{concave}$. Generally, the matrix in the parenthesis is not a negative M-matrix, but its non-positive off-diagonal elements are small, and it remains diagonally dominant.

For the case of NL models discussed in Chapter 3, the matrix $\tilde{\mathbf{A}}$ *is* negative semi-definite. The proofs of Theorem 3.5 and Theorem 4.16 bear some similarity and make the same assumptions on $\mathbf{B}$, but the former is dealing with the original profit function whereas we are analyzing a reformulation in this section. Concavity of Π is a stronger statement than concavity of $\Psi - \lambda\mu G$. It often does not hold for CNL models even when the transformed objective is concave. Concavity of the transformed objective implies only that Π is *quasi*-concave, as discussed in Section 4.6.

4.7 The Original Profit Function

Instead of defining $\mathbf{q}$ and Ψ, the derivations of Section 4.5 and Section 4.6 can be done with respect to the original, normalized demand function $p(\mathbf{z})$ and the original profit function $\Pi(\mathbf{p})$ as a function of $\mathbf{p}$. However, the expressions involved become significantly more complicated.

Nevertheless, the sensitivity of demand to prices is of interest. In this section, we recover the Jacobian of the demand, and state the Hessian of the profit, in terms of $\mathbf{p}$. We use the same notation as for the special case of NL demand models in Chapter 3, although we can only recover a closed form for the *inverse* of the Jacobian for the general case.

4.7.1 The Jacobian of the Demand

We define the analogue of the Jacobian matrix $\mathbf{LD}$ and write its inverse,

$$\mathbf{J_z} \triangleq \mathbf{JD} = \left[\frac{\partial z_j}{\partial p_i}\right]_{ij} \qquad \text{and} \qquad \mathbf{J_z^{-1}} = \mathbf{D}^{-1}\mathbf{J}^{-1} = \left[\frac{\partial p_j}{\partial z_i}\right]_{ij},$$

where the Jacobian between $\mathbf{y}$ and $\mathbf{p}$ is

$$\mathbf{J} = \left[\frac{\partial y_j}{\partial p_i}\right]_{ij}.$$

We consider the full (singular) Jacobian matrix of $\bar{\mathbf{p}}$ with respect to $\bar{\mathbf{y}}$, analogous to the (non-singular) matrix $\bar{\mathbf{L}}^{-1}$:

$$\bar{\mathbf{J}}^{-1} = \begin{bmatrix} & & & \vdots \\ & \mathbf{J}^{-1} & & \frac{\partial p_{n+1}}{\partial y_i} \\ & & & \vdots \\ \cdots & \frac{\partial p_j}{\partial y_{n+1}} & \cdots & \frac{\partial p_{n+1}}{\partial y_{n+1}} \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

This matrix is singular because the entries of each row must sum to zero so that $\mathbf{p}$ remains a probability vector, when some y_i is increased. Then $\bar{\mathbf{J}}^{-1}\mathbf{e} = \mathbf{0}$ and $\mathbf{e}$ is a zero-eigenvector. When $\mu = 1$, as is common, the submatrix $\mathbf{J}^{-1}$ is an M-matrix.

Proposition 4.17. *The Jacobians with respect to* $\mathbf{p}$ *and* $\mathbf{q}$ *are related by*

$$\bar{\mathbf{J}}^{-1} = \frac{1}{\mu G(\mathbf{y})}\bar{\mathbf{L}}^{-1} - \bar{\mathbf{Y}}^{-1}\bar{\mathbf{p}}\bar{\mathbf{p}}^\top.$$

If $\mu \geq 1$ *and* $\mathbf{y} > \mathbf{0}$, *then* $\mathbf{J}^{-1}$ *is a strictly row diagonally dominant* M*-matrix, and* $\bar{\mathbf{J}}^{-1}$ *is (weakly) row diagonally dominant.*

Proof. Computing the derivatives of $p_j = y_j G_j(\mathbf{y})/\mu G(\mathbf{y})$, we have, for $i \neq j$

$$\left[\bar{\mathbf{J}}^{-1}\right]_{ij} = \frac{y_j G_{ij}(\mathbf{y})}{\mu G(\mathbf{y})} - \frac{y_j G_i(\mathbf{y}) G_j(\mathbf{y})}{(\mu G(\mathbf{y}))^2} = \frac{y_j G_{ij}(\mathbf{y})}{\mu G(\mathbf{y})} - \frac{p_i p_j}{y_i}.$$

If $i = j$, we must also add the term $G_j(\mathbf{y})/\mu G(\mathbf{y}) = p_j/y_j$. More concisely,

$$\begin{aligned}\bar{\mathbf{J}}^{-1} &= \frac{1}{\mu G(\mathbf{y})}\mathbf{G}\bar{\mathbf{Y}} + \bar{\mathbf{Y}}^{-1}(\bar{\mathbf{p}} - \bar{\mathbf{p}}\bar{\mathbf{p}}^\top) \\ &= \frac{1}{\mu G(\mathbf{y})}\left(\mathbf{G}\bar{\mathbf{Y}} + \operatorname{diag}(\mathbf{g})\right) - \bar{\mathbf{Y}}^{-1}\bar{\mathbf{p}}\bar{\mathbf{p}}^\top \\ &= \frac{1}{\mu G(\mathbf{y})}\bar{\mathbf{L}}^{-1} - \bar{\mathbf{Y}}^{-1}\bar{\mathbf{p}}\bar{\mathbf{p}}^\top,\end{aligned}$$

as claimed. Referring to the proof of Lemma 4.8, we have that $\bar{\mathbf{L}}^{-1}\mathbf{e} = \mu\bar{\mathbf{g}} \geq 0$ and that the off-diagonal entries of $\bar{\mathbf{L}}^{-1}$ are non-positive. The row sums are

$$\bar{\mathbf{J}}^{-1}\mathbf{e} = \frac{1}{\mu G(\mathbf{y})}\mu\bar{\mathbf{g}} + \bar{\mathbf{Y}}^{-1}\bar{\mathbf{p}}(\bar{\mathbf{p}}^{\top}\mathbf{e}) = \frac{1}{\mu G(\mathbf{y})}\left(\mu\bar{\mathbf{g}} - \bar{\mathbf{g}}\right) \geq 0,$$

because $\mu \geq 1$ by assumption. Therefore each diagonal element of $\mathbf{J}^{-1}$ must be non-negative and at least as large in magnitude as the sum of the off-diagonal elements in its row. Thus, $\bar{\mathbf{J}}^{-1}$ is weakly diagonally dominant. By Assumption 4.3, $\mathbf{p} > 0$ and the off-diagonal entries are negative. Removing the last row and column yields a strictly diagonally dominant M-matrix. This completes the proof. □

By the chain rule, the Jacobian of $\mathbf{p}$ with respect to $\mathbf{z}$ is

$$\mathbf{J}_{\mathbf{z}}^{-1} = \mathbf{D}^{-1}\mathbf{J}^{-1} = -\mathbf{B}\mathbf{Y}\mathbf{J}^{-1}.$$

Then the Jacobian matrix $-\mathbf{J}_{\mathbf{z}}^{-1}$ of the *negated* demand function is a positive diagonal scaling of an M-matrix, and therefore it is also an M-matrix [43, Theorem 1.2.3]. As such, it satisfies the requirements of both versions of the Gale-Nikaido Univalence Theorem and its generalizations [30, 54, 8]. That is, the negated demand is an injective (one-to-one) function. We have already proved a stronger result in Theorem 4.9, which also establishes surjectivity of the demand function onto the interior of the simplex and does not require that $\mu \geq 1$.

4.7.2 The Hessian of the Profit

Deriving an expression for the Hessian of the profit in terms of $\mathbf{p}$ is complicated because the second partial derivatives of $\mathbf{p}$ with respect to $\mathbf{y}$ (the analogue to $\mathcal{K}$) involve many terms. We only state the final result,

$$\frac{\partial^2\Pi}{\partial\mathbf{p}\partial\mathbf{p}} = \mathbf{J}\left(\mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} - \mathbf{J}\mathbf{D}\mathbf{p}\cdot\mathcal{K} - \mathbf{D}\,\mathrm{diag}\left(\mathbf{g}\right) + \mathbf{p}^{\top}\mathbf{J}\mathbf{D}\mathbf{p}\right)\mathbf{J}^{\top}, \qquad (4.5)$$

where we use both Jacobian matrices $\mathbf{L}$ and $\mathbf{J}$. We compare this expression to that of Lemma 4.14. The quantity in the brackets is identical except for the last term, and the substitution of $\mathbf{LDq} \cdot \mathcal{K}$ by $\mathbf{JDp} \cdot \mathcal{K}$. However, in light of the discussion in Section 4.6, the function Ψ may be concave at a given point while Π is not concave.

Our computational experiments in Chapter 5 give rise to some such instances[11]. Briefly, for large ratios between the price-sensitivity parameters, Π is not always concave for CNL models. However, when the parameters $\mu_1, \ldots, \mu_M$ of the CNL model take on moderate values, the function Ψ does remain concave. In our experiments, both functions Π and Ψ do in fact remain concave when the condition on the price sensitivity parameters is satisfied (their ratios are less than two).

Our practical observations and experimental results show that the condition of Theorem 4.16 is far from necessary for concavity. The theorem relies on characterizations of negative M-matrices to show negative-definiteness of the Hessian, but the negative M-matrix property is not a necessary condition for negative-definiteness. Thus even if it were possible to construct instances that violate this property, the objective function might remain concave. (It often does in practice.) Moreover, concavity of Ψ is not a necessary condition for the optimal solution of the pricing problem to be unique. We explore different sufficient conditions for uniqueness of the optimal solution in Chapter 5.

We already derived a closed-form expression for the Hessian of the profit under NL models in Chapter 3 and showed that it is negative definite under the condition of Theorem 4.16. Our results in this chapter suggest that a similar result does not hold for GEV models in general. However, the experimental results we just mentioned only imply that it does not hold when the condition on the price sensitivity parameters is already violated. Under what circumstances Π is concave remains an open question. Nevertheless, the fractional programming reformulation of Section 4.6 suggests that the profit Π is *almost* quasi-concave in terms of the market-shares, becauze Ψ is *almost* concave.

[11]See Table 5.1 in Section 5.8.

4.8 Conclusions

In this chapter, we have developed a formulation for maximizing the profit when pricing under general GEV models over the space of market shares. We showed that the formulation is nearly concave in terms of the unnormalized choice probabilities, if the price sensitivity parameters for each product are sufficiently close to one another. More importantly, we have shown that the profit is in fact well-defined over of the market shares in that the demand function is invertible. We conjecture that the optimal prices often remain unique even when the reformulated objective function is not concave.

We have also stated the Jacobian and Hessian matrices of the original profit function in terms of market shares under general GEV models. The near-concavity of our reformulation suggests that the profit function itself is nearly quasi-concave.

Chapter 5

Practical Pricing under GEV and MMNL Models

5.1 Introduction

In this chapter, we propose a first-order nonlinear optimization method to solve the pricing problem when customer demand is represented by a model from the *generalized extreme value* (GEV) family, or by the *mixed logit* (MMNL) demand model. A seller selects the prices $x_i, x_2, \ldots, x_n \in \mathbb{R}$ of n differentiated substitutable products. We aim to maximize the profit,

$$\max_{\mathbf{x} \in \mathbb{R}^n} \left\{ \Pi(\mathbf{x}) = \sum_{i=1}^{n} x_i p_i(\mathbf{x}) \right\}, \tag{5.1}$$

where the function $p_i(\mathbf{x})$ denotes the demand for product i. Both demand models can be adjusted to take into account different profit margins or production costs for each product: the reparameterization in Section 3.5.1 generalizes in a straightforward manner.

The profit is not a (quasi-)concave function in the vector $\mathbf{x}$ even under the basic *multinomial logit* (MNL) model (see Appendix A.2). Under MMNL models, the profit is the sum of the profits arising under the individual MNL models in a mixture, and it may have numerous local maxima [40].

However, in the preceding chapters we have shown that the local maximum of the profit is often unique under members of the GEV family of models. MMNL models do not belong to this family, but Hanson and Martin [40] devised an algorithm for finding "good" local optima under MMNL models. In practice, their method often appears to find the global maximum. Motivated by these two facts, we propose a simple, practical algorithm for maximizing the profit under both classes of models.

Leveraging our results from Chapter 4, we begin by analyzing the necessary first-order optimality conditions (FOCs) of problem (5.1) for GEV models, when it is expressed as an optimization problem in terms of the (unnormalized) demands. We argue that the FOCs often define a *nonlinear (generalized) diagonally dominant* function. The solution of such systems of equations is unique, and can be computed using the nonlinear Jacobi or Gauss-Seidel method [62, 28]. Unfortunately, this requires finding the root of a nonlinear equation for each variable at every step. Each solution in turn requires multiple costly evaluations of the partial derivatives of the prices $\mathbf{x}$ (with respect to the demands).

On the other hand, we give closed-form expressions for the partial derivatives of the demands $p(\mathbf{x})$ with respect to the prices $\mathbf{x}$. We use their properties to show that a linearization of the FOCs can be solved with *linear* Jacobi iterations. The iteration matrix obtained is *sub-stochastic* for any step size $\alpha \leq 1$. This suggests rapid convergence, though we observe empirically that the step size must be reduced somewhat to ensure convergence despite the linearization. The iteration matrix on each step of our algorithm is computed rapidly in closed form. We do not require costly the costly numerical matrix inversions needed to apply the nonlinear Jacobi or Gauss-Seidel method mentioned above.

Our general algorithm for GEV models can in particular be used to maximize the profit under the *cross-nested logit* (CNL) demand model. The CNL model is a member of the GEV family, but appears similar to the MMNL model in that a number of nests may represent different segments of a population, who are all offered the same set of products. Within each nest of both the CNL and MMNL, customers choose according to an MNL model. The difference is that an arriving customer belongs to each nest with a *fixed* probability under the MMNL, whereas this probability also depends on the prices in the CNL. Neither of these models is, strictly speaking, more general than the other.[1]

We show that MMNL models can be approximated *locally* by CNL models with not only the same choice probabilities in each nest, but also with arbitrarily close substitution behavior. That is, the partial derivatives of the demands become arbitrarily close under the two models as we let the parameters of the CNL approximation become large.

This motivates us to apply our algorithm to MMNL models directly. There is no need to explictly construct the CNL approximation in practice, since the required partial derivatives can be computed directly from the MMNL model. We observe similar convergence as for CNL models.

Unfortunately, convergence for MMNL models is only to one of the often many local maxima. Combining our algorithm with the path-following method of Hanson and Martin [40] finds a good local maximum in roughly the same time it takes to solve the CNL pricing problem. Based on our experiments, the maximum found appears to be globally optimal as long as the step size is chosen to be sufficiently small. We observe fast convergence even for extreme values of the demand model parameters.

In Section 5.2, we review the literature on pricing with multiple customer segments. We discuss the optimality condittions under GEV models in Section 5.4, and present the nonlinear and linearized Jacobi algorithms in Sections 5.5 and 5.6, respectively. We then present the CNL and MMNL models and discuss their relationship in Section 5.7. Our experimental results are presented in Section 5.8.

[1] Mcfadden and Train [58] show that a MMNL model with a large enough number of segments ("nests") can approximate any GEV model. However, a very large number of segments could potentially be required. We will show that the reverse, approximating a MMNL model with a CNL model, can be accomplished *locally* without increasing the number of segments.

5.2 Literature Review: Pricing under CNL and MMNL Models

The pricing problem under the MMNL model was studied by Hanson and Martin [40]. They propose the path-following method that we incorporate into our algorithm, but they suggested using conjugate-gradient or second-order methods to maximize $\Pi(\mathbf{x})$ as a function of the prices $\mathbf{x}$. Because the profit is not concave, the Hessian is relatively costly to compute, and these methods require line searches at each step, they are slow to converge and they are prone to finding the nearest local maximum to the starting point. We propose a method that requires no line search and uses only first-order information. It appears less prone to get stuck in local minima.

To our knowledge, there has been no work on pricing under the CNL model except for the slight generalizations of the NL model discussed in Section 4.3. Work on NL models, where each product is only available to one customer segment, and on single-segment MNL models is discussed in the preceding chapters.

We argue that the solution to the FOCs often appears to be unique on the basis of the nonlinear Jacobi and Gauss-Seidel methods. These algorithms are straighforward generalizations of the linear Jacobi and Gauss-Seidel methods for solving linear systmes of equations. Each equations in a *nonlinear* system is solved for one variable at each step, keeping the other variables fixed. Depending on the method, some or all of the variables are updated between successive steps. The convergence of theses methods for diagonally-dominant functions was established by More [62]. It was extended to generalized diagonally dominant functions by Frommer [28], who conjectured that functions with generalized diagonally dominant *Jacobian matrices* belonged to this class. The conjecture was verified by Gan et al. [37].

The first-order method we propose defines a non-linear operator for which the algorithm converges to a fixed point. We defer our discussion of approaches for showing that a given nonlinear operator defines a contraction mapping to Section 5.6.1.

5.3 Outline of the Proofs

In Section 5.4, we state the FOCs of the reformulated pricing problem under GEV models from Chapter 4. We show that they are equivalent to the FOCs of the original pricing problem for an appropriate value of the parameter λ (Proposition 5.2). Lemma 5.1 presents the FOCs in two different but equivalent ways. The first statement of the FOCs is explored in Section 5.5. Lemma 5.3 computes the Jacobian matrix of the FOCs to explain why they often appear to define a *(generalized) diagonally dominant function* of the prices. Theorem 5.4 states that, if this is the case, then the FOCs admit a unique solution, which necessarily corresponds to the global maximum. In fact, this unique solution can in principle be computed with the well known nonlinear Jacobi and Gauss-Seidel methods mentioned above, by repeatedly solving the one-dimensional equations obtained by fixing all but one variable at a time. Unfortuntately, such an approach may be computationally inefficient because it requires the inversion of large matrices at each step.

Instead, in Section 5.6, we use the second statement of the FOCs from Lemma 5.1 to develop a new algorithm based on *linear* Jacobi iterations for solving systems of linear equations. Lemma 5.5 states the linear Jacobi iteration for the FOCs when an important quantity (the matrix $\mathbf{S}$) is held fixed, and shows that it converges to the solution of the system obtained in this way. It forms the basis of our algorithm, which we present in Section 5.6.3. Lemma 5.6 states the iteration used in our algorithm, where the matrix $\mathbf{S}$ is updated based on the value of the current iterate. It further shows that the *iteration matrix* at each step (denoted by $\mathbf{M}_k$) is strictly sub-stochastic. This fact explains the rapid convergence of the algorithm in our computational experiments. Theorem 5.7 verifies that the optimal solution of the GEV pricing problem is indeed a fixed point of the algorithm. Lemma 5.8 briefly states a slightly modified iteration that appears somewhat simpler, and shows that the iteration matrix is also sub-stochastic, for an appropriate step size. Although we have ultimately been unable to show any stronger theoretical results for this version of our algorithm, the simpler form suggests that it may be more amenable to further analysis. Specifically, we conjecture that both of the iterations we propose define nonlinear operators with a unique fixed point. If this is the case, the fixed point corresponds to the global optimum of the GEV pricing problem.

In Section 5.7, we formally define the CNL and MMNL models, and discuss their relationship. Lemmas 5.9 and 5.10 give simplified expressions for the Jacobian matrices of the choice probabilities with respect to the prices under the two models. These matrices are required to implement our algorithm. Theorem 5.11 uses the similar structure of the two Jacobian matrices to show that a CNL model can locally approximate a MMNL model, in that both models posess the same choice probabilities, the same conditional choice probabilities and arbitrarily close Jacobian matrices for a vector of prices. We end the section by explaining how, together with a known path-following heuristic [40], our algorithm can be applied to solve the pricing problem under MMNL models.

In Section 5.8, we present experimental results that explore the performance of our algorithm. We also empirically test the properties of the pricing problem reformulation from Chapter 4 when the assumptions of Theorem 4.16 are violated. It turns out that the concavity properties discussed earlier in this thesis appear to persist even for relatively extreme values of the parameters. The (generalized) diagonal dominance property persists for even more extreme values of the parameters.

5.4 The First-order Optimality Conditions

In this section, we state the FOCs for the pricing problem under GEV demand models. Our approach is based on the results in Chapter 4. We refer the reader to that chapter for details about the reformulated pricing problem, parameterized by λ, which we restate[2] here:

$$\max_{\mathbf{q}>\mathbf{0}} \{F(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda G(y(\mathbf{q}))\} \tag{5.2}$$

[2]We previously defined the objective as $\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \mu\lambda G(y(\mathbf{q}))$. Because the constant $\mu > 0$ can be absorbed into the parameter $\lambda > 0$ without loss of generality, we can omit it from the objective of (5.2) in this chapter. In fact, it is often chosen to be $\mu = 1$ in MNL, NL and CNL models.

The optimization is over the vector of *unnormalized demands* $\mathbf{q} = \mu G(y) \cdot \mathbf{p}$, and we refer to Definition 4.2 of the *GEV generating function* $G(\mathbf{y})$, with $\mathbf{y} > \mathbf{0}$.

The current chapter focuses on the sub-stochastic matrix $\mathbf{S}$ introduced in Section 4.6.4 to analyze the gradient and Hessian matrix of $\Gamma(\mathbf{q})$. Recall that its inverse is expressed in terms of the first and second partial derivatives of $G(\mathbf{y})$ as

$$\mathbf{S}^{-1} = (\mathbf{L}\,\mathrm{diag}\,(\mathbf{g}))^{-1} = \mathrm{diag}\,(\mathbf{g})^{-1}\,\mathbf{L}^{-1} = \left(\mathbf{I} + \mathrm{diag}\,(\mathbf{g})^{-1}\,\mathbf{G}\mathbf{Y}\right), \tag{5.3}$$

where the matrices $\mathrm{diag}\,(\mathbf{g})^{-1}$ and $\mathbf{Y} = \mathrm{diag}\,(\mathbf{y})$ are positive diagonal, and where $\mathbf{G}$ has positive diagonal and non-positive off-diagonal entries by definition. The matrix $\mathbf{L}$ is the Jacobian matrix of the vector $\mathbf{y}$ with respect to the unnormalized demands $\mathbf{q}$, and can also be expressed in terms of the partial derivatives of $G(\mathbf{y})$ (see Section 4.5.1). Our new results follow from Lemma 4.15, which established that the matrix $\mathbf{S}$ is a sub-stochastic inverse-M-matrix. We no longer require the characterization of the second partial derivatives of the *profit* developed earlier, except briefly in Lemma 5.3 below.

Recall that $\mathbf{z}$ is an affine transformation of the original price vector $\mathbf{x}$ to allow for varying profit margins and production costs accross products (see Section 3.5.1). Its components are related to those of $\mathbf{x}$ in terms of the profit margins a_i and price sensitivity parameters b_i specific to each product i by

$$z_i \triangleq a_i(x_i - c_i), \qquad i = 1, \ldots, n.$$

The transformation does not affect the results in this chapter at all. We could, without loss of generality, assume that the production costs are zero and that the profit margins are the same across products. Then, $\mathbf{x} = \mathbf{z}$. We continue to denote the prices by $\mathbf{z}$ only for consistency. In Theorem 4.9, we showed that the mapping between $\mathbf{z}$ and $\mathbf{q}$ is invertible. The same holds for the mappings between $\mathbf{y}$ and $\mathbf{q}$, and between $\mathbf{y}$ and $\mathbf{p}$.

We begin with the transformed problem (5.2). The FOCs can be stated in terms of $\mathbf{z}$, the matrix $\mathbf{S}$ and a vector $\mathbf{w}$ that depends only on the scalar parameter λ.

Lemma 5.1. *The stationary points of* $F(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda G(y(\mathbf{q}))$ *are the solutions of*

$$F'(\mathbf{z}) = \mathbf{z} - \mathbf{S}\mathbf{w} = \mathbf{0}$$

and, equivalently,

$$\mathbf{S}^{-1}\mathbf{z} - \mathbf{w} = \mathbf{0}, \tag{5.4}$$

where the vector $\mathbf{w} \triangleq (\mathbf{B}^{-1} + \lambda\mathbf{I})\mathbf{e}$ *is constant.*

Proof. The first-order optimality conditions follow from setting the derivatives to zero. Lemma 4.13 gives the gradient of Ψ, and the gradient of $G(y(\mathbf{q}))$ is obtained by applying the chain rule with the Jacobian matrix $\mathbf{L}$.

$$\frac{\partial \Psi}{\partial \mathbf{q}} - \lambda \frac{\partial G}{\partial \mathbf{q}} = \mathbf{z} + \mathbf{L}\mathbf{D}\mathbf{q} - \lambda\mathbf{L}\mathbf{g} = \mathbf{z} - \mathbf{S}\mathbf{B}^{-1}\mathbf{e} - \lambda\mathbf{S}\mathbf{e} = \mathbf{z} - \mathbf{S}(\mathbf{B}^{-1} + \lambda\mathbf{I})\mathbf{e} = \mathbf{0}$$

The equivalence of equation (5.4) follows becase $\mathbf{S}^{-1}$ is an M-matrix and thus has non-zero eigenvalues. We can therefore multiply the first statement through by $\mathbf{S}^{-1}$. □

We will consider two types of algorithms for computing a stationary point $\mathbf{z}$ of $F(\cdot)$, and argue that $\mathbf{z}$ is likely the *unique* stationary point. If that is the case, then it is also a global maximizer of $F(\cdot)$.

By Theorem 4.10, if $\lambda = \lambda^* = \Pi(\mathbf{z}^*)$ is a global maximum of the profit, then a global maximizer of $F(\cdot)$ is also a global maximizer of $\Pi(\cdot)$. The maximum value of $F(\cdot)$ is $F(\mathbf{z}^*) = 0$. For $\lambda > \lambda^*$, the maximum of $F(\cdot)$ is negative. For $\lambda < \lambda^*$, it is positive. This follows since F is clearly decreasing in λ, uniformly over all values of $\mathbf{y}$.

Of course, λ^* is not known *a priori*. In principle, we can perform a line search over values of λ, maximizing $F(\cdot)$ and checking whether the resulting maximum $F(\mathbf{z})$ is positive, for each value of λ. This is the approach one would use for the first class of algorithms that we discuss in Section 5.5.

However, for the iterative algorithm that we will present later in Section 5.6, it turns out that updating the value of λ with the profit $\Pi(\mathbf{z})$ at the current point $\mathbf{z}$ is sufficient. In fact, substituting $\lambda = \Pi(\mathbf{z}) = \mathbf{p}^\top\mathbf{z}$ into the expression for $F'(\mathbf{z})$ in

Lemma 5.1 yields the first-order optimality conditions for the original problem of maximizing $\Pi(\mathbf{z})$. Therefore, our algorithm in Section 5.6 can also be seen as directly solving the first-order optimality conditions of the original, un-transformed problem.

Proposition 5.2. *Under a GEV demand model, setting* $\lambda = \Pi(\mathbf{z}) = \mathbf{p}^\top \mathbf{z}$*, the FOCs of the parametric reformulation in* (5.4) *are equivalent to the FOCs*

$$\frac{\partial \Pi}{\partial \mathbf{z}} = -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \mathbf{w}\right) = \mathbf{0}$$

of the unconstrained pricing problem (5.1).

Proof. See Appendix D.2. □

5.5 Nonlinear Jacobi Iterations

We first argue that $F'(\mathbf{z}) = \mathbf{0}$ is likely to have a unique solution because $F'(\cdot)$ often belongs to the class of *diagonally dominant functions.* We will apply two results from the literature guaranteeing that the stationary point of $F(\cdot)$ is unique when this is the case. To this end, we compute the Jacobian matrix of F' with respect to $\mathbf{z}$ in the next lemma. It is related to the Hessian matrix of F with respect to $\mathbf{q}$ computed in Chapter 4, but different since the second differentiation is now with respect to the vector $\mathbf{z}$ of prices rather than with respect to the vector $\mathbf{q}$ of un-normalized demands.

Lemma 5.3. *The Jacobian of* $F'(\mathbf{z})$ *is*

$$\frac{\partial F'}{\partial \mathbf{z}} = \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^\top \mathbf{BL}^\top - \left(\tilde{\mathbf{L}}^{-1}\right)^\top \mathbf{L}^\top - \tilde{\mathbf{E}},$$

with the last term

$$\tilde{\mathbf{E}} = \mathbf{BY}\left(\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathrm{e} \cdot \mathcal{K} - (1 + \lambda\mu)\tilde{\mathbf{G}}\right)\mathbf{L}^\top,$$

and where the matrix

$$\left(\tilde{\mathbf{L}}^{-1}\right)^\top = \left(\mathrm{diag}\left(\mathbf{g}\right) + \mathbf{BYG}\right)$$

is equal to $(\mathbf{L}^{-1})^\top$ *when* $\mathbf{B} = \mathbf{I}$.

Proof. See Appendix D.1. $\square$

Consider the Jacobian matrix in Lemma 5.3 when all the price sensitivity parameters are equal, that is, when $\mathbf{B} = \mathbf{I}$. Then the second and third term are both equal to the identity matrix, and it reduces to

$$\frac{\partial F'}{\partial \mathbf{z}} = \mathbf{I} - \tilde{\mathbf{E}} = \mathbf{I} + (1 + \lambda\mu)\mathbf{Y}\tilde{\mathbf{G}}\mathbf{L}^\top$$

The remaining term $\tilde{\mathbf{E}}$ is similar to the non-negative-definite term $\tilde{\mathbf{A}}$ in Theorem 4.16. (There, we expanded the matrix $\tilde{\mathbf{G}}$ using Lemma 4.15.) Like the case there, the off-diagonal entries of $\tilde{\mathbf{E}}$ are often small in practice. Clearly, if the off-diagonal entries of $\tilde{\mathbf{E}}$ are small, then the Jacobian is diagonally dominant. This is often the case in our experiments. Generalized diagonal dominance is a weaker condition which only requires the matrix

$$\left(\frac{\partial F'}{\partial \mathbf{z}}\right)\mathbf{U}$$

to be row diagonally dominant, for some diagonal matrix $\mathbf{U}$. Testing for this property is non-trivial, but we often observe in practice that the matrix happens to be diagonally dominant with the specific choice $\mathbf{U} = \operatorname{diag}(\mathbf{q})$. Iterative algorithms to find a suitable scaling $\mathbf{U}$ can be found in the literature and are employed in our computational experiments on random instances [49]. We observed that generalized diagonal dominance appears to persist even when the price sensitivity parameters vary greatly across products ($\mathbf{B}$ is very different from $\mathbf{I}$), and even in many cases where the entries of $\tilde{\mathbf{E}}$ become large.

We apply two results from the literature guaranteeing that the stationary point of $F(\cdot)$ is unique, if either the diagonal dominance or the generalized diagonal dominance property holds.

Theorem 5.4. *If the Jacobian of* F' *is a (generalized) diagonally dominant matrix,* F' *is injective and* $F(\mathbf{z}) = \mathbf{0}$ *has at most one solution.*

Proof. The domain of F is clearly convex. If the Jacobian is (generalized) diagonally dominant, then F' is a nonlinear (generalized) diagonally dominant function [62, 37]. As such, it is injective[62, 28]. In particular, there is at most one value of $\mathbf{z}$ for which $F(\mathbf{z}) = \mathbf{0}$. □

Nonlinear generalized diagonally dominant equations can be solved using the well-known nonlinear Jacobi and Gauss-Seidel methods discussed at the end of the literature review (Section 5.2). However, applying them in practice can be computationally intensive. Evaluating F and F' requires the computation of $\mathbf{S}$. Except for special cases like the NL model, this would be accomplished by evaluating $\mathbf{S}^{-1}$ and then computing its inverse. Unless the matrix is sparse or well structured, the inversion is costly. (For example, under the CNL models discussed in Section 5.7.1, the matrix is sparse if the parameters α_{im} are mostly zero. This implies that the nesting structure is also sparse, in that products belong to few nests. If the number of nests is small, then the inversion can be accelerated using the Woodbury matrix identity [38].) It is not practical to computationally invert dense matrices of dimension greater than a few hundred at each iteration of an algorithm. To handle models with such a large number of products, we turn our attention to the equivalent statement of the first-order optimality conditions in terms of $\mathbf{S}^{-1}$ in the next section.

5.6 A Simple Iterative Algorithm

In this section, we propose an iterative algorithm to solve the FOCs (5.4). Although it is easy to verify that the resulting stationary point is a *local* maximum in practice (on the basis of the Hessian, say), if the condition of Theorem 5.4 is satisfied, the stationary point is also the global maximum. Specifically, we propose an iteration of the form

$$\mathbf{z}^{t+1} = T(\mathbf{z}^t)$$

where $T(\cdot)$ is a nonlinear operator that we fully define below. There has been much work on establishing the convergence of such procedures to a fixed point $\mathbf{z}^*$ such that

$T(\mathbf{z}^*) = \mathbf{z}^*$, and showing that the fixed point is unique. Usually one proves that $T(\cdot)$ or a related operator defines a contraction mapping in some norm. Unfortunately, we are unable to determine whether this is the case for our algorithm.

Nevertheless, we will show at the end of the present section, that (a slightly modified version of) our operator takes the particular form

$$\mathbf{z}^{t+1} = T(\mathbf{z}^t) = \mathbf{R}_t \mathbf{z}_t + \mathbf{w}$$

when solving the FOCs stated in Lemma 5.1. The *iteration matrix* $\mathbf{R}_t$ on the t^{th} step, which depends on $\mathbf{z}_t$ through $\mathbf{S}^{-1}$, is *strictly sub-stochastic.* The sub-stochasticity of $\mathbf{R}_t$ follows because $\mathbf{S}^{-1}$ is a strictly diagonally dominant M-matrix. We have already used this fact, in a quite different way, to establish that $\mathbf{S}$ itself is also sub-stochastic.

Since the optimal value of λ can be found via a line search, showing that $T(\cdot)$ converges to a unique fixed point for constant $\mathbf{w}$ would be sufficient to show that the optimal solution to the pricing problem is unique. Before proceeding to develop our algorithm, we outline some approaches that we have pursued on the basis of this observation.

5.6.1 Literature Review: Contraction Mappings

Uniqueness of the fixed point has been established for a number of related problems. We briefly mention some techniques that we have explored and provide relevant references in the literature.

First, note that the iterate $\mathbf{z}^t$ obtained on the t^{th} iteration results from the product of a sequence of t stochastic matrices. Such products are well-studied and are known to converge rapidly under quite general conditions. Convergence is usually much faster than suggested by the *spectral radius*[3] of the iteration matrix (denoted $\mathbf{R}_t$ in the case discussed above), and can be qualified in terms of various *coefficients of*

[3]The spectral radius of a matrix is the norm of its (real or complex) eigenvalues. For square, substochastic matrices, it is necessarily less than one. It turns out that the spectral radius is submultiplicative, and can be used to characterize the convergence of powers and products of matrices. See the cited papers for details.

ergodicity [70, 42].

If $\mathbf{R}_t$ were constant rather than depending on $\mathbf{z}_t$, the operator $T(\cdot)$ would correspond to the stochastic dynamic programming operator with transition matrix $\mathbf{R}_t$ and cost vector $\mathbf{w}$. This operator can be shown to converge to a fixed point via a contraction mapping in the *span semi-norm* [10]. In fact, the usual proof technique for showing uniqueness and existence of a fixed point is to establish that $\mathbf{T}$ is a *contraction mapping* in an appropriate norm, meaning that applying the operator to two points reduces the distance between them. Ćirić [18] discusses a generalization of this concept called a *quasi-contraction mapping.* It is also sufficient to show that applying the operator a constant number of times reduces the distance (*an n-step contraction mapping*), since this simply defined a new operator. One particularly well-known contraction-mapping result is the Perron-Frobenius theorem for homogeneous linear operators, including square matrices with positive entries [45].

There has been work on extending the Perron-Frobenius theorem to non-linear homogeneous operators [46, 80, 44, 48]. Although $T(\cdot)$ is clearly not homogeneous in our case, there is a bijection between the prices $\mathbf{z}$ and the vectors $\mathbf{y}$ occurring in the definion of GEV models (see Section 4.2). We can re-express $T(\cdot)$ in terms of the vector $\bar{\mathbf{y}}$ that includes the attraction of the outside alternative. The demand model (and therefore all the quantities involved in our analysis) are invariant to scaling of $\bar{\mathbf{y}}$, and the re-expressed operator *is* homogeneous. The most common proof of the Perron-Frobenius theorem and the proof of its nonlinear extensions rely on showing a contraction mapping in *Hilbert's projective metric* [45]. It turns out that this metric on the the space $\mathcal{R}_+^{n+1}$ containing $\bar{\mathbf{y}}$ corresponds to the metric induced by the span semi-norm on $\mathcal{R}^n$ containing $\mathbf{z}$.

We have not been able to determine whether the nonlinear operator $T(\cdot)$ is always a quasi-contraction in either of these norms, but we have observed that it need not always be a contraction mapping for CNL models. While the results on the convergence of products of matrices suggest that our algorithm converges quickly, we have not been able to show that they imply that even multiple iterations $T^n(\cdot)$ of the operator yield a contraction mapping. Nevertheless, all these results motivate our belief that the fixed point is often, if not always, unique. They also help explain why our algorithm converges rapidly in computational experiments (see Section 5.8).

5.6.2 The Jacobi Iteration for Linear Systems

Although the nonlinear Jacobi method discuseed in Section 5.5 is impractical, we can use *linear* Jacobi iterations to define the operator $T(\cdot)$ mentioned above. For a fixed value of λ and a *fixed* matrix $\mathbf{S}^{-1}$, the system (5.4) is a strictly diagonally dominant system of *linear* equations. The Jacobi and Gauss-Seidel methods are known to converge for such linear systems. These methods depend on a *splitting* of the matrix $\mathbf{S}^{-1}$ into a diagonal part $\mathbf{E}$ and a full matrix $\mathbf{F}$ such that

$$\mathbf{S}^{-1} = \mathbf{E} + \mathbf{F}.$$

In the linear Jacobi method, one simply lets $\mathbf{E}$ be equal to the diagonal part of $\mathbf{S}^{-1}$. Then $\mathbf{F}$ has zeros on the diagonal. The solution of (5.4) satisfies

$$\begin{aligned} \mathbf{S}^{-1}\mathbf{z} - \mathbf{w} &= \mathbf{0} \\ (\mathbf{E} + \mathbf{F})\mathbf{z} - \mathbf{w} &= \mathbf{0} \\ \mathbf{E}\mathbf{z} &= \mathbf{w} - \mathbf{F}\mathbf{z} \\ \mathbf{z} &= \mathbf{E}^{-1}(\mathbf{w} - \mathbf{F}\mathbf{z}) \end{aligned} \tag{5.5}$$

It is then easy to compute the diagonal matrix $\mathbf{E}^{-1}$ and update $\mathbf{z}$ according to the last equation on each iteration. The convergence rate is often improved by picking a step size less than one, thereby only doing a partial update at each iteration. We apply the following well-known convergence result when $\mathbf{S}^{-1}$ is fixed before presenting our algorithm for solving the original nonlinear system.

Lemma 5.5. *Fix* $\mathbf{S}^{-1}$. *Define the splitting* $(\mathbf{E}, \mathbf{F})$ *where the matrix* $\mathbf{E}$ *is positive diagonal and* $\mathbf{F}$ *has non-positive off-diagonal elements and zeros on the diagonal,* *such that* $\mathbf{S}^{-1} = \mathbf{E} + \mathbf{F}$. *Then the damped Jacobi iteration, with* $0 < \alpha \leq 1$,

$$\mathbf{z}^{t+1} = (1-\alpha)\mathbf{z}^t + \alpha\mathbf{E}^{-1}\left(\mathbf{w} - \mathbf{F}\mathbf{z}^t\right) = \mathbf{z}^t + \alpha\mathbf{E}^{-1}\left(\mathbf{w} - \mathbf{S}^{-1}\mathbf{z}^t\right)$$

converges to the unique solution of the linear system (5.4).

Proof. In Lemma 4.8 we showed that $\mathbf{L}^{-1}$ is a strictly row-diagonally-dominant M-matrix. The matrix $\mathbf{S}^{-1} = \operatorname{diag}(\mathbf{g})^{-1}\mathbf{L}^{-1}$ is a strictly row-diagonally dominant M-matrix by Lemma 4.8, because we have only scaled the rows of $\mathbf{L}^{-1}$ by a positive diagonal matrix (see Lemma 4.15 or [43, Theorem 1.2.3]). As such, its diagonal $\mathbf{E}$ is positive and $\mathbf{F}$ has non-positive entries. The convergence of the Jacobi method for any diagonally dominant systems of equation is well known [6, and references therein]. □

5.6.3 A Linearized Jacobi Iteration for the FOCs

Our approach essentially consists in applying linear Jacobi iterations to the nonlinear system (5.4). We can solve the linear system of the preceding section for a given value of $\mathbf{S}^{-1}$, say $\mathbf{S}_k^{-1}$, corresponding to the iterate $\mathbf{z}^k$. Then we could update the value of $\mathbf{S}^{-1}$ based on the new value of $\mathbf{z}^{k+L}$ after a suffciently large number of steps L to ensure appropriate convergence. Instead, we prefer to update $\mathbf{S}_k^{-1}$ more frequently, or even at each iteration (in which case $t = k$ always). In a similar manner, rather than explicitly perform a line search over λ as discussed above, we can update its value periodically before convergence is achieved.

We propose the following algorithm, parameterized by an integer $L \geq 1$ and a step size $0 < \alpha \leq 1$:

1. Choose $\mathbf{z}^0 \in \mathbb{R}^n$. Set $k = 0$.

2. Let $\lambda = \Pi(\mathbf{z}^k) = (\mathbf{p}^k)^\top \mathbf{z}^k$.

3. Set $\mathbf{S}_k^{-1} = \mathbf{S}^{-1}$ corresponding to z^k.

4. For $t = k+1, \ldots, k+L$, update z^t according to (5.6) below. That is, let

$$\mathbf{z}^{t+1} = \mathbf{z}^t + \alpha \mathbf{E}_k^{-1}\left(\mathbf{w} - \mathbf{S}_k^{-1}\mathbf{z}^t\right).$$

5. If $\{z^t\}$ has converged, stop. Otherwise set k to $k+L$ and go to Step 2.

The starting point in Step 1 is not important, but Theorem 5.7 below defines a box which contains the optimal solution. Choosing $\mathbf{z}^0$ in this set will yield faster convergence.

There is no reason to ever decrease the value of λ in Step 2 if the objective value decreases on a given iteration, because the profit for the prices $\mathbf{z}^t$ at any iteration t is clearly a lower bound on the optimal profit. However, we find empirically that testing for such a decrease in Step 2 is not necessary. Omitting such a check will allow us to use the same algorithm later when we approximate MMNL models, and the profit at an iteration is no longer a lower bound on the optimum because of approximation error.

In practice, there seems to be no advantage to choosing $L > 1$ in Step 4. In our experiments, we terminate the procedure in Step 5 whenever the change in λ would be sufficiently small.

The following lemma defines the Jacobi iterations that we apply, and characterizes the iteration matrix.

Lemma 5.6. *Define the splittings $(\mathbf{E}_k, \mathbf{F}_k)$ where the matrix $\mathbf{E}_k$ is the positive diagonal of $\mathbf{S}_k^{-1}$ and $\mathbf{F}_k$ is negative off-diagonal such that $\mathbf{S}_k^{-1} = \mathbf{E}_k + \mathbf{F}_k$. Consider the damped Jacobi iteration, with $0 < \alpha \leq 1$, defined by*

$$\mathbf{z}^{t+1} = (1-\alpha)\mathbf{z}^t + \alpha \mathbf{E}_k^{-1}\left(\mathbf{w} - \mathbf{F}_k \mathbf{z}^t\right) = \mathbf{z}^t + \alpha \mathbf{E}_k^{-1}\left(\mathbf{w} - \mathbf{S}_k^{-1}\mathbf{z}^t\right). \tag{5.6}$$

Then the iteration matrices $\mathbf{M}_k = \mathbf{I} - \alpha \mathbf{E}_k^{-1}\mathbf{S}_k^{-1}$ are strictly row-sub-stochastic. We may write

$$\mathbf{z}^{t+1} = T(\mathbf{z}^t) \triangleq \mathbf{M}_k \mathbf{z}^t + \alpha \mathbf{E}_k^{-1}\mathbf{w}.$$

Moreover, the diagonal entries of $\mathbf{E}_k^{-1}$ are less than one.

Proof. For any k, the splitting is the same as in Lemma 5.5. Rewriting the iteration in terms of $\mathbf{M}_k$ consists of simple substitution. We must show that $\mathbf{M}_k$ is sub-stochastic.

Because $\mathbf{E}_k$ is the positive diagonal of the strictly diagonally dominant M-matrix $\mathbf{S}_k^{-1}$, the matrix $\mathbf{E}_k^{-1}\mathbf{S}_k^{-1}$ has a diagonal of all ones and non-positive off-diagonal entries in the range $(-1, 0]$. Moreover, the off-diagonal elements of each row sum to less than one in magnitude because row-diagonal-dominance is preserved by the positive scaling due to $\mathbf{E}_k^{-1}$. It immediately follows that $\mathbf{M}_k$ is strictly sub-stochastic for $\alpha = 1$, since its diagonal elements are all zero. For any value of $0 < \alpha \leq 1$, $\mathbf{M}_k$ is a convex combination of the matrix just discussed with the identity, and therefore also sub-stochastic.

The diagonal entries of $\mathbf{E}_k$ are simply those of $\mathbf{S}_k^{-1}$. But they are all greater than one by its definition (5.3) in terms of the partial derivatives of $G(\mathbf{y})$. Then the entries of $\mathbf{E}_k^{-1}$ are all less than one. □

The stepsize in Lemma 5.5 affects the rate of convergence but not *whether* the iterates converge. In contrast, for the method in Lemma 5.6, we must be careful to choose a sufficiently small stepsize α when updating $\mathbf{S}_k^{-1}$. Even if the step size is too large, the method does not diverge. The iteration matrices are sub-stochastic and therefore have spectral radius less than one. However, the iterates may oscillate. This situation is easily detected and the step size can be reduced accordingly.

The following theorem verifies that the optimal solution to the pricing problem is indeed a fixed point of our algorithm. The theorem also provides a box in $\mathbb{R}^n$ that must contain the optimal $\mathbf{z}^*$. Recall that each entry of $\mathbf{z}$ is simply an affine transformation of the corresponding entry of $\mathbf{x}$ to account for varying profit margins and production costs across products. Therefore, there is a corresponding box in $\mathbb{R}^n$ that must contain $\mathbf{x}^*$.

Theorem 5.7. *Let $\mathbf{x}^*$ be the optimal solution to the pricing problem* (5.1) *under a GEV model. Then the corresponding $\mathbf{z}^*$ is a fixed point of the algorithm defined in this section. Morever, $\mathbf{z}^* \in \left[0, \lambda^* + max_i \frac{1}{b_i}\right]^n \subset \mathbb{R}^n$.*

Proof. If we chose $\mathbf{z}^0 = \mathbf{z}^*$, then in Step 2 of the algorithm, we set $\lambda = \Pi(\mathbf{z}^*) = \lambda^*$.

When $\lambda = \lambda^*$, the optimal solution $\mathbf{z}^*$ satisfies the system of equations (5.5) by Theorem 4.10, and Lemma 5.1. This system defines the Jacobi iteration, so $\mathbf{z}^*$ is clearly a fixed point of the update in Step 4. Therefore $\mathbf{z}^*$ is a fixed point of the algorithm.

By the first equation in Lemma 5.1, any stationary point must satisfy $\mathbf{z} = \mathbf{S}\mathbf{w}$. This holds, in particular, for the global maximum $\mathbf{z}^*$. However, we have already shown in Lemma 4.15 that $\mathbf{S}$ is a sub-stochastic matrix. Therefore $\mathbf{z}^*$ must belong to the specified set since

$$\mathbf{0} \leq \mathbf{S}(\mathbf{B}^{-1} + \lambda^*\mathbf{I})\mathbf{e} \leq \left(\max_i \frac{1}{b_i} + \lambda^*\right)\mathbf{e}.$$

□

5.6.4 A Simpler Fixed Point Iteration

A slightly simpler iteration can be obtained by replacing $\mathbf{E_k}$ with the identity matrix, and choosing a smaller step size. In Lemma 5.8 below, the additive term on each step is constant for a fixed value of λ, rather than just bounded as in Lemma 5.6. The corresponding iterative update for linear systems is known as the modified Richardson iteration [38].

Lemma 5.8. *Consider also the damped Richardson iteration*

$$\mathbf{z}^{t+1} = T(\mathbf{z}^t) \triangleq \mathbf{R}_k\mathbf{z}^t + \alpha\mathbf{w} \tag{5.7}$$

with iteration matrix $\mathbf{R}_k = \mathbf{I} - \alpha\mathbf{S}_k^{-1}$. *For step size* $0 < \alpha \leq \frac{1}{\max_i(\mathbf{E}_k)_{ii}}$, *the matrices* $\mathbf{R}_k$ *are sub-stochastic.*

Proof. The proof is essentially the same as for Lemma 5.6, replacing the diagonal $\mathbf{E}_k$ of $\mathbf{S}_k^{-1}$ simply by $\mathbf{I}$. Consider the matrix $\alpha\mathbf{S}_k^{-1}$ after it is scaled by α. By the condition on α, the diagonal entries of this matrix are in the range $(0, 1]$. As before, diagonal dominance and the M-matrix property is preserved by positive scaling of the rows. Then the sum of the (non-positive) off-diagonal entries in each row is strictly less in magnitude than than the diagonal entry.

Restating the diagonal dominance property in terms of the entries of $\mathbf{R}_k = \mathbf{I} - \alpha\mathbf{S}_k^{-1}$, and writing $[\mathbf{R}_k]_{ij}$ for the (i,j) entry of $\mathbf{R}_k$, we have that for each row i,

$$\sum_{j \neq i} [\mathbf{R}_k]_{ij} < 1 - [\mathbf{R}_k]_{ii}.$$

The left hand side is the sum of the magnitues of the (non-negative) off-diagonal entries of $\mathbf{R}_k$. The right-hand-size corresponds to the diagonal entry of $\alpha\mathbf{S}_k^{-1} = \mathbf{I} - \mathbf{R}_k$. Adding $[\mathbf{R}_k]_{ii}$ on both sides yields that the entries in each row sum to less than one. The off-diagonal entries are non-negative because $\alpha\mathbf{S}_k^{-1}$ is an M-matrix. The diagonal entries are non-negative since those of $\alpha\mathbf{S}_k^{-1}$ are less than one. Therefore $\mathbf{R}_k$ is strictly row-sub-stochastic. □

Using the damped Richardson iteration is more restrictive in theory, since the step size must be chosen as a function of $\mathbf{S}^{-1}$ to ensure that the iteration matrix is actually sub-stochastic. On the other hand, the additive term is now constant and the operator (5.7) has the form of the stochastic dynamic programming operator mentioned in Section 5.6.1. In practice, the two updates appear to perform similarly with properly adjusted constant step sizes [4], even if the actual Richardson iteration matrix is sometimes not actually substochastic.

5.7 The CNL and MMNL Discrete Choice Models

In this section, we first formally define the CNL and MMNL models, and characterize the Jacobian matrix of the prices under both of them via two lemmas. Our algorithm only applies to GEV models, and the MMNL models do not belong to this family. We will show, however, that MMNL models can be locally approximated arbitrarily closely by CNL models, which do belong to the GEV family.

[4] By this we mean that a stepsize of $\alpha = 1$ for (5.6) corresponds to the maximal stepsize allowed for (5.7). Smaller stepsizes for (5.6) correspond to similarly scaled-stepsizes for (5.7).

Therefore, we *can* in fact apply our algorithm to MMNL models via this approximation. There remains the problem that, under MMNL models, the profit often has many local maxima. We show that we can incorporate the path-following method of Hanson and Martin [40] into our algorithm to nonetheless find good local maxima (which we in fact believe to be globally optimal).

5.7.1 The Cross-nested Logit model

We will define the cross-nested logit (CNL) model, and derive expressions for the matrix $\mathbf{S}^{-1}$ in terms of the conditional choice probabilities within each nests. We will use it in our experiments, and to motivate our approach for MMNL models. The CNL model is defined by the generating function

$$G(\mathbf{y}) = \sum_{m=1}^{M} \left(\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{5.8}$$

with conditions on the parameters

1. $\alpha_{im} \geq 0, \forall i, m,$
2. $\sum_m \alpha_{im} > 0, \forall i,$ and
3. $0 < \mu < \mu_m, \forall m.$

Since not all parameters are identified when estimating the model from data, the normalization $\mu = 1$ and $\sum_m \alpha_{im} = 1, \forall i$ are common. Ben-Akiva and Lerman [7] and Train [77] provide details on specific estimation procedures. Bierlaire [11] provides an up-to-date discussion of the CNL model and the various formulations found in the literature.

The partial derivatives of (5.8) with respect to the elements of $\mathbf{y}$ are

$$G_i(\mathbf{y}) = \frac{\partial G(y)}{\partial y_i} = \mu \sum_{m=1}^{M} \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \left(\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1}. \tag{5.9}$$

Thus, the choice probabilities are

$$p_i = \frac{y_i G_i(\mathbf{y})}{\mu G(\mathbf{y})} = \frac{1}{G(\mathbf{y})} \sum_{m=1}^{M} \left(\frac{\alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m}}{\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m}} \right) \left(\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}. \tag{5.10}$$

The choice probabilities can also be re-expressed as

$$p_i = \sum_{m=1}^{M} p_{i|m} q_m,$$

where

$$q_m \triangleq \frac{1}{G(\mathbf{y})} \left(\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \quad \text{and} \quad p_{i|m} \triangleq \left(\frac{\alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m}}{\sum_{j=1}^{n+1} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m}} \right). \tag{5.11}$$

denote the probability that a customer choosing nest m, and the probability that such a customer then chooses alternative i, respectively. As for the NL model of Chapter 3, we define the matrix $\bar{\mathbf{P}} \in \mathbb{R}^{(n+1)\times(n+1)}$ with the choice probabilities p_i on the diagonal, and the matrix $\mathbf{Q} \in \mathbb{R}^{M\times M}$ with the nest probabilities on the diagonal.

The following lemma provides an expression for $\mathbf{S}^{-1}$ under the CNL model. It also characterizes the Jacobian of the choice probabilities with respect to prices, denoted $\mathbf{J}_{\mathbf{z}}^{-1}$ as in Chapters 3 and 4. The latter matrix will allow us to relate the CNL model with the mixed logit model.

Like the matrix $\bar{\mathbf{L}}^{-1}$ compared to $\mathbf{L}^{-1}$, the matrices $\bar{\mathbf{S}}^{-1}$ and $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ have an extra row and column corresponding to the outside alternative. The matrix $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ is actually singular because the total change in demand as a price changes, when including p_{n+1}, must be zero. This implies that the row sums of $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ are all zero. The abuse of the inverse notation is for consistency.

By considering $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ instead of $\mathbf{J}_{\mathbf{z}}^{-1}$, we relax the assumption that the price for the "product" corresponding to the outside alternative is fixed. We define the diagonal matrix $\bar{\mathbf{B}}$, which extends $\mathbf{B}$ with the price sensitivity parameter b_{n+1} for the $(n+1)^{\text{th}}$ product. Similarly, we define the new product's quality parameter d_{n+1} and its price z_{n+1}. Our assumption of a no-purchase option amounts to setting both of these parameters to zero, and removing the last row of $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ containing the derivatives with respect to z_{n+1}. (To recover the square $\mathbf{J}_{\mathbf{z}}^{-1}$, we must also remove the last column corresponding to p_{n+1}. The column is redundant, because, as discussed above, the row sums of $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ are zero.)

Lemma 5.9. *For the cross-nested logit (CNL) model,*

$$\bar{\mathbf{S}}^{-1} = \operatorname{diag}(\mathbf{NWe}) - \mathbf{NWP}_{|}{}^{\top} + \mu\mathbf{NP}_{|}{}^{\top},$$

where $\mathbf{W} = \operatorname{diag}(\mu_1, \ldots, \mu_M)$ *is a diagonal matrix,* $\mathbf{P}_{|} \in \mathbb{R}^{n+1 \times m}$ *is the matrix of conditional choice probabilities with* $[\mathbf{P}_{|}]_{im} = p_{i|m}$ *in the* (i, m) *component, and* $\mathbf{N} = \bar{\mathbf{P}}^{-1}\mathbf{P}_{|}\mathbf{Q}$. [5]

Moreover, the Jacobian matrix of the choice probabilities with respect to prices is

$$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = -\bar{\mathbf{B}}\left(\mathbf{P}\mathbf{S}^{-1} - \mathbf{p}\mathbf{p}^{\top}\right) = -\bar{\mathbf{B}}\left(\operatorname{diag}(\mathbf{Ae}) - \mathbf{A}\right) \tag{5.12}$$

where $\mathbf{A} \in \mathbb{R}^{n+1,n+1}$ *is the symmetric matrix*

$$\mathbf{A} = \begin{bmatrix} \mathbf{P}_{|} & \bar{\mathbf{p}} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{|}{}^{\top} \\ \bar{\mathbf{p}}^{\top} \end{bmatrix}.$$

Proof. See Appendix D.3. □

5.7.2 The Mixed Logit Model

Under MMNL models, the choice probabilities are

$$p_i = \sum_m \gamma_m p_{i|m} \triangleq \sum_m \gamma_m \frac{e^{d_{im} - b_{\cdot|m} z_i}}{\sum_{j=1}^{n+1} e^{d_{jm} - b_{\cdot|m} z_j}},$$

for constant nest sizes $\gamma_1, \ldots, \gamma_M$ such that $\sum_m \gamma_m = 1$. That is, the choice probabilites for each product result from a mixture of MNL models defining the probabilities

[5] For the NL models with no cross-nesting, $\mathbf{N}$ was the incidence matrix of products to nests. It is no longer constant for CNL models, although the equation given here obviously holds in the special case of the NL model.

withing each nest.

For simplicity and without loss of generality, we again include the "price" of the outside alternative as $z_{n+1} = 0$, and the quality parameters $d_{(n+1),1}, \ldots, d_{(n+1),M} = 0$, like in the last section. That is, the last term in the sums over j is equal to $e^0 = 1$ for each nest. This is analogous to setting $y_{n+1} = 1$ for a GEV model, as can be seen if we consider the MNL models obtained when $M = 1$. Like Hanson and Martin [40], we assume that the price sensitivity parameters $b_{\cdot|m}$ are constant for all products within the MNL model representing each segment.[6]

We derive the Jacobian matrix of the choice probabilities for MMNL models, and compare it with the Jacobian for CNL models derived in the preceding lemma.

Lemma 5.10. *Under mixed logit (MMNL) models, the Jacobian of the choice probabilities with respect to prices is given by*

$$\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1} = -\left(\operatorname{diag}\left(\mathbf{A}^{MIX}\mathbf{e}\right) - \mathbf{A}^{MIX}\right)$$

$$\mathbf{A}^{MIX} = \mathbf{P}_{|}\mathbf{W}^{MIX}\mathbf{P}_{|}^{\top}$$

where $\mathbf{W}^{MIX} = \operatorname{diag}\left(b_{\cdot|1}\gamma_1, \ldots, b_{\cdot|M}\gamma_M\right)$ *is a diagonal matrix with the price sensitivity parameters for each nest scaled by the nest sizes, and* $\mathbf{P}_{|} \in \mathbb{R}^{n+1\times m}$ *is the matrix of choice probabilities* $p_{i|m}$.

Proof. See Appendix D.3 □

5.7.3 Approximation of MMNL Models by CNL Models

In Section 2.4 of Chapter 2 we have already provided a local approximation of mixed MNL models by GA demand models. The major problem when using these approximations for optimization is that a stationary point of the MMNL profit is usually not

[6]This is not the same as assuming constant price sensitivity within a nest for the CNL model. The CNL model has one sensitivity parameter b_i per product, and one quality parameter d_i per product. For each nest m, both parameters are affected by the scale parameter μ_m. We assume that the MMNL model has one sensitivity parameter $b_{\cdot|m}$ per *nest*, but the quality parameters d_{im} are specific to each product *and* each nest.

a stationary point of the GA approximation. While they may provide good estimates of the demand, this makes them less suitable than CNL approximations.

The CNL model is extremely flexible because the parameters α_{im} can be chosen arbitraritly. It is relativly straightforward to *locally* approximate a MMNL model by a CNL model with identical choice probabilities *and* similar partial derivatives. This is shown in the next theorem, by comparing $\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1}$ and $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$. The result can potentially be strengthened to have the Jacobians be exactly equal, if we allow the conditional choice probabilities $\mathbf{P}_{|}$ to differ (but not $\bar{\mathbf{p}}$) [7].

Theorem 5.11. *A MMNL model with a constant price sensitivity parameter for each nest can be approximated locally, at a give price vector* $\mathbf{z}$*, by a CNL model the same choice probabilities* $\bar{\mathbf{p}}$*, the same conditional choice probabilities* $\mathbf{P}_{|}$ *and an arbitrarily close Jacobian matrix* $\mathbf{J}_{\mathbf{z}}^{-1}$ *of the choice probabilities with respect to the prices.*

Proof. See Appendix D.1 □

In fact, since we only require the value of the Jacobian matrix in our algorithm, we may bypass the approximation altogether and obtain $\bar{\mathbf{S}}^{-1}$ directly from the Jacobian of the MMNL model. Solving for $\bar{\mathbf{S}}^{-1}$ in the expression for $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ of Lemma 5.9, we have

$$\bar{\mathbf{S}}^{-1} = -\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1}\bar{\mathbf{J}}_{\mathbf{z}}^{-1} + \mathbf{e}\bar{\mathbf{p}}^{\top}.$$

We can then replace $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ with $\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1}$, and recover an approximation of the required matrix $\bar{\mathbf{S}}^{-1}$. The matrix $\mathbf{B}$ is not defined for MMNL models, since they do not have per-product price sensitivity parameters. (The per-nest parameters are captured by the μ_m parameters of the CNL model in the approximation. See the proof of Theorem 5.11.) Therefore we take $\mathbf{B}$ to be the identity, and simply use

$$-\bar{\mathbf{P}}^{-1}\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1} + \mathbf{e}\bar{\mathbf{p}}^{\top}$$

[7] The matrix $[\mathbf{P}_{|}\bar{\mathbf{p}}]$ in Lemma 5.9 is rank-deficient because the last column $\bar{\mathbf{p}} = \mathbf{P}_{|}\mathbf{q}$ is a linear combination of the conditional choice probability vectors. A change of basis can be performed while maintaining positivity of the vectors in the basis: write $\mathbf{A} = \mathbf{P}_{|}(\mathbf{W} + \mathbf{q}\mathbf{q}^{\top})\mathbf{P}_{|}{}^{\top} = \mathbf{P}_{|}\mathbf{X}\mathbf{X}^{\top}\mathbf{P}_{|}{}^{\top}$, where the Cholesky decomposition factor $\mathbf{X}$ has all positive entries (this is easy to show from the Cholesky decomposition algorithm [38]). After appropriate scaling we recover an expression of the form $\mathbf{A}_{MIX}$. The reverse process is similar, but requires an appropriate choice of $\mathbf{q}$ to ensure positivity while maintaining that $\mathbf{P}_{|}\mathbf{q} = \mathbf{p}$.

in the algorithm.

5.7.4 The Path-following Method of Hanson and Martin

Even though we can *locally* approximate MMNL models, the profit often has many local maxima. Therefore we risk converging to a suboptimal solution.

Hanson and Martin [40] already proposed a heuristic for finding good local optima of the profit under MMNL models when optimizing over prices. Their approach serves both as motivation and as a basis of comparison for our own heuristics. Their key observation is that, even though the profit is not concave in terms of the prices $\mathbf{z}$ in general, it becomes concave as the mean utilities for all products are scaled down. Specifically, as *both* the parameters b_i and d_i approach zero. In the utility theory maximization interpretation [77], this is equivalent to increasing the variance in each customer's utility. Intuitively, as the parameters are scaled down and the prices are held constant, the "randomness" in each customer's utility increases, and the probability of choosing each product becomes essentially the same.

Their heuristic first solves the problem with the parameters scaled down sufficiently so that the profit is concave. The parameters are then gradually increased back to their original values on each iteration. The solution of the previous iteration is used as the starting point in the subsequent iteration. This procedure generates a path of solutions, which may avoid getting trapped in local maxima of the original profit function. A number of paths can be generated by selecting different schedules for scaling the parameters.

In the computational experiments of Hanson and Martin [40] (as well as in our own), their approach finds better local solutions more quickly than simply choosing different starting points. However, it still requires solving optimization problems in terms of prices. As we showed in Chapter 2, this can be quite inefficient even for single-segment MNL models. Fortunately, we can now implicitly use a local CNL approximation to the MMNL model, as shown at the end of the preceding section, and apply our algorithm of Section 5.6. In addition to being faster and simpler than the conjugate gradient method sugested by Hanson and Martin [40] to solve the sequence of MMNL models, our algorithm performs well with a more agressive schedule of parameter scalings in practice.

5.8 Experimental Results

5.8.1 Small Scale Instances

Table 5.1 shows the number of steps needed for convergence to maximize various randomly generated CNL profit function, with $n = 10$ products and $m = 4$ nests. The price sensitivity b_i parameters are chosen such that their maximum ratio is as indicated, and the largest value is 1. The scale parameters μ_m for each nests are randomly sampled in the range $[1, \max_m \mu_m]$. The parameter ρ indicates how close the nesting structure is to a NL model. For $\rho = 1$, the α_{im} are sampled randomly. For $\rho = 0$, each product is assigned to a single nest. Intermediate values of ρ represent a convex combination of the the two extremes.

The step size α is chosen as a power of $\frac{1}{2}$ sufficiently small such that a tolerance of 10^{-6} is achieved for the value of λ in less than 1000 iterations. The number of steps needed before convergence is also indicated. The iterates never diverge, but may oscillate or converge very slowly if the step size is too large.

The last four columns represent our attempt to empirically determine whether the functions Π and Ψ are concave, and whether F' is a (generalized) diagonally dominant function. This is accomplished by randomly sampling 1000 prices $\mathbf{z} \in [0, \lambda + \max_i b_i^{-1}]^n$ with the final value of λ. (All stationary points must lie in this set by Theorem 5.7). We then check whether the eigenvalues of the Hessians are negative at all these points. We also check whether the Jacobian of F' scaled by Y is row-diagonally dominant. The last column shows our attempt to find some other scaling vector $\mathbf{u}$ such that the Jacobian is a generalized diagonally-dominant matrix. We use the algorithm of Li [49] with parameter $\theta = 0.001$ and at most 100 iterations. A zero in the last column indicates that, for at least one point, the algorithm failed to find an appropriate scaling of the Jacobian and the result is inconclusive. The function may yet be generalized diagonally-dominant.[8] A zero in the preceeding three columns represents a conclusively negative result, since those conditions are easy to check.

b_{max}/b_{min}	$\max_m \mu_m$	ρ	α	Steps	Π concave	Ψ concave	F' DD	F' GDD
2	5	0.5	0.5	15	1	1	1	1
5	5	0.5	0.5	13	1	1	1	1
10	5	0.5	0.5	13	0	1	1	1
15	5	0.5	1	60	0	0	1	1
20	5	0.5	1	78	0	0	1	1
30	5	0.5	1	254	0	0	1	1
40	5	0.5	0.5	15	0	0	1	1
50	5	0.5	0.5	14	0	0	1	1
10	10	0.5	0.5	13	0	1	1	1
10	15	0.5	0.5	17	0	0	1	1
10	20	0.5	0.5	33	0	0	0	1
10	30	0.5	0.5	31	0	0	0	0
10	40	0.5	0.125	77	0	0	0	0
10	50	0.5	0.125	194	0	0	0	0
10	5	0.1	1	9	0	1	1	1
10	5	0.25	1	8	0	1	1	1
10	5	0.5	0.5	13	0	1	1	1
10	5	0.75	0.5	16	0	1	1	1
10	5	1	1	396	0	1	1	1

Table 5.1: Small-scale random CNL instances.

In the first 8 lines of the table, we vary the ratio between the price sensitivity parameters. Theorem 4.16 shows that Ψ is concave when the ratio is less than 2. However concavity appears to persist for much larger ratios. Even the profit function Π appears to remain concave with ratios up to 5. The Jacobian F' appears to remain diagonally dominant (column DD) even as the ratio becomes large.

Next, we vary the value of the μ_m parameters. Large values indicate that the choices within each nest become almost deterministic in the price. That is, the demand shifts abruptly from one product to another, within each nest, as prices change. As these parameters become large, the Jacobian is no longer DD, and for very large values of the μ_m we cannot verify that it is generalized diagonally dominant at

[8]In principle, trying different parameters for the agorithm of Li [49] should find an appropriate scaling to show geneneralized diagonal-dominance of the Jacobian, or should conclusively show that it is not GDD. However, the matrices where we fail to obtain a result in our experiments are so badly scaled to begin with that it is difficult to choose reliable parameters to use across all instances.

the sampled points. We also observe that a smaller step size is needed for convergence as the μ_m become large.

Finally, we vary the parameter ρ in the last six lines. We note that for all the values Ψ appears to be concave, but Π does not. In Theorem 3.5 we showed that Π is concave when the ration b_{max}/b_{min} is less than 2 for NL models ($\rho = 0$), but it is fixed at 10 for these experiments, so they are consistent with our theoretical results. We observe that a smaller stepsize is more appropriate for CNL models with strong cross-nesting (ρ close to 1), since the second and third row from the bottom required a reduction in step-size before converging. The botton row would likely have exhibited faster convergence with a smaller step size.

In general, intermediate step sizes seem to work best. Even for the instances where the ratios of the price sensitivities are large, a step size of $\alpha = 1/8$ work well. On the other hand, even for the first three rows, a reduction to $\alpha = 0.5$ was required before convergence was achieved.

5.8.2 Large Scale Instances

In Table 5.2 we increase the number of products to $n = 1000$ and the number of nests to $m = 100$. We use a step size of at most 0.5, and reduce the tolerance to 10^{-4}. We indicate the time to convergence with a MATLAB implementation running on an Intel Core 2 Duo laptop. The running time is roughly proportional to the number of iterations needed, and most of the time is spent computing the value of the matrix $\mathbf{G}$. We are able to solve these instances in less than 20 seconds, except when the values of the μ_m are very large (corresponding almost to a deterministic choice model; see above.) We remark that Jacobi-type iterations like our algorithm naturally lend themselves to parallelization, although we have not pursued this. Undoubtedly, an optimized native implementation would be orders of magnitude faster than our MATLAB code.

Interestingly, it is possible to use a larger step size for some of the instances that for the corresponding small-scale instances with similar parameters.

b_{max}/b_{min}	$\max_m \mu_m$	ρ	α	Steps	Time (s)
2	5	0.5	0.5	19	9.19
5	5	0.5	0.5	19	9.5
10	5	0.5	0.5	18	9.01
15	5	0.5	0.5	18	8.88
20	5	0.5	0.5	18	8.99
30	5	0.5	0.5	17	8.24
40	5	0.5	0.5	17	8.77
50	5	0.5	0.5	17	8.52
10	10	0.5	0.5	21	10.42
10	15	0.5	0.5	25	12.63
10	20	0.5	0.5	35	17.93
10	30	0.5	0.5	31	18.06
10	40	0.5	0.25	45	34
10	50	0.5	0.5	127	193.67
10	5	0.1	0.25	40	19.24
10	5	0.25	0.5	20	9.83
10	5	0.5	0.5	18	8.81
10	5	0.75	0.5	18	8.67
10	5	1	0.5	18	9.02

Table 5.2: Large-scale random CNL instances.

5.8.3 Heuristics for MMNL Instances

Table 5.3 and Table 5.4 show the value of the local optima and the running times achieved by four heuristics. HM represents the heuristic of Hanson and Martin [40] with the most aggressive parameter choices they suggested ($\lambda = 0.5$, 20 scaling steps). The optimizations are performed using the pre-conditioned conjugate-gradient (PCG) method implemented in the MATLAB Optimization Toolbox. This method often seems to find the global maximum, and we normalize the objective values by Π_{HM} to facilitate comparison.

The value Π_{rand} is achieved by starting the PCG algorithm at 20 random starting points (to make it comparable with the 20 steps of the HM method). Π_{CNL} represents the result of starting the CNL approximation based heuristic at the same starting points. We report the mean value achieved, and the total running times. All optimizations are stopped when a tolerance of 10^{-6} is achieved or when 10000 iterations are reached.

Finally, Π_{HM-CNL} represents the value achieved when using the same scaling as for

Π_{HM}, but performing a limited number of iterations (50) of the CNL approximation based heuristic for each step.

The third column of both tables represents the scale parameter of the MMNL model. That is, the actual values of the price sensitivity parameters $b_{\cdot|m}$ for each nest are on the order of μ^{-1}. The scale of the μ_m in the CNL approximation of Theorem 5.11 is then on the order of μ^{-1}. That is, if we were sampling comparable instances for the preceding two tables, the value in the column $\max_m \mu_m$ would be approximately $\frac{1}{\mu}$. The largest ratio of the price sensitivity parameters is given in the fourth column. The last column of Table 5.4 shows the maximum number of iterations needed for convergence to within a tolerance 10^{-6} for our algorithm.

n	m	μ	$b_{\cdot\|max}/b_{\cdot\|min}$	Π_{HM}	Π_{rand}	Π_{CNL}	Π_{HM-CNL}
15	10	0.10	2	1.00	1.00	1.00	1.00
15	10	0.05	2	1.00	0.99	0.99	1.00
15	10	0.01	2	1.00	0.90	0.90	1.00
100	10	0.05	2	1.00	1.00	1.00	1.00
100	10	0.01	2	1.00	0.97	1.00	1.00
100	50	0.05	2	1.00	1.00	1.00	1.00
100	50	0.01	2	1.00	0.86	0.87	1.00
100	50	0.05	10	1.00	0.96	0.95	0.93
100	50	0.01	10	1.00	0.74	0.75	0.93

Table 5.3: Comparison of heuristic solutions for randomly generated MMNL instances. The profit achieved by each method is normalized by Π_{HM}.

n	m	μ	$b_{\cdot\|max}/b_{\cdot\|min}$	t_{HM}	t_{rand}	t_{CNL}	t_{HM-CNL}	Max. Iters.
15	10	0.10	2	1.47	2.68	0.18	0.12	42
15	10	0.05	2	1.28	3.65	0.99	0.32	251
15	10	0.01	2	1.17	7.62	2.05	0.30	625
100	10	0.05	2	27.25	84.22	16.95	2.16	525
100	10	0.01	2	24.58	381.19	24.26	2.03	778
100	50	0.05	2	42.17	125.23	22.06	2.67	454
100	50	0.01	2	36.92	748.54	42.17	2.73	1244
100	50	0.05	10	55.58	194.99	33.22	2.78	905
100	50	0.01	10	52.27	473.25	409.85	2.82	10000

Table 5.4: Comparison of the running times of the four heuristics for randomly generated MMNL instances.

First, we observe that Π_{HM} and Π_{HM-CNL} both appear to achieve the global optimum in most cases. Of course, we cannot guarantee this in general. However, Hanson and Martin [40] showed that there is a unique local (and thus global) optimum for sufficiently large values of μ. This is consitent with our results: for $\mu = 0.10$ all the random starting points result in the same solution as the HM heuristic. For the last two rows, Π_{HM-CNL} is sub-optimal, but this could likely be resolved by using a smaller step size α. (The step size was fixed at $\alpha = 0.1$ in these experiments for consistency. A smaller step size was clearly needed for the last row because our algorithm failed to converge even in 10000 iterations. This indicates that the iterates were oscillating as discussed above.)

Randomly selecting enough starting points usually also finds a good local optimum even for smaller values of μ. However, his may be prohibitive for large instances. Clearly, 20 points is insuficient for $n = 100$ when $\mu = 0.01$. Our algorithm performs only slightly better that the PCG method, but is somewhat faster. Therefore more trials could be performed at the same computational cost.

The difference between the methods lies in their running times. Since our method essientially performs a scaled gradient ascent step on each iteration without performing a line search, it is very fast compared to a PCG method. Indeed, using the HM scaling along with our algorithm is similar to solving a simple CNL model except that we are slighly changing the problem as we go along.

5.9 Conclusions

We have provided a simple gradient ascent algorithm to solve the pricing problem for GEV models. Our experiments show that it converges rapidly for CNL models under a wide range of parameters. We provide theoretical insights into why our algorithm may be expected to converge quickly, and why it is reasonable to expect the solution found to be globally optimal. We verify sufficient conditions for uniqueness experimentally for a wide range of parameters.

We also adapted our algorithm to solve the pricing problem under MMNL models. Our experiments show that it finds good local optima as effectively has the algorithm of Hanson and Martin [40], at a fraction of the computational cost. Because it is a first-order method that computes good search directions at each step, there is no need to resort to second-order or conjugate-gradient methods which may not scale well to large instances.

Chapter 6

Conclusions

In this thesis, we have studied the pricing problem under discrete choice models of demand ranging from the basic MNL model to the MMNL and GEV models used in current practice. The question of how to efficiently price a set of substitutable products offered to customers is fundamental in revenue management and marketing. Our results show that optimal solutions can be found reliably and efficiently even in the presence of variable profit margins, production costs and inventory constraints. We have demonstrated that this is true not only for the MNL and the generalized attraction models that have been popular in the revenue management literature due to their simplicity and attractive theoretical properties, but also for the rich customer demand models used in current practice.

From a theoretical standpoint, we have identified the structure of the optimization problem arising when pricing under MNL models, we have explained how it enabled existing work in the literature, and we have extended the same approach to all GA models. We showed that, surprisingly, even NL models share the same concavity properties under a certain range of parameters. More generally, we showed that the profit under NL models is a *bi-concave* function. For the far more general class of GEV models, we exposed the relationship of the pricing problem to quasi-concavity results for fractional programming. Even though known techniques cannot be used to establish the uniqueness of the optimal solution in general, we applied linear algebra results to show that a certain reformulation often gives rise to an auxiliary concave maximization problem. We then applied our results to show that the first-order optimality conditions are often defined nonlinear generalized diagonally dominant functions, which admit a unique solution.

We have proposed efficient algorithms for all of the models we considered. We developped compact closed-form expressions for the Hessian matrices and gradients required to numerically solve the convex optimization problems arising from our reformulations of the MNL, GA and NL pricing problems. For the MNL, a solution can be computed in a polynomial number of iterations. Our structural results for NL models allowed us to apply an iterative optimization algorithms for bi-concave maximization problems where each iterate can essentially be computed in closed form, so that no general-purpose solver is required. We also provided readily-computed expressions for the Hessian matrices and gradients needed to solve the princing problem under CNL and other GEV models. We showed that the well-known nonlinear Jacobi and Gauss-Seidel methods can be used to solve the first-order optimality conditions, and we provided expressions to compute the required quantities. Most importantly, we developped a simple, intuitive and efficient algorithm for pricing under any GEV or MMNL model that does not require any second-order knowledge of the profit, and does not perform expensive line-searches at each iteration. Our method shows that these seemingly complex and difficult optimization problems can in fact be solved rapidly in practice. Our algorithm allows for easy implementation and opens the door for further research on price-based revenue management to parallel recent developments in *multi-period* quantity-based revenue management under customer choice. Since even the quantity-based methods appear intractable for large, complex demand models, such an alternative approach shows great promise, especially in applications where prices are easily updated and represent a driving factor in potential customers' purchasing decisions.

Appendix A

GA Models and the Dual Pricing Problem

A.1 Background on Attraction Demand Models

The class of attraction demand models subsumes a number of important customer choice models by retaining only their fundamental properties. Namely, the form of the demands (2.1) ensures that they are positive and sum to one. A related feature is the well known *independence from irrelevant alternatives* (IIA) property which implies that the demand lost from increasing the price of one product is distributed to other alternatives proportionally to their initial demands.

The attraction functions $f_i(\cdot), i = 1, \ldots, n$ may depend on a number of product attributes in general, but we limit our attention to the effect of price. The requirements of Assumption 2.1 are mild. The positivity assumption and (i) imply that demand for a product is smoothly decreasing in its price but always positive. The requirement (ii) implies that the demand grows to 1 if the price is sufficiently negative, and ensures that increasing the price eventually becomes unprofitable for a seller. As we demonstrate for specific instances below, if the latter two assumptions are not satisfied, the attraction functions can be suitably modified.

Though the class of attraction demand models is very general, certain instances are well studied and admit straightforward estimation methods to calibrate their parameters. This is the case for the MNL and MCI demand models [57, 63]. On the other hand, if assumptions have been made on customers responses to price changes, appropriate attraction functions can be defined to model the desired behavior. Examples of this approach include the linear attraction demand model, and the mixtures of attraction functions discussed in Section 2.4.

A.1.1 Multinomial Logit (MNL) Models

The MNL demand model is a discrete choice model founded on utility theory, where $d_i(\mathbf{x}_i)$ is interpreted as the *probability* that a utility-maximizing consumer will elect to purchase product i. The utility a customer derives from buying product i is $U_i = V_i + \epsilon_i$whereas making no purchase is has utility $U_0 = \epsilon_0$. The V_i terms are deterministic quantities depending on the product characteristics (including price) and the random variables ϵ_i are independent with a standard Gumbel distribution. It can be shown that the probability of product i having the highest realized utility is then in fact given by $d_i(\mathbf{x})$ in (2.1), with $f_i(x_i)$ replaced by e^{V_i}. To model the impact of pricing, we let, for each alternative $i = 1, \ldots, n$,

$$V_i \triangleq V_i(\hat{x}_i) = \beta_{0,i} - \beta_{1,i}\hat{x}_i, \tag{A.1}$$

where $\beta_{0,i} > 0$ represents the quality of product i and $\beta_{1,i} > 0$ determines how sensitive a customer is to its price, denoted here by $\hat{x}_i$. When there is a population of consumers with independent utilities, the fractions $d_i(\mathbf{x})$ represent the portion of the population opting for each product in expectation. For ease of notation, we re-scale the true price $\hat{x}_i$ by $\beta_{1,i}$ to obtain the single-parameter attraction functions (2.2), with $v_i = e^{\beta_{0,i}}$ and $x_i = \beta_{1,i}\hat{x}_i$, rather than using the form of the exponents (A.1) directly. These functions clearly satisfy Assumption 2.1.

Parameters for the demand model used in the experiments of Section 2.6 are generated by sampling the mean linear utilities $V_i(\hat{x}_i)$ in equation (A.1) for each product i. Specifically, $V_i(0)$ and $V_i(x_{max})$ are chosen uniformly over $[2\sigma, 4\sigma]$ and $[-4\sigma, -2\sigma]$ respectively, where $\sigma = \pi/\sqrt{6}$ is the standard deviation of the random

Gumbel-distributed customer utility terms ϵ_i. The parameters $\beta_{0,i}$ and $\beta_{1,i}$ are set accordingly. Recall that the mean utility of the outside alternative is fixed at $V_0 = 0$. The choice of parameters thus ensures that purchasing each product is preferred with large probability when its price is set to 0, and that no purchase is made with large probability when the (unscaled) prices $\hat{x}_i$ are near x_{max}.

A.1.2 Multiplicative Competitive Interaction (MCI) Models

Another common choice of attraction functions is Cobb-Douglas attraction functions $\hat{f}_i(x_i) = \alpha_i x_i^{-\beta_i}$, with parameters $\alpha_i > 0$ and $\beta_i > 1$. It yields the *multiplicative competitive interaction* (MCI) model. Since the attraction is not defined for negative prices, we use its linear extension below a small price ϵ. Let

$$f_i(x_i) = \begin{cases} \alpha_i \epsilon^{-\beta_i} - (x_i - \epsilon)\alpha_i \beta_i \epsilon^{-\beta_i - 1} & \text{if } x_i < \epsilon, \\ \alpha_i x_i^{-\beta_i} & \text{otherwise.} \end{cases} \tag{A.2}$$

This is a mathematical convenience, since one would expect problems involving MCI demand to enforce positivity of the prices. The approximation can be made arbitrarily precise by reducing ϵ.

A.1.3 Linear Attraction Models

This demand model approximates a linear relationship between prices and demands, while ensuring that the demands remain positive and sum to less than one. The attraction function for the i^{th} product is $\hat{f}_i(x_i) = \alpha_i - \beta_i x_i$, with parameters $\alpha_i, \beta_i > 0$. An appropriate extension is needed to ensure positivity. For instance, by choosing the upper bound $\bar{x}_i = \alpha_i/\beta_i - \epsilon$, the following attraction function satisfies our assumptions:

$$f_i(x_i) = \begin{cases} \alpha_i - \beta_i x_i & \text{if } x_i \leq \overline{x}_i, \\ \beta_i e^{-(x_i - \overline{x}_i)/\epsilon} & \text{otherwise.} \end{cases} \tag{A.3}$$

A.2 Pricing under Attraction Demand Models

A.2.1 Non-Convexity of the Naive MNL Pricing Problem

This sections illustrates why the pricing problem (P) is difficult to solve directly in terms of prices, as claimed in Section 2.2. Figure A-1 shows the profit in terms of the prices under an MNL demand model when the number of products is $n = 1$ and $n = 2$. The dashed line in the first plot shows the demand as a function of the price. The profit function is not concave even for a single product. With multiple products, the level sets of the objective are not convex, i.e., the objective is not even quasi-concave.

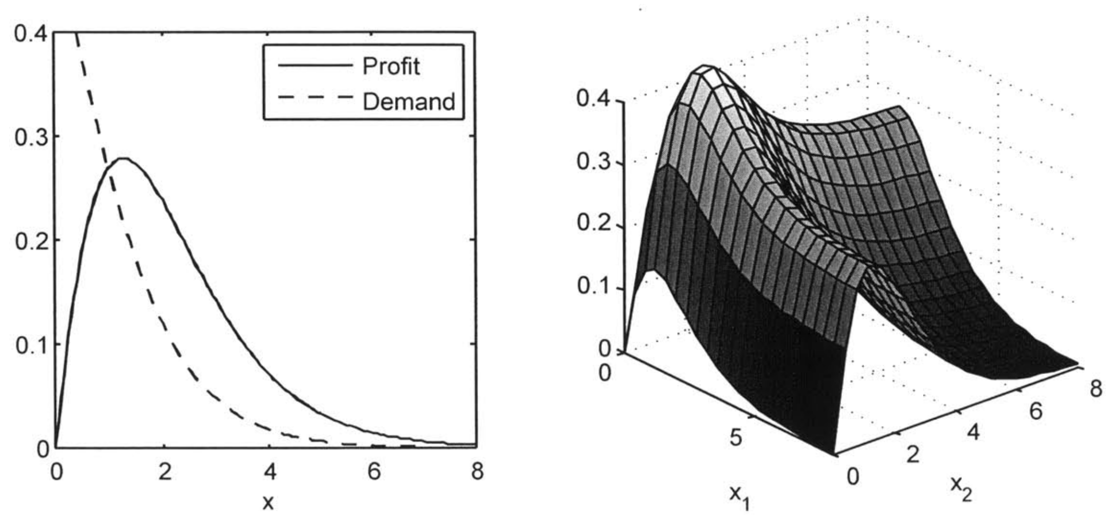

Figure A-1: The objective functions of (P) with a single product (left) and two-products (right).

Furthermore, combining nonlinear constraints with a non-quasi-concave objective function introduces additional complications. First, it is easy to see that, because the objective is not quasi-concave, even a linear inequality constraint in terms of prices could exclude the global maximum in the right panel of Figure A-1, and thus give rise to a local maximum on each of the ridges leading to the peak. Secondly, the feasible region of (P) is in general not convex. Figure A-2 illustrates the constraints of problem (P) with data

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & -\frac{1}{2} \end{bmatrix}, \qquad \mathbf{u} = \begin{bmatrix} 0.6 \\ -0.4 \\ 0.4 \end{bmatrix}.$$

The left panel shows the feasible region in terms of the prices, and the right panel shows the polyhedral feasible region in terms of the demands. Observe that the last two constraints are clearly non-convex in the space of prices. On the other hand, the first constraint happens to belong to the class of convex constraints characterized by Proposition 2.2.

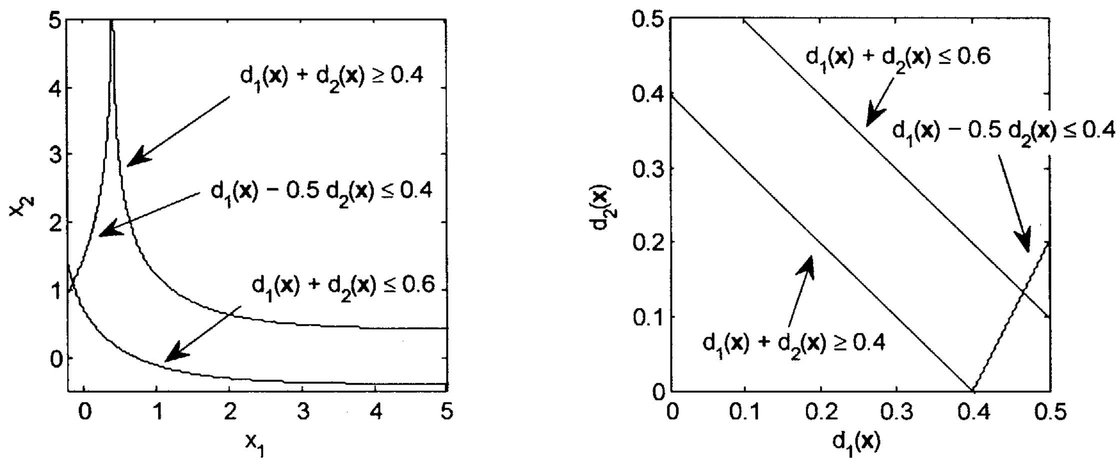

Figure A-2: The feasible region of (P) with two products and three constraints on the demands.

A.2.2 Representing Joint Price Constraints

This section shows how to incorporate certain joint price constraints into the formulations we have proposed. Under MNL demand models, it is natural to assume that the consumer's utility (A.1) is equally sensitive to the price regardless of the alternative she considers. That is, $\beta_{1,i} = \beta_{1,j}$. Then the constraint (2.4) can be expressed as

$$d_i(\mathbf{x}) \leq \frac{f_i(x_j + \delta_{ij})}{f_j(x_j)} d_j(\mathbf{x}) = \frac{v_i f_j(x_j + \delta_{ij})}{v_j f_j(x_j)} d_j(\mathbf{x}) = \left(\frac{v_i}{v_j} e^{-\delta_{ij}} \right) d_j(\mathbf{x}).$$

The assumption regarding the sensitivity to price is required so that the same scaling described in Appendix A.1.1 is used to relate x_i and x_j with $\hat{x}_i$ and $\hat{x}_j$ of equation (A.1), respectively. This allows f_i to be replaced with f_j in the preceding equation. The resulting constraint is evidently linear in terms of the demands and is captured by the formulation (P). This transformation depends on the relationship between the attraction functions for different products and is thus specific to the MNL model. A similar transformation is possible for the linear attraction demand model with the analogous uniform price sensitivity assumption, $\beta_i = \beta_j$ in (A.3). From (2.4), we then have

$$f_i(x_i)d_0(\mathbf{x}) \le f_i(x_j + \delta_{ij})d_0(\mathbf{x}) \quad \Leftrightarrow$$
$$d_i(\mathbf{x}) \le (\alpha_i - \beta_i(x_j + \delta_{ij}))d_0(\mathbf{x}) = (\alpha_i - \alpha_j - \beta_j\delta_{ij})d_0(\mathbf{x}) + d_j(\mathbf{x}),$$

where the $d_0(\mathbf{x})$ terms can be substituted out using the simplex constraint (2.8).

A.2.3 Convexity of (COP) Under Common Attraction Models

Corollary A.1. *Under the linear, MNL and MCI attraction demand models, the objective of* (COP) *is a concave function and any local maximum of either* (COP) *or* (P) *is also a global maximum.*

Proof. For each model, we verify the condition (2.12). For the MNL model (2.2), we have

$$g_i(y) = -\log\frac{y}{v_i},\ g_i'(y) = \frac{-1}{y},\ g_i''(y) = \frac{1}{y^2},\ \text{and}\ 2g_i'(y) + yg_i''(y) = \frac{-2}{y} + \frac{y}{y^2} = \frac{-1}{y} < 0.$$

Now consider the attractions (A.3) for the linear model. For $x > \overline{x}_i$ (i.e., $y < f_i(\overline{x}_i)$) we have the MNL attraction function so the condition (2.12) holds as shown above. Elsewhere, when $x \le \overline{x}_i$,

$$g_i(y) = \frac{\alpha_i - y}{\beta_i}, \quad g_i'(y) = \frac{-1}{\beta_i}, \quad g_i''(y) = 0, \quad \text{and} \quad 2g_i'(y) + yg_i''(y) = \frac{-2}{\beta_i} + 0 = \frac{-2}{\beta_i} < 0.$$

as desired. For the MCI attraction functions (A.2), we have the linear attraction function for $x_i < \epsilon$, otherwise

$$g_i(y) = \left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}}, \quad g_i'(y) = \frac{-1}{\alpha_i\beta_i}\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-1}, \quad g_i''(y) = \frac{-1}{\alpha_i^2\beta_i}\left(\frac{-1}{\beta_i}-1\right)\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-2},$$

$$\begin{aligned} 2g_i'(y) + y g_i''(y) &= \frac{-2}{\alpha_i\beta_i}\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-1} + y\frac{-1}{\alpha_i^2\beta_i}\left(\frac{-1}{\beta_i}-1\right)\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-2} \\ &= \left(-2+\frac{1}{\beta_i}+1\right)\frac{1}{\alpha_i\beta_i}\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-1} = \left(\frac{1}{\beta_i}-1\right)\frac{1}{\alpha_i\beta_i}\left(\frac{y}{\alpha_i}\right)^{\frac{-1}{\beta_i}-1} < 0 \end{aligned}$$

where the inequality uses that $\beta_i > 1$. So the condition (2.12) is also satisfied. □

A.3 The Dual Market Share Problem

Proposition A.2. *The dual of* (COP) *is given by* (DCOP). *For any* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *and* $\mu \in \mathbb{R}$, *there exist optimal* $y_i^* > 0$, *for* $i = 1, \dots, n$, *so that* $\phi_i(y_i^*, \boldsymbol{\lambda}, \mu) > 0$ *in each of the inner maximization problems that appear in the equality constraint of* (DCOP). *Furthermore, when condition* (2.12) *(or equivalently, condition* (2.13)*) is satisfied and* $(\boldsymbol{\lambda}, \mu)$ *is an optimal solution of* (DCOP), *a primal optimal solution* $\boldsymbol{\theta}^*$ *of* (COP) *is given by*

$$\theta_0^* = \frac{1}{1+\sum_{i=1}^n y_i^*}, \qquad \text{and} \qquad \theta_i^* = \frac{y_i^*}{1+\sum_{i=1}^n y_i^*}, \quad i = 1, \dots, n. \tag{A.4}$$

Proof. For each $i = 1, \dots, m'$, let λ_i be the Lagrange multiplier associated with the i^{th} constraint in (COP). Let μ be the multiplier associated with the equality constraint. The Lagrangian is

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{\lambda}, \mu) &= \sum_{i=1}^n a_i\theta_i g_i\left(\frac{\theta_i}{\theta_0}\right) - \sum_{k=1}^{m'}\lambda_k\left(\sum_{i=1}^n A_{ki}\theta_i - u_k\right) - \mu\left(\sum_{i=0}^n \theta_i - 1\right) \\ &= \sum_{i=1}^n \theta_i\left(a_i g_i\left(\frac{\theta_i}{\theta_0}\right) - \sum_{k=1}^{m'}\lambda_k A_{ki} - \mu\right) - \mu\theta_0 + \mu + \sum_{k=1}^{m'}\lambda_k u_k. \end{aligned}$$

Taking the supremum successively over the different variables, we obtain the dual function

$$
\begin{aligned}
L^*(\boldsymbol{\lambda},\mu) \triangleq \sup_{\boldsymbol{\theta}>0} L(\boldsymbol{\theta};\boldsymbol{\lambda},\mu) &= \sup_{\theta_0>0}\left\{\sup_{\theta_1\ldots\theta_n>0} L(\boldsymbol{\theta};\boldsymbol{\lambda},\mu)\right\} \\
&= \mu + \sum_{k=1}^{m'}\lambda_k u_k + \sup_{\theta_0>0}\left\{-\mu\theta_0 + \sup_{\theta_1\ldots\theta_n>0}\sum_{i=1}^{n}\theta_i\left(a_i g_i\left(\frac{\theta_i}{\theta_0}\right) - \sum_{k=1}^{m'}\lambda_k A_{ki} - \mu\right)\right\} \\
&= \mu + \sum_{k=1}^{m'}\lambda_k u_k + \sup_{\theta_0>0}\left(-\mu\theta_0 + \sup_{\theta_1\ldots\theta_n>0}\sum_{i=1}^{n}\theta_0\phi_i(\frac{\theta_i}{\theta_0},\boldsymbol{\lambda},\mu)\right) \\
&= \mu + \sum_{k=1}^{m'}\lambda_k u_k + \sup_{\theta_0>0}\theta_0\left(-\mu + \sum_{i=1}^{n}\sup_{\theta_i>0}\phi_i(\frac{\theta_i}{\theta_0},\boldsymbol{\lambda},\mu)\right), \qquad \text{(A.5)}
\end{aligned}
$$

where $\phi_i(y,\boldsymbol{\lambda},\mu)$ is defined as in (2.29). The value of θ_0 has no impact on the value of the inner supremums in (A.5) since the optimization is over the ratio $\frac{\theta_i}{\theta_0}$ with the numerator free to take any positive value. Thus we may write the dual problem as

$$
\inf_{\boldsymbol{\lambda}\geq 0,\mu} L^*(\boldsymbol{\lambda},\mu) = \inf_{\boldsymbol{\lambda}\geq 0,\mu}\left\{\mu + \sum_{k=1}^{m'}\lambda_k u_k + \sup_{\theta_0>0}\theta_0\left(-\mu + \sum_{i=1}^{n}\sup_{y_i>0}\phi_i(y_i,\boldsymbol{\lambda},\mu)\right)\right\}.
$$

At optimality, the quantity in the inner parentheses must be non-positive, so we may write

$$
\begin{aligned}
\inf \quad & \mu + \sum_{k=1}^{m'}\lambda_k u_k \\
\text{s.t.} \quad & \mu \geq \sum_{i=1}^{n}\sup_{y_i>0}\phi_i(y_i,\boldsymbol{\lambda},\mu) \\
& \boldsymbol{\lambda} \geq 0.
\end{aligned}
$$

The inequality constraint is tight at optimality, because $\phi_i(y_i,\boldsymbol{\lambda},\mu)$ are strictly decreasing in μ.

We now show that $\phi_i(y_i,\boldsymbol{\lambda},\mu)$ achieves a maximum at some $y_i = y_i^* > 0$, for any fixed $\boldsymbol{\lambda}$ and μ. For ease of notation, we fix i and drop the subscript. Let

$$
\phi(y) \triangleq \phi_i(y,\boldsymbol{\lambda},\mu) = y\left(g(y) - \nu\right), \qquad \text{(A.6)}
$$

where we define $g(y) \triangleq a_i g_i(y)$ and $\nu \triangleq \sum_{k=1}^{m'} \lambda_k A_{ki} + \mu$. By Assumption **2**.1, there exists a value $\hat{y} > 0$ for which $g(\hat{y}) = \nu$, since the attraction $f_i(\cdot)$ is defined everywhere on $\mathbb{R}$ and $g(\cdot)$ is its inverse. Moreover, $\phi(\cdot)$ is strictly positive on the interval $(0, \hat{y})$ and strictly negative on $(\hat{y}, \infty)$ since $g(\cdot)$ is strictly decreasing. Also by Assumption 2.1 (ii) ,

$$\lim_{y \downarrow 0} \phi(y) \triangleq \lim_{y \downarrow 0} (y g(y) - y\nu) = \lim_{y \downarrow 0} a_i y g_i(y) = \lim_{x \to \infty} a_i x f_i(x) = 0.$$

We consider the continuous extension of ϕ, with $\phi(0) = 0$, without loss of generality. Then, the continuous function $\phi(\cdot)$ achieves a maximum y_i^* on the closed interval $[0, \hat{y}]$ by Weierstrass' Theorem. We have $0 < y_i^* < \hat{y}$ and $\phi(y_i^*) > 0$, since $\phi(0) = \phi(\hat{y}) = 0$ and ϕ is strictly positive on the interval.

Suppose now that condition (2.12) (equivalently, condition (2.13)) holds. Then (COP) has a concave objective, a bounded polyhedral feasible set and a finite maximum (because the feasible set is bounded). Then the dual (DCOP) has an optimal solution and there is no duality gap. Consider now an optimal dual solution $(\boldsymbol{\lambda}^*, \mu^*)$ and corresponding maximizers $y_1^*, \ldots, y_n^*$. Then (A.4) is a primal optimal solution, since it maximizes the Lagrangian $L(\boldsymbol{\theta}; \boldsymbol{\lambda}, \mu)$ by definition of the dual: we have only made the change of variable $y_i = \frac{\theta_i}{\theta_0}$. □

A.3.1 The Dual Problem under MNL Demand Models

Proposition A.3. *The dual problem* (DCOP) *for the special case of MNL attraction functions* (2.2) *is given by* (DMNL).

Proof. The inverse attraction functions for the MNL model (2.2) and their derivatives are $g_i(y) = -\log \frac{y}{v_i}$, and $g_i'(y) = \frac{-1}{y}$, respectively. Then the first order necessary optimality condition for the i^{th} inner maximization in (DCOP) is

$$\frac{\partial \phi_i}{\partial y} = \left(a_i g_i\left(y\right) - \sum_{k=1}^{m'} \lambda_k A_{ki} - \mu \right) + a_i y g_i'(y) = 0 \quad \Leftrightarrow$$

$$y = v_i \exp\left\{ -1 - \frac{\sum_{k=1}^{m'} \lambda_k A_{ki} + \mu}{a_i} \right\}.$$

The preceding line gives the unique maximizer since one exists by Proposition A.2. Substituting the optimal value of y back into (2.29) yields that

$$\phi_i(y_i^*, \boldsymbol{\lambda}, \mu) = a_i v_i \exp\left\{ -1 - \frac{\sum_{k=1}^{m'} \lambda_k A_{ki} + \mu}{a_i} \right\},$$

which can in turn be substituted into (DCOP) to obtain (DMNL). The constraint may be relaxed to an inequality which is tight at optimality, since the right hand side is decreasing in μ. □

A.3.2 Solving the Dual Problem in General

More generally, there may not exist a closed form solution for the values $\phi_i(y_i^*, \boldsymbol{\lambda}, \mu)$. Then the dual problem may not reduce to a tractable optimization problem. If there is no closed form inverse for the attraction functions, not even the primal market share problem (COP) can be solved directly, even if it has a concave objective function. This is notably the case for the demand models discussed in Section 2.4 (although we have shown that the primal objective function's gradient and Hessian can nevertheless be computed efficiently).

In this section, we present a column generation algorithm to solve the dual which avoids both of these difficulties. It is more general than solving either of the formulations (COP) and (DCOP) directly, since it does not require the convexity of the primal objective function assumed in Theorem 2.4, and it does not require a closed form solution for the inner maximizations of the dual problem.

In the dual (DCOP), fixing the variables $\boldsymbol{\lambda}$ uniquely determines the value of the remaining variable μ, because of the equality constraint. Notice that the right hand side of the constraint is decreasing in μ, because all functions $\phi_i(y, \boldsymbol{\lambda}, \mu)$ are decreasing in μ for any value of y. Furthermore, any feasible μ is positive since the maxima $\phi_i(y_i^*, \boldsymbol{\lambda}, \mu)$ are positive by Proposition A.2. We define $\mu(\boldsymbol{\lambda})$ as the unique root of equation

$$F_{\boldsymbol{\lambda}}(\mu) = \mu - \sum_{i=1}^{n} \max_{y_i > 0} \phi_i(y_i, \boldsymbol{\lambda}, \mu) = 0. \tag{A.7}$$

Its value may be computed by a line search which computes the maximizers y_i^* at each evaluation. When these one-dimensional maximizations are tractable, it is possible to evaluate the dual objective efficiently, and the Dantzig-Wolfe column generation scheme can be applied to solve (COP). (See, for instance, [9] for details.) Specifically, we propose the following algorithm:

1. **Initialization:** Set lower and upper bounds $LB = -\infty$ and $UB = \infty$.

2. **Master Problem:** Given market share vectors $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{L-1}$, solve the following linear program over the variables $\xi_0, \xi_1, \ldots, \xi_{L-1}$:

$$\begin{aligned} \gamma^L = \max \quad & \sum_{\ell=0}^{L-1} \xi^\ell \Pi(\boldsymbol{\theta}^\ell) \\ \text{s.t.} \quad & \sum_{i=1}^{n} A_{ki} \left(\sum_{\ell=0}^{L-1} \xi^\ell \theta_i^\ell \right) \le u_k \qquad k = 1 \ldots m' \\ & \sum_{\ell=0}^{L-1} \xi^\ell = 1, \qquad \xi^\ell \ge 0, \quad \ell = 0, \ldots, L-1. \end{aligned} \tag{LP}$$

   Let $\boldsymbol{\lambda}^L$ be the vector of optimal dual variables associated with the inequality constraints. The master problem solves (COP) with the feasible region restricted to the convex hull of the demand vectors $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{L-1}$. If the optimal value γ^L of (LP) exceeds the lower bound LB, update $LB := \gamma^L$.

3. **Dual Function Evaluation:** Compute the root $\mu(\boldsymbol{\lambda}^L)$ of the dual equality constraint $F_{\boldsymbol{\lambda}^L}(\mu)$ shown in (A.7), and let $\boldsymbol{\theta}^L$ be the primal solution (A.4) corresponding to the maximizers $\{y_i^*, i = 1, \ldots, n\}$. If the dual objective value $L(\boldsymbol{\theta}^L; \boldsymbol{\lambda}^L, \mu(\boldsymbol{\lambda}^L))$ is less than the upper bound, set $UB := L(\boldsymbol{\theta}^L; \boldsymbol{\lambda}^L, \mu(\boldsymbol{\lambda}^L))$.

4. **Termination:** If $(UB-LB)$ is below a pre-specified tolerance, stop. Otherwise, let $L := L+1$ and go to Step 2.

This algorithm requires at least one initial feasible solution $\boldsymbol{\theta}^0$, which can be found by solving any linear program with the constraints of (COP). It does not require (COP) to have a concave objective, since it computes an optimal solution to its dual, which is always a convex minimization problem. Moreover, it can be used even if there is no closed form for the inverse attraction functions $g_i(\cdot)$. Indeed, we can equivalently represent the functions $\phi_i(y, \boldsymbol{\lambda}, \mu)$ in terms of the original attraction functions $f_i(x_i)$, as

$$\psi_i(x_i, \boldsymbol{\lambda}, \mu) \triangleq f_i(x_i) \left(a_i x_i - \sum_{k=1}^{m'} \lambda_k A_{ki} - \mu \right). \tag{A.8}$$

Then the maximization can be performed over the price x_i, and the optimal price for given dual variables $(\boldsymbol{\lambda}, \mu)$ is

$$x_i^* \triangleq \arg\max_{x_i} \psi_i(x_i, \boldsymbol{\lambda}, \mu) = g_i(y_i^*).$$

The maximum is guaranteed to exist since y_i^* exists by Proposition A.2. It can be computed via a line search if it is the unique local maximum. The unimodality of ϕ_i (and equivalently, of ψ_i) is guaranteed, for instance, by the assumption of Theorem 2.4, or more generally, by the assumption of Proposition A.4 below. In the column generation algorithm, the objective of (LP) depends on the prices $x_i = g_i\left(\theta_i^0/\theta_0^0\right)$ corresponding to the initial feasible point. Because they must satisfy $f_i(x_i) = \theta_i^0/\theta_0^0$ and $f_i(x_i)$ is monotone, they can also be found using line search procedures in practice. For each new point $\boldsymbol{\theta}^L$, corresponding prices x_i^* are computed in the maximizations of ψ_i over x_i.

Finally, we remark that it is not necessary to dualize the price bounds represented by the constraints $k = (m+1), \ldots, m'$ defined in (2.9). These constraints may be omitted if the price bounds $\underline{x}_i \leq x_i \leq \overline{x}_i$ are instead enforced when computing the maximizers x_i^* (or, equivalently, the bounds $f_i(\underline{x}_i) \geq y_i \geq f_i(\overline{x}_i)$ are enforced when computing y_i^*). This modification reduces the number of constraints from $m' = (m + 2n)$ to m.

The algorithm just described may also be viewed as a generalization of the procedure presented by Gallego and Stefanescu [31] to general attraction demand models and arbitrary linear inequality constraints. (Although they arrive at their method by taking the dual of the price-based formulation (P) for the special case of MNL demand.) Because convergence of column generation algorithms is often slow near the optimum, we expect that directly solving (COP) or (DCOP) will be more efficient when it is possible. This, for example, is the case with the MNL demand models considered by [31]. However, the column generation algorithm applies to demand models where it is not possible to solve the other formulations. It can provide an upper bound on the optimal profit when the objective function of (COP) is not concave, and can often compute an approximate solution quickly (accurate within a few percent in relatively few iterations, as shown in our experiments).

We end this section with the following proposition providing a sufficient condition on the inverse attractions guaranteeing unique maximizers y_i^*. It requires that the inverse attraction functions are "sufficiently concave" (though not necessarily concave) up until some $\bar{y}$, and then "sufficiently convex" afterward. Omitting the ratio $\frac{x}{y}$, conditions (A.9) and (A.10) below correspond to strict concavity and strict convexity, respectively. However, the first requirement is weaker, and the second is stronger, because this ratio is less than one. (Recall that $g_i'(x) < 0, \forall x$ since f_i and g_i are decreasing.) We note that the proposition allows $\bar{y} = 0$ or $\bar{y} = \infty$, in which case one of the assumptions holds trivially.

Proposition A.4. *If for each $i = 1, 2, \ldots, n$, there exists a point $\bar{y}_i \in [0, \infty]$ such that*

$$g_i(y) < g_i(x) + \frac{x}{y}(y - x)g_i'(x), \qquad \forall x, y \in (0, \bar{y}_i], x < y, \qquad \textit{and} \tag{A.9}$$

$$g_i(y) > g_i(x) + \frac{x}{y}(y - x)g_i'(x), \qquad \forall x, y \in [\bar{y}_i, \infty), x < y, \tag{A.10}$$

then the maximizers $\{y_i^*, i = 1, \ldots, n\}$ are unique for any values of $\boldsymbol{\lambda}$ and μ.

Proof. We fix i and use the simplified notation defined in (A.6). From Proposition A.2, the maximizer $y_i^* > 0$ exists, and it must be a stationary point of ϕ. We will show that the rightmost stationary point to the left of $\bar{y}_i$ maximizes $\phi(y)$ over $(0, \bar{y}_i]$, and that the leftmost stationary point to the right of $\bar{y}_i$ maximizes $\phi(y)$ over $[\bar{y}_i, \infty)$, if they exist. At least one of them must exist since we know a maximum is attained. If both exist, we deduce that there is an additional stationary point between them by applying the mean value theorem. This contradicts the fact that they are the rightmost and leftmost stationary points on their respective intervals, proving uniqueness of the maximizer.

Suppose $y \in (0, \bar{y}_i]$ is a stationary point of $\phi(\cdot)$, i.e.

$$\phi'(y) = g(y) - \nu + yg'(y) = 0 \qquad \Rightarrow \qquad \nu = g(y) + yg'(y). \tag{A.11}$$

We will show that for any other point $x \in (0, y)$, whether or not it is a stationary point,

$$\begin{aligned}
&\phi(x) < \phi(y) \quad \Leftrightarrow \quad x(g(x) - \nu) < y(g(y) - \nu) \quad \Leftrightarrow \\
&x\,(g(x) - g(y) - yg'(y)) < -y^2 g'(y) \quad \Leftrightarrow \quad x(g(x) - g(y)) < (x - y)\,yg'(y) \quad \Leftrightarrow \\
&\qquad\qquad g(x) - g(y) < (x - y)\frac{y}{x}g'(y) \quad \Leftrightarrow \quad \frac{x}{y}\left(\frac{g(x) - g(y)}{x - y}\right) > g'(y),
\end{aligned}$$

where we used (A.11). Having fixed y, we denote the left hand side as a function of x by

$$h(x) = \frac{x}{y}\left(\frac{g(y) - g(x)}{y - x}\right)$$

and note that $\lim_{x \uparrow y} h(x) = g'(y)$. Thus, to prove the inequality, it is sufficient to show that the continuous function $h(x)$ is decreasing in x on the interval $(0, y)$. We consider the derivative with respect to x

$$\begin{aligned}
h'(x) &= \frac{g(x)-g(y)}{y(x-y)} + \frac{x}{y}\left(\frac{g'(x)}{x-y} - \frac{g(x)-g(y)}{(x-y)^2}\right)\\
&= \frac{1}{(x-y)^2}\left(\frac{x-y}{y}\left(g(x)-g(y)\right) + \frac{x}{y}(x-y)g'(x) - \frac{x}{y}\left(g(x)-g(y)\right)\right)\\
&= \frac{1}{(x-y)^2}\left(\frac{x}{y}(x-y)g'(x) - g(x) + g(y)\right)\\
&= a_i\frac{1}{(x-y)^2}\left(-\frac{x}{y}(y-x)g_i'(x) - g_i(x) + g_i(y)\right) < 0.
\end{aligned} \tag{A.12}$$

The assumption (A.9) implies that the above derivative is negative, where we have substituted $g_i(\cdot)$ back in, and thus $h(x)$ is decreasing.

A similar argument shows the analogous result for stationary points to the right of $\bar{y}_i$. Take instead $x \in (\bar{y}_i, \infty)$ to be the leftmost stationary point in the half-line, and let $y \in (x, \infty)$ be some other stationary point. We still have that $x < y$, but now $\phi(x) > \phi(y) \Leftrightarrow h(x) < g'(y)$, because $h(x)$ is *increasing* in x. This is implied by the assumption (A.10), which shows that the derivative in (A.12) is now positive. □

A.3.3 Performance of the Column Generation Algorithm

Table A.1 shows the accuracy achieved and the running time in seconds after a fixed number of iterations of the column generation algorithm, when applied to randomly generated problem instances with four overlapping customer segments, using the approximation of Section 2.4. Only the most recently active 512 columns are retained in the master problem (LP). We have no closed form for the inner maximizers y_i^* and instead use a numerical minimization algorithm based on Brent's method to compute them. Brent's method (see Brent [15]) is also used to solve (A.7) numerically. The algorithm was halted if six significant digits of accuracy were achieved.

As is often the case for column generation algorithms, we observe fast convergence early on. After 500 iterations, most of the instances are solved to within 10 percent of the optimal objective value. Quadrupling the number of iterations further reduces the duality gap to a few percentage points in all but the largest instances. The solution

Products	Constraints	100 Iteration		500 Iterations		2000 Iterations	
(n)	(m)	Duality Gap	Time	Duality Gap	Time	Duality Gap	Time
16	256	0.03%	0.41	(< 1e-6)	0.86	(< 1e-6)	0.86
64	256	19.63%	1.10	0.68%	11.83	(< 1e-6)	90.05
256	256	31.28%	2.49	4.18%	25.99	0.19%	198.75
512	256	25.77%	4.05	4.75%	38.30	0.92%	263.76
1,024	256	14.39%	8.74	4.12%	65.91	2.71%	393.85
2,048	256	4.45%	18.31	1.25%	117.06	1.15%	605.92
4,096	256	1.35%	35.50	0.52%	212.99	0.52%	1,150.80
256	16	0.03%	1.49	0.01%	9.54	0.01%	55.87
256	64	6.72%	1.36	0.10%	12.74	0.07%	68.91
256	256	31.28%	2.49	4.18%	25.99	0.19%	198.75
256	512	237.23%	4.02	12.77%	75.98	1.79%	638.32
256	1,024	215.69%	9.98	24.42%	250.18	5.17%	1,882.00
256	2,048	161.09%	20.73	26.25%	583.72	6.80%	4,636.20
256	4,096	178.72%	64.33	28.96%	1,561.70	7.76%	10,944.00

Table A.1: Duality gap as a percentage of LB and running time in seconds for the column generation algorithm.

times compare favorably with the price formulation (P) and the dual formulation (DCOP) for the single-segment case, but are significantly slower than for the market-share formulation (COP). Of course, the latter formulation requires the custom objective evaluation code described in the preceding section when multiple segments are being approximated. We conclude that the column generation method offers a viable alternative when the other formulations cannot be applied easily, and only limited accuracy is needed.

Appendix B

Derivations for the Nested Logit Model

B.1 Explicitly Inverting the Choice Model

Lemma 3.1. *Under a NL model, for* $i = 1, \dots, n+1$,

$$y_i = \frac{p_{i|m_i}^{\frac{1}{\mu_{m_i}}} Q_{m_i}}{p_{n+1|m^*}^{\frac{1}{\mu_{m^*}}} Q_{m^*}} = \frac{p_i^{\frac{1}{\mu_{m_i}}} Q_{m_i}^{1-\frac{1}{\mu_{m_i}}}}{p_{n+1}^{\frac{1}{\mu_{m^*}}} Q_{m^*}^{1-\frac{1}{\mu_{m^*}}}}.$$

Therefore the demand function $p : \mathbb{R}^n \to \Delta_{n+1}$ *is invertible (one-to-one and onto). Moreover, both* $p(\cdot)$ *and* $p^{-1}(\cdot)$ *are differentiable.*

Proof. In general, the choice probabilities under GEV models, such as the NL, are defined by the partial derivatives $G_i(\mathbf{y}) = \left.\frac{\partial G}{\partial y_i}\right|_{\mathbf{y}}$ of the function $G(\mathbf{y})$ stated in (3.4). Specifically,

$$p_i = \frac{y_i G_i(\mathbf{y})}{\mu G(\mathbf{y})}, \qquad i = 1, \dots, n,$$

where we have implicitly set $\mu = 1$ for the NL model. (See Section 4.2 for details.) For consistency in the proofs, we use the slightly more general definition of the function $G(\mathbf{y})$ for the *cross*-nested logit (CNL) model presented in Section 5.7.1. Setting $\mu = 1$ and restricting the α_{im} to be binary parameters reduces the function in Equation (5.8) exactly to function we have stated for the NL model in (3.4).

The partial derivatives $G_i(\mathbf{y})$ are stated in (5.9). We can rewrite them slightly and substitute in the expressions for the conditional probabilities of (5.11) (which also reduce to those for the NL model).

$$\begin{aligned}
G_i(\mathbf{y}) &= \mu \sum_{m=1}^{M} \left(\alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m}\right)^{1-\frac{1}{\mu_m}} \alpha_{im}^{\frac{1}{\mu}} \left(\sum_{j=1}^{n} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m}\right)^{\frac{\mu-1}{\mu_m}-(1-\frac{1}{\mu_m})}, \\
&= \mu \sum_{m=1}^{M} \alpha_{im}^{\frac{1}{\mu}} p_{i|m}^{1-\frac{1}{\mu_m}} \left(\sum_{j=1}^{n} \alpha_{jm}^{\frac{\mu_m}{\mu}} y_j^{\mu_m}\right)^{\frac{\mu-1}{\mu_m}}, \\
&= \mu^{\frac{1}{\mu}} \sum_{m=1}^{M} \alpha_{im}^{\frac{1}{\mu}} p_{i|m}^{1-\frac{1}{\mu_m}} \left(Q_m G(\mathbf{y})\right)^{1-\frac{1}{\mu}} \\
&= \left(G(\mathbf{y})\right)^{1-\frac{1}{\mu}} \mu^{\frac{1}{\mu}} \sum_{m=1}^{M} \alpha_{im}^{\frac{1}{\mu}} p_{i|m}^{1-\frac{1}{\mu_m}} Q_m^{1-\frac{1}{\mu}}.
\end{aligned}$$

Since α_{im} is nonzero for a single nest m in the NL model, the summation in the last line reduces to a single term. The expression simplifies to

$$G_i(\mathbf{y}) = \left(G(\mathbf{y})\right)^{1-\frac{1}{\mu}} \mu^{\frac{1}{\mu}} p_{i|m_i}^{1-\frac{1}{\mu_{m_i}}} Q_{m_i}^{1-\frac{1}{\mu}}.$$

We can now solve for y_i in the definition of p_i above:

$$\begin{aligned}
y_i &= \frac{\mu G(\mathbf{y})}{G_i(\mathbf{y})} p_i \\
&= \frac{\mu G(\mathbf{y})}{G^{1-\frac{1}{\mu}}(\mathbf{y}) \mu^{\frac{1}{\mu}} p_{i|m_i}^{1-\frac{1}{\mu_{m_i}}} Q_{m_i}^{1-\frac{1}{\mu}}} p_{i|m_i} Q_m \\
&= \mu^{1-\frac{1}{\mu}} G^{\frac{1}{\mu}}(\mathbf{y}) p_{i|m_i}^{\frac{1}{\mu_{m_i}}} Q_{m_i}^{\frac{1}{\mu}}
\end{aligned}$$

Since $y_{n+1} = 1$ is fixed and the no-purchase option is in nest $m_{n+1} = m^*$, we have

$$G^{\frac{1}{\mu}}(\mathbf{y}) = \mu^{\frac{1}{\mu}-1} p_{n+1|m^*}^{-\frac{1}{\mu_{m^*}}} Q_{m^*}^{-\frac{1}{\mu}}.$$

Substituing this quantity back into the expression for any of the y_i, we obtain

$$y_i = \left(\frac{p_{i|m_i}}{p_{n+1|m^*}}\right)^{\frac{1}{\mu_{m_i}}} \left(\frac{Q_{m_i}}{Q_{m^*}}\right)^{\frac{1}{\mu}} p_{n+1|m^*}^{\frac{1}{\mu_{m_i}} - \frac{1}{\mu_{m^*}}}.$$

The first equality of the statement follows. The second equality follows immediately because $p_i = p_{i|m_i} Q_{m_i}$. Invertibility and differentiability are also immediate becase we have given differentiable close-form expressions for y_i and p_i. The mapping between the y_i and the prices $\mathbf{x}$ is clearly invertible and differentiable. □

B.2 The Jacobian of the Prices

In the following proofs, we use the shorthand $\mathbf{q} \triangleq \mathbf{Q_m}$ to denote the vector of nest probabilities $\mathbf{Q_m} = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_M \end{bmatrix}^\mathsf{T}$, for clarity. This is not to be confused with the vector $\mathbf{q}$ defined in Chapter 4. The matrix $\mathbf{Q}$ has the vector $\mathbf{q}$ on the diagonal.

We find it necessary to deal with functions of $\mathbf{p}$, $\bar{\mathbf{p}}$ or the pair $(\bar{\mathbf{p}}, \mathbf{q})$ at different times. As a rule, quantities related to $\bar{\mathbf{p}}$ are also marked with a bar, and quantities related to $(\bar{\mathbf{p}}, \mathbf{q})$ are subscripted. Rather than work directly with $\mathbf{J_z}$, we will define the Jacobians of $\mathbf{y}$ with respect to $\mathbf{p}$, $\bar{\mathbf{p}}$ and $(\bar{\mathbf{p}}, \mathbf{q})$, and denoted them by $\mathbf{J}$, $\bar{\mathbf{J}}$ and $\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}$, respectively.

We denote the vector of all ones by $\mathbf{e}$ and the identity matrix by $\mathbf{I}$ regardless of dimension. The vectory $\mathbf{e}_i$ represents the vector of all zeros except for a 1 in the i^{th} entry. We also define the diagonal matrices $\bar{\mathbf{P}}$, $\mathbf{P}$, $\bar{\mathbf{Y}}$ and $\mathbf{Y}$ with the vectors $\bar{\mathbf{p}}$, $\mathbf{p}$, the entire vector $\bar{\mathbf{y}} = \begin{bmatrix} y_1, y_2, \ldots, y_n, y_{n+1} = 1 \end{bmatrix}^\mathsf{T}$ and the first n entries of $\mathbf{y}$ on their respective diagonals. The diagonal Jacobian matrix $\mathbf{D} = -\mathbf{B}^{-1}\mathbf{Y}^{-1}$ of $\mathbf{z}$ with respect to $\mathbf{y}$ is defined in Lemma 3.4 and used throughout.

Lemma 3.4. *Let $\bar{\mathbf{P}} \in \mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{Q} \in \mathbb{R}^{m \times m}$ be the positive diagonal matrices with the probabilities $\bar{\mathbf{p}}$ and $Q_1, \ldots, Q_m$ on their respective diagonals. Define the diagonal matrices of parameters $\bar{\mathbf{U}} \in \mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{V} \in \mathbb{R}^{M \times M}$ with entries*

$$\bar{U}_{ii} = \frac{1}{\mu_{m_i}} > 0 \qquad \textit{and} \qquad V_{mm} = \left(1 - \frac{1}{\mu_m}\right) \geq 0.$$

Then the partial derivatives of $\mathbf{z}$ *with respect to* $\mathbf{p}$ *are*

$$\mathbf{J_z} = \left[\frac{\partial z_j}{\partial p_i}\right]_{ij} = -\mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} + \mathbf{N}\mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top\right)\mathbf{M}^\top\mathbf{B}^{-1}.$$

Proof. We assume without loss of generality that the outside alternative belongs to the last nest $m_{n+1} = m^* = M$. The partial derivatives of the vector $\mathbf{y}$ can be computed directly from the expression in Lemma 3.1. Then, by the chain rule,

$$\mathbf{J_z} = \mathbf{J}\mathbf{D}, \qquad \text{where} \qquad \mathbf{J} = \left[\frac{\partial y_j}{\partial p_i}\right]_{ij},$$

and where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is the diagonal Jacobian matrix with entries

$$D_{ii} = \left[\frac{\partial y_i}{\partial z_i}\right]_{ij} = -\mathbf{B}^{-1}\mathbf{Y}^{-1} = \frac{-1}{b_i' y_i}.$$

This follows directly by differentiating Equation (3.7) in Section 3.5.1 with respect to z_i, and taking the inverse of $\partial z_i/\partial y_i = -b_i' y_i$. We now compute the entries of $\mathbf{J}$.

The key observation is that, because y_i Lemma 3.1 has the form of a monomial (with negative and fractional exponents) in the four variables p_i, Q_{m_i}, p_{n+1} and Q_{m*}, the partial derivatives have the same form. One can then factor out y_i. For example,

$$\frac{\partial y_i}{\partial p_i} = \frac{1}{\mu_i p_i} y_i. \tag{B.1}$$

All the partial derivatives of $\mathbf{y}$ with respect to $(\bar{\mathbf{p}}, \mathbf{q})$ are

$$\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} = \begin{bmatrix} \mathbf{U}\mathbf{P}^{-1} \\ -(\mu_{n+1} p_{n+1})^{-1}\mathbf{e}^\top \\ \mathbf{V}\mathbf{Q}^{-1}(\mathbf{N}^\top - \mathbf{e}_M\mathbf{e}^\top) \end{bmatrix} \mathbf{Y} \in \mathbb{R}^{(n+1+M)\times n},$$

where the matrix $\mathbf{N}^\top - \mathbf{e}_M\mathbf{e}^\top$ is the matrix $\mathbf{N}$ with the vector of all ones subtracted from the last column, transposed. It is straighforward to compute the entries one-by-one, as we have done above for diagonal elements in the first n rows. The $(n+1)^{\text{th}}$ row is dense, because all the y_j involve p_{n+1}. For the next-to-last $(M-1)$ rows, there is a non-zero entry $(n+1+m, j)$ if product j belongs to nest m (y_j depends on Q_m). The last row has a non-zero entry if product j is *not* in the last nest $M = m_*$, because all the y_j depend on Q_M except if product j is in nest M (and the Q_M factors cancel out).

Observe that

$$\mathbf{N}^\top \mathbf{M}^\top = \mathbf{N} \begin{bmatrix} \mathbf{I} & -\mathbf{e} \end{bmatrix} = \mathbf{N}^\top - \mathbf{e}_M \mathbf{e}^\top$$

because only the last M^{th} column of $\mathbf{N}$ has a one in the last coordinate. We can simplify the expression above to obtain

$$\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} = \begin{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} \\ \mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{e}^\top \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} \\ \mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top \end{bmatrix} \mathbf{M}^\top \mathbf{Y} \tag{B.2}$$

Applying the chain rule, we obtain the partial derivatives with respect to $\mathbf{p}$,

$$\bar{\mathbf{J}} = \begin{bmatrix} \mathbf{I} & \mathbf{N} \end{bmatrix} \mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} = \left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} + \mathbf{N}\mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top\right) \mathbf{M}^\top \mathbf{Y}. \tag{B.3}$$

Applying the chain rule again, $\mathbf{J} = \mathbf{M}\bar{\mathbf{J}}$. Substituting back into the expression of $\mathbf{J}_\mathbf{z} = -\mathbf{J}\mathbf{Y}^{-1}\mathbf{B}^{-1}$ yields the result. □

B.3 The Hessian of the Profit

B.3.1 A Chain Rule for Hessians

We shall need the following chain rule for Hessians. This and similar results are common in the literature. For this section, we may assume that the tensors of second partial derivatives $\mathcal{H}$ and $\mathcal{K}$ are zero, since we apply the chain rule only to linear functions. The full statement will be required in Appendix C to prove the results of Chapter 4.

Proposition B.1 (Chain rule for Hessians). *Consider the twice-differentiable function $\Pi(\mathbf{y}) : \mathbb{R}^n \to \mathbb{R}$, and the continuously differentiable mapping $y(\mathbf{p}) : \mathbb{R}^n \to \mathbb{R}^n$ from vectors $\mathbf{p}$ to vectors $\mathbf{y}$. Let $\mathbf{g}$ and $\mathbf{H}_\mathbf{y}$ denote the gradient and Hessian of $\Pi(\mathbf{y})$,* *respectively. Let* $\mathbf{J}$ *be the Jacobian of* $y(\mathbf{p})$*. Then the Hessian of* $\Pi(y(\mathbf{p}))$ *is given by*

$$\mathbf{H_p} = \mathbf{J}\mathbf{H_y}\mathbf{J}^\top + \mathbf{g} \cdot \mathcal{H},$$

where $\mathbf{g} \cdot \mathcal{H}$ *denotes the tensor product* $\mathbf{g} \cdot \mathcal{H} = \sum_i g_i \mathcal{H}^i$*, and* $\mathcal{H}^i$ *is the Hessian of* y_i *with respect to* $\mathbf{p}$*. Furthermore, if* $\mathbf{J}$ *is invertible, then*

$$\mathbf{H_p} = \mathbf{J}\left(\mathbf{H_y} - \mathbf{J}\mathbf{g} \cdot \mathcal{K}\right)\mathbf{J}^\top,$$

where $\mathcal{K}^j$ *is the Hessian of* p_j *with respect to* $\mathbf{y}$*.*

Proof. The (i, j) element of $\mathbf{H_p}$ is

$$h_{ij} = \frac{\partial^2 \Pi}{\partial p_i \partial p_j} = \frac{\partial}{\partial p_i}\left(\frac{\partial \Pi}{\partial y_j}\right) = \frac{\partial}{\partial p_i}\left(\sum_k \frac{\partial \Pi}{\partial y_k}\frac{\partial y_k}{\partial p_j}\right) = \sum_k \frac{\partial}{\partial p_i}\left(\mathbf{g}_k \mathbf{J}_{jk}\right).$$

By applying the scalar product rule, we obtain the two terms,

$$\begin{aligned} h_{ij} &= \sum_k \frac{\partial \mathbf{g}_k}{\partial p_i}\mathbf{J}_{jk} + \sum_k \mathbf{g}_k \frac{\partial \mathbf{J}_{jk}}{\partial p_i} \\ &= \sum_k \mathbf{J}_{jk}\left(\sum_\ell \frac{\partial \mathbf{g}_k}{\partial y_\ell}\frac{\partial y_\ell}{\partial p_i}\right) + \sum_k \mathbf{g}_k \frac{\partial^2 y_k}{\partial p_i \partial p_j} \\ &= \sum_k \mathbf{J}_{jk}\left(\sum_\ell \frac{\partial \Pi}{\partial y_k \partial y_\ell}\mathbf{J}_{i\ell}\right) + \sum_k \mathbf{g}_k \mathcal{H}^k_{ij} \\ &= \sum_k \sum_\ell \mathbf{J}_{jk}\mathbf{J}_{i\ell}\mathbf{H}_{k\ell} + \sum_k \mathbf{g}_k \mathcal{H}^k_{ij}. \end{aligned}$$

This proves the first equality in the statement.

If $\mathbf{J}^{-1}$ exists, then, by the inverse function theorem, the continuously differentiable inverse mapping $p(\mathbf{y}) = y^{-1}(\mathbf{y})$ exists in a neighboorhood of the point where the Hessians are being evaluated. Moreover, its Jacobian is $\mathbf{J}^{-1}$. Applying the chain rule we have just proved, we have that

$$\mathbf{H_y} = \mathbf{J}^{-1}\mathbf{H_p}\left(\mathbf{J}^{-1}\right)^\top + \mathbf{J}\mathbf{g} \cdot \mathcal{K},$$

where $\mathbf{Jg}$ is the gradient of $\Pi(y(\mathbf{p}))$ by the chain rule. Substituting the last expression for $\mathbf{H_y}$ back into the first equality, we obtain

$$\mathbf{H_p} = \mathbf{H_p} + \mathbf{J}(\mathbf{Jg} \cdot \mathcal{K})\mathbf{J}^\top + \mathbf{g} \cdot \mathcal{H}.$$

Subtracting $\mathbf{H_p}$ on both sides, we have that $\mathbf{g} \cdot \mathcal{H} = -\mathbf{J}(\mathbf{Jg} \cdot \mathcal{K})\mathbf{J}^\top$. Substituting this last expression in the first equality of the statment yields the second equality. □

B.3.2 An Explicit Expression for the Hessian

The following lemma gives a first expression for the Hessian. We develop it further after the proof.

Lemma B.2. *The gradient of the profit* $\Pi = \mathbf{p}^\top \mathbf{z}$ *is*

$$\nabla_{\mathbf{p}} \Pi = \mathbf{z} + \mathbf{JDp},$$

and its Hessian matrix is

$$\mathbf{H} = \mathbf{DJ}^\top + \mathbf{JD} + \mathbf{JY}^{-1}\mathbf{PB}^{-1}\mathbf{Y}^{-1}\mathbf{J}^\top + \mathbf{Dp} \cdot \mathcal{H}. \tag{B.4}$$

Proof. We will show that the gradient and Hessian of the function

$$\Pi(\bar{\mathbf{p}}) = \bar{\mathbf{p}}^\top \begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix}$$

with respect to $\bar{\mathbf{p}}$ are, respectively,

$$\nabla_{\bar{\mathbf{p}}} \Pi = \begin{bmatrix} \mathbf{z} \\ 0 \end{bmatrix} + \bar{\mathbf{J}}\mathbf{Dp},$$

and

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{D}\bar{\mathbf{J}}^\top \\ 0 \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{J}}\mathbf{D} & 0 \end{bmatrix} + \bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{PB}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \mathbf{Dp} \cdot \bar{\mathcal{H}}.$$

The conclusion then follows directly from the vector chain rule and from the chain rule for Hessians of Proposition B.1, because the Jacobian of $\bar{\mathbf{p}}$ with respect to $\mathbf{p}$ is $\mathbf{M}^\top$ and their relationship is affine.

The expression for the gradient follows directly from the chain rule, since $\mathbf{D}$ is the Jacobian of the prices $\mathbf{z}$ with respect to the vector $\mathbf{y}$. The first term of the Hessian follows similarly. We still need to differentiate

$$\left[\bar{\mathbf{J}}\mathbf{D}\mathbf{p}\right]_i = \sum_{k=1}^{n} \frac{\partial y_k}{\partial p_i} \cdot \left(\frac{-p_k}{y_k b_k}\right)$$

with respect to the vector p_j. From the product rule, we have, for $j \leq n$,

$$\frac{\partial}{\partial p_j}\left[\bar{\mathbf{J}}\mathbf{D}\mathbf{p}\right]_i = \sum_{k=1}^{n} \frac{\partial y_k}{\partial p_i \partial p_j} \cdot \left(\frac{-p_k}{y_k b_k}\right) + \sum_{k=1}^{n} \frac{\partial y_k}{\partial p_i} \cdot \frac{\partial y_k}{\partial p_j} \cdot \left(\frac{p_k}{y_k^2 b_k}\right) + \frac{\partial y_j}{\partial p_i}\left(\frac{-1}{y_j b_j}\right).$$

If $j = n+1$, the expression is the same but the last term is omitted since the summation is from 1 to n only. In vector notation,

$$\frac{\partial}{\partial \bar{\mathbf{p}}}\bar{\mathbf{J}}\mathbf{D}\mathbf{p} = \mathbf{D}\mathbf{p} \cdot \bar{\mathcal{H}} + \bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \begin{bmatrix}\bar{\mathbf{J}}\mathbf{D} & 0\end{bmatrix}.$$

This completes the derivation. □

B.3.3 Evaluating the Tensor Product

The next lemma gives an alternate expression for the last two terms in the expression for the Hessian we have just derived.

Lemma B.3. *For NL models, denote the Hessian of y_k with respect to the vector of probabilities $\bar{\mathbf{p}}$ by $\bar{\mathcal{H}}^k$. Then the tensor product $\mathbf{D}\mathbf{p} \cdot \bar{\mathcal{H}}$ evaluates to*

$$\mathbf{D}\mathbf{p} \cdot \bar{\mathcal{H}} = \bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top - \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} - \mathbf{N}\mathbf{Q}^{-1}\mathbf{W}\mathbf{N}^\top, \tag{B.5}$$

where

$$\mathbf{W} = \mathbf{V}\mathbf{N}^\top\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{N}\mathbf{Q}^{-1}.$$

Proof. We begin by showing that the Hessians of each y_k with respect to the vector of probabilities and nest probabilitiess $(\bar{\mathbf{p}}, \mathbf{q})$ is

$$\mathcal{H}^k_{\bar{\mathbf{p}}\mathbf{q}} = \mathbf{J}^k_{\bar{\mathbf{p}}\mathbf{q}} \left(\frac{1}{y_k}\right) (\mathbf{J}^k_{\bar{\mathbf{p}}\mathbf{q}})^\top - y_k \begin{bmatrix} \bar{\mathbf{P}}^{-1} & \\ & \mathbf{Q}^{-1} \end{bmatrix} \operatorname{diag}\left(\mathbf{J}^k_{\bar{\mathbf{p}}\mathbf{q}}\right), \tag{B.6}$$

where $\mathbf{J}^k_{\bar{\mathbf{p}}\mathbf{q}} = \frac{\partial y_k}{\partial(\bar{\mathbf{p}},\mathbf{q})}$ denotes the k^{th} column of $\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}$. As shown in Lemma 3.4, the first partial derivatives of y^k have the form of monomials in the four variables p_k, Q_{m_k}, p_{n+1} and Q_{m^*}. Taking the example in (B.1),

$$\frac{\partial y_k}{\partial p_k} = \frac{1}{\mu_k p_k} y_k,$$

the second partial derivative is, if $j \neq k, 0$

$$\frac{\partial^2 y_k}{\partial p_k \partial p_j} = \frac{1}{\mu_k p_k} \frac{\partial y_k}{\partial p_j} = \frac{1}{y_k} \frac{\partial y_k}{\partial p_k} \frac{\partial y_k}{\partial p_j}.$$

A similar statements holds for any off-diagonal term of the Hessian. This yields the first term of (B.6). Note that for the example presented here, one of $\frac{\partial y_k}{\partial p_k}$ or $\frac{\partial y_k}{\partial p_j}$ is zero, so the (k, j) entry of the Hessian is zero. This is not the case, say, for $\frac{\partial^2 y_k}{\partial p_k \partial p_{n+1}}$, since y^k depends on both p_k and p_{n+1}. Because y_k depends only on four variables, the Hessian is sparse.

For the diagonal terms of the Hessian, we have, for example,

$$\frac{\partial^2 y_k}{\partial p_k \partial p_k} = \frac{1}{\mu_k p_k} \frac{\partial y_k}{\partial p_j} - \frac{1}{\mu_k p_k} y_k \frac{1}{p_k} = \frac{1}{y_k} \left(\frac{\partial y_k}{\partial p_k}\right)^2 - \frac{1}{p_k} \frac{\partial y_k}{\partial p_k}.$$

That is, we must subtract and additional term, whence the second term in (B.6).

We now show that

$$\mathbf{Dp} \cdot \mathcal{H}_{\bar{\mathbf{p}}\mathbf{q}} = -\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} \mathbf{Y}^{-1} \mathbf{P} \mathbf{B}^{-1} \mathbf{Y}^{-1} \mathbf{J}^\top_{\bar{\mathbf{p}}\mathbf{q}} + \begin{bmatrix} \bar{\mathbf{P}}^{-1} & \\ & \mathbf{Q}^{-1} \end{bmatrix} \operatorname{diag}\left(\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} \mathbf{Y}^{-1} \mathbf{B}^{-1} \mathbf{p}\right). \tag{B.7}$$

Recall that

$$\mathbf{Dp}\cdot\mathcal{H}_{\bar{\mathbf{p}}\mathbf{q}} = -\left(\mathbf{B}^{-1}\mathbf{Y}^{-1}\mathbf{p}\right)\cdot\mathcal{H}_{\bar{\mathbf{p}}\mathbf{q}} = -\sum_k \frac{p_k}{y_k b_k}\mathcal{H}^k_{\bar{\mathbf{p}}\mathbf{q}}.$$

The two terms in (B.7) correspond to the weighted sums of the terms in (B.6). The factor $\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}$ in (B.7) is the diagonal matrix with entries

$$\frac{p_k}{y_k b_k}\cdot\frac{1}{y_k} = \frac{p_k}{y_k^2 b_k}.$$

These entries correspond to the weights in the summation, and the $1/y_k$ factors in (B.6). The second term of (B.7) is the diagonal matrix from the second term of (B.6), times the diagonal matrix with the weighted sum of the columns of $\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}$ on the diagonal.

We now apply the chain rule for Hessians of Proposition B.1 to each term in the summation for (B.7) to obtain

$$\mathbf{Dp}\cdot\bar{\mathcal{H}} = \sum_k \frac{p_k}{y_k b_k}\bar{\mathcal{H}}^k = \sum_k \frac{p_k}{y_k b_k}\begin{bmatrix}\mathbf{I} & \mathbf{N}\end{bmatrix}\mathcal{H}^k_{\bar{\mathbf{p}}\mathbf{q}}\begin{bmatrix}\mathbf{I}\\ \mathbf{N}^\top\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{N}\end{bmatrix}\left(\mathbf{Dp}\cdot\mathcal{H}_{\bar{\mathbf{p}}\mathbf{q}}\right)\begin{bmatrix}\mathbf{I}\\ \mathbf{N}^\top\end{bmatrix}.$$

There are no second order terms, since the relationship between the vectors $(\bar{\mathbf{p}},\mathbf{q})$ and $\bar{\mathbf{p}}$ is linear. Then

$$\begin{aligned}
\mathbf{Dp}\cdot\bar{\mathcal{H}} &= -\begin{bmatrix}\mathbf{I} & \mathbf{N}\end{bmatrix}\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\mathbf{J}^\top_{\bar{\mathbf{p}}\mathbf{q}}\begin{bmatrix}\mathbf{I}\\ \mathbf{N}^\top\end{bmatrix}\\
&\quad+\begin{bmatrix}\mathbf{I} & \mathbf{N}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{P}}^{-1} & \\ & \mathbf{Q}^{-1}\end{bmatrix}\operatorname{diag}\left(\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}\mathbf{Y}^{-1}\mathbf{B}^{-1}\mathbf{p}\right)\begin{bmatrix}\mathbf{I}\\ \mathbf{N}^\top\end{bmatrix}\\
&= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \begin{bmatrix}\mathbf{I} & \mathbf{N}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{P}}^{-1} & \\ & \mathbf{Q}^{-1}\end{bmatrix}\operatorname{diag}\left(\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}\mathbf{Y}^{-1}\mathbf{B}^{-1}\mathbf{p}\right)\begin{bmatrix}\mathbf{I}\\ \mathbf{N}^\top\end{bmatrix}
\end{aligned}$$

In order to simplify the last term, we use the extended matrix $\mathbf{B}^{-1}$ defined in Section 3.7. We may rewrite

$$\mathbf{M}^\top\mathbf{B}^{-1}\mathbf{p} = \mathbf{M}^\top\mathbf{B}^{-1}\mathbf{P}\mathbf{e} = \mathbf{M}^\top\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e} = (\mathbf{I}-\mathbf{e}_{n+1}\mathbf{e}^\top)\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e} = \bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e},$$

because we chose b_{n+1} such that

$$\mathbf{e}^\top \bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e} = \mathbf{e}^\top \mathbf{B}^{-1}\mathbf{p} + b_{n+1}^{-1} p_{n+1} = 0.$$

Substituting the value of $\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}$ from (B.2), we have

$$\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}}\mathbf{Y}^{-1}\mathbf{B}^{-1}\mathbf{p} = \begin{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} \\ \mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top \end{bmatrix} \mathbf{M}^\top \mathbf{B}^{-1}\mathbf{p} = \begin{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} \\ \mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top \end{bmatrix} \bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e}$$

and then, using the two facts stated after the equation,

$$\begin{aligned}
\mathbf{D}\mathbf{p} \cdot \bar{\mathcal{H}} &= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \begin{bmatrix} \mathbf{I} & \mathbf{N} \end{bmatrix} \operatorname{diag}\left(\begin{bmatrix} \bar{\mathbf{U}}\bar{\mathbf{P}}^{-2} \\ \mathbf{V}\mathbf{Q}^{-2}\mathbf{N}^\top \end{bmatrix} \bar{\mathbf{B}}^{-1}\bar{\mathbf{P}} \right) \begin{bmatrix} \mathbf{I} \\ \mathbf{N}^\top \end{bmatrix} \\
&= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \bar{\mathbf{U}}\bar{\mathbf{P}}^{-2}\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}} + \mathbf{N}\operatorname{diag}\left(\mathbf{V}\mathbf{Q}^{-2}\mathbf{N}^\top\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e}\right)\mathbf{N}^\top \\
&= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} + \mathbf{N}\mathbf{Q}^{-1}\mathbf{V}\operatorname{diag}\left(\mathbf{N}^\top\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{e}\right)\mathbf{Q}^{-1}\mathbf{N}^\top \\
&= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} + \mathbf{N}\mathbf{Q}^{-1}\mathbf{V}\mathbf{N}^\top\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{N}\mathbf{Q}^{-1}\mathbf{N}^\top \\
&= -\bar{\mathbf{J}}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\bar{\mathbf{J}}^\top + \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} + \mathbf{N}\mathbf{Q}^{-1}\mathbf{W}\mathbf{N}^\top.
\end{aligned}$$

In the third equality, we use that for any vector $\mathbf{x}$ and diagonal matrix $\mathbf{V}$,

$$\operatorname{diag}(\mathbf{V}\mathbf{x}) = \mathbf{V}\operatorname{diag}(\mathbf{x}) = \operatorname{diag}(\mathbf{x})\mathbf{V}.$$

For the fourth equality, we note that for any vector $\mathbf{x}$,

$$\operatorname{diag}\left(\mathbf{N}^\top\mathbf{x}\right) = \mathbf{N}^\top\operatorname{diag}(\mathbf{x})\mathbf{N},$$

because each row of $\mathbf{N}$ is all zero except for one entry.

Applying the chain rule to each term of the sum, as above,

$$\begin{aligned}
\mathbf{D}\mathbf{p} \cdot \mathcal{H} &= \mathbf{M}(\mathbf{D}\mathbf{p} \cdot \bar{\mathcal{H}})\mathbf{M}^\top \\
&= -\mathbf{J}\mathbf{Y}^{-1}\mathbf{P}\mathbf{B}^{-1}\mathbf{Y}^{-1}\mathbf{J}^\top + \mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} + \mathbf{N}\mathbf{Q}^{-1}\mathbf{W}\mathbf{N}^\top\right)\mathbf{M}^\top.
\end{aligned}$$

This completes the proof. $\square$

B.3.4 Concavity of the Profit

Theorem 3.5. *Under NL models, the Hessian of* $\Pi(\mathbf{p})$ *is*

$$\mathbf{H} = \mathbf{J_z} + \mathbf{J_z}^\top - \tilde{\mathbf{J}}_\mathbf{z}$$

where

$$\tilde{\mathbf{J}}_\mathbf{z} = -\mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}\bar{\mathbf{B}}^{-1} + \mathbf{N}\mathbf{Q}^{-1}\mathbf{W}\mathbf{N}^\top\right)\mathbf{M}^\top.$$

and the matrix $\mathbf{W}$ *is related to* $\mathbf{V}$ *by*

$$\mathbf{W} = \mathbf{V}\mathbf{N}^\top\bar{\mathbf{B}}^{-1}\bar{\mathbf{P}}\mathbf{N}\mathbf{Q}^{-1}.$$

If $\frac{1}{2} \leq b_1', b_2', \dots, b_n' \leq 1$, *then* $\Pi(\mathbf{p})$ *is strictly concave.*

Proof. Substituting (B.5) into (B.4) yields the equality. The following matrix is symmetric, because all the matrices are diagonal except for $\mathbf{M}$ and $\mathbf{N}$:

$$\mathbf{J_z}\mathbf{B} = \mathbf{J}\mathbf{D}\mathbf{B} = \mathbf{J}(-\mathbf{Y}^{-1}\mathbf{B}^{-1})\mathbf{B} = -\mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1} + \mathbf{N}\mathbf{V}\mathbf{Q}^{-1}\mathbf{N}^\top\right)\mathbf{M}^\top.$$

Adding and subtracting this quantity to the Hessian twice, we have

$$\begin{aligned}\mathbf{H} &= \mathbf{J_z}(\mathbf{I}-\mathbf{B}) + (\mathbf{I}-\mathbf{B})\mathbf{J_z}^\top + (\mathbf{J_z}\mathbf{B} + \mathbf{B}\mathbf{J_z}^\top - \tilde{\mathbf{J}}) \\ &= \mathbf{J_z}(\mathbf{I}-\mathbf{B}) + (\mathbf{I}-\mathbf{B})\mathbf{J_z}^\top + (2\mathbf{J_z}\mathbf{B} - \tilde{\mathbf{J}})\end{aligned}$$

The matrix $(\mathbf{I}-\mathbf{B})$ is a strictly positive diagonal matrix by the assumption on the entries of $\mathbf{B}$. Then by Proposition 3.3, the first two terms are negative inverse M-matrices, and negative-definite. We need only show that the last term is negative semi-definite to show strict concavity of the profit.

We first prove an equality we shall use immediately below. Observe that $\mathbf{N}^\top\bar{\mathbf{P}}\mathbf{N} = \mathbf{Q}$, since $Q_{mm} = Q_m = \sum_{i=1}^{n+i} \alpha_{im}p_m$ is just the total choice probability in nest m.

Then

$$\begin{aligned}\mathbf{W} - 2\mathbf{V} &= \mathbf{V}(\mathbf{N}^\top\bar{\mathbf{P}}\bar{\mathbf{B}}^{-1}\mathbf{N}\mathbf{Q}^{-1} - 2\mathbf{I})\\ &= \mathbf{V}(\mathbf{N}^\top\bar{\mathbf{P}}\bar{\mathbf{B}}^{-1}\mathbf{N}\mathbf{Q}^{-1} - 2\mathbf{N}^\top\bar{\mathbf{P}}\mathbf{N}\mathbf{Q}^{-1})\\ &= \mathbf{V}\mathbf{N}^\top\bar{\mathbf{P}}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})\mathbf{N}\mathbf{Q}^{-1}\\ &= \mathbf{N}^\top(\mathbf{I} - \bar{\mathbf{U}})\bar{\mathbf{P}}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})\mathbf{N}\mathbf{Q}^{-1}.\end{aligned}$$

For the last line, we have used the fact that $\mathbf{V}\mathbf{N}^\top = \mathbf{N}^\top(\mathbf{I} - \bar{\mathbf{U}})$. To see this, observe that if $N_{im} = 1$, then product i in nest m, $V_{mm} = 1 - \frac{1}{\mu_{m_i}} = 1 - \bar{U}_{ii} \geq 0$.

Expanding the last term of the Hessian matrix $\mathbf{H}$, we have

$$\begin{aligned}2\mathbf{J_z}\mathbf{B} - \tilde{\mathbf{J}} &= \mathbf{M}\left(\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I}) + \mathbf{N}(\mathbf{W} - 2\mathbf{V})\mathbf{Q}^{-1}\mathbf{N}^\top\right)\mathbf{M}^\top\\ &= \mathbf{M}\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})\mathbf{M}^\top + \mathbf{M}\mathbf{N}\mathbf{Q}^{-1}(\mathbf{W} - 2\mathbf{V})\mathbf{N}^\top\mathbf{M}^\top\\ &= \mathbf{M}\bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})\mathbf{M}^\top +\\ &\qquad \mathbf{M}\mathbf{N}\mathbf{Q}^{-1}\mathbf{N}^\top(\mathbf{I} - \bar{\mathbf{U}})(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})\bar{\mathbf{P}}\mathbf{N}\mathbf{Q}^{-1}\mathbf{N}^\top\mathbf{M}^\top \qquad \text{(B.8)}\end{aligned}$$

By the assumption on the entries of $\mathbf{B}$, then diagonal entries of $\mathbf{B}^{-1}$ are less than 2. The same holds for the entries of $\bar{\mathbf{B}}^{-1}$ because b_{n+1} was chosen to be *negative.* Then the matrix $(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})$ has strictly negative diagonal entries. On the other hand, the matrix $(\mathbf{I} - \bar{\mathbf{U}})$ has non-negative entries. The diagonal matrices $\bar{\mathbf{U}}$, $\bar{\mathbf{P}}^{-1}$ and $\mathbf{Q}^{-1}$ have positive diagonal entries.

It follows that (B.8) is negative semi-definite, because both terms are of the form $\mathbf{A}^\top\mathbf{C}\mathbf{A}$, where $\mathbf{C}$ is a negative diagonal matrix. In fact, the matrix is also negative definite because, for the first term, the matrix $\mathbf{A} = \mathbf{M}$ has full rank and $\mathbf{C} = \bar{\mathbf{U}}\bar{\mathbf{P}}^{-1}(\bar{\mathbf{B}}^{-1} - 2\mathbf{I})$ is negative definite. It follows that the profit is globally strictly concave in $\mathbf{p}$. □

Appendix C

Derivations for GEV models

C.1 Almost-Concavity of the Reformulation

C.1.1 Partial Derivatives of the Unnormalized Demands

Lemma 4.12. *The Hessian matrix of q_k with respect to* $\mathbf{y}$*, for $k = 1, \ldots, n+1$, is*

$$\bar{\mathcal{K}}^k_{ij} = y_k \bar{\mathcal{G}}^k + \bar{\mathbf{G}}^k \mathbf{e}_k^\top + \mathbf{e}_k (\bar{\mathbf{G}}^k)^\top,$$

where $\bar{\mathbf{G}}^k = \bar{\mathbf{G}}\mathbf{e}_k$ is the k^{th} column (or row, transposed) of the symmetric matrix $\bar{\mathbf{G}}$. Removing the last row and column, the Hessian with respect to $\mathbf{y}$ *is*

$$\mathcal{K}^k_{ij} = y_k \mathcal{G}^k + \mathbf{G}^k \mathbf{e}_k^\top + \mathbf{e}_k (\mathbf{G}^k)^\top.$$

Moreover, these Hessian matrices $\bar{\mathcal{K}}^k$ are related to the Hessian matrix $\bar{\mathbf{G}}$ of $G(\cdot)$ by

$$\mathbf{e} \cdot \bar{\mathcal{K}} = \sum_{k=1}^{n+1} \bar{\mathcal{K}}^k = \mu \bar{\mathbf{G}}.$$

Proof. From the scalar product rule, with $i \neq j$,

$$\frac{\partial q_j}{\partial y_i} = \frac{\partial y_j G_j(\mathbf{y})}{\partial y_i} = y_j G_{ij}(\mathbf{y}), \qquad \text{and} \qquad \frac{\partial q_j}{\partial y_j} = \frac{\partial y_j G_j(\mathbf{y})}{\partial y_j} = G_j(\mathbf{y}) + y_j G_{jj}(\mathbf{y}).$$

These are the entries of $\mathbf{L}^{-1}$.

Differentiating again, with $i \neq \ell \neq j$,

$$\frac{\partial^2 q_j}{\partial y_i \partial y_\ell} = y_j G_{ij\ell}(\mathbf{y}), \qquad \frac{\partial^2 q_j}{\partial y_j \partial y_\ell} = G_{j\ell}(\mathbf{y}) + y_j G_{ij\ell}(\mathbf{y}), \qquad \text{and}$$

$$\frac{\partial^2 q_\ell}{\partial y_\ell \partial y_\ell} = \frac{\partial}{\partial y_\ell}\frac{\partial q_\ell}{\partial y_\ell} = \frac{\partial}{\partial y_\ell}\left(G_\ell(\mathbf{y}) + y_\ell G_{\ell\ell}(\mathbf{y})\right) = 2G_{\ell\ell}(\mathbf{y}) + y_j G_{\ell\ell\ell}(\mathbf{y}).$$

That is,

$$\mathcal{K}^j_{i\ell} = y_j \mathcal{G}^j + \mathbf{G}^j \mathbf{e}_j^\top + \mathbf{e}_j (\mathbf{G}^j)^\top,$$

where the matrix $\mathbf{G}^j \mathbf{e}_j^\top$ (or $\mathbf{e}_j(\mathbf{G}^j)^\top$) correspond to $\mathbf{G}$ with all but the j^{th} column (or row) set to zero. Renaming the indices for consistency, we obtain the desired expression.

Applying Corollary 4.7 of Euler's theorem,

$$\mathbf{e} \cdot \bar{\mathcal{K}} = (\mu - 2)\bar{\mathbf{G}} + \bar{\mathbf{G}} + \bar{\mathbf{G}} = \mu \bar{\mathbf{G}}, \tag{C.1}$$

as claimed. $\square$

C.1.2 The Hessian of the Unnormalized Profit

Lemma 4.14. *The gradient of $\Psi(\mathbf{y})$ with respect to $\mathbf{q}$ is*

$$\frac{\partial \Psi}{\partial \mathbf{q}} = \mathbf{z} + \mathbf{LDq}$$

and its Hessian is

$$\frac{\partial^2 \Psi}{\partial \mathbf{q} \partial \mathbf{q}} = \mathbf{L}\left(\mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^\top - \mathbf{LDq} \cdot \mathcal{K} - \mathbf{D}\,\mathrm{diag}\left(\mathbf{g}\right)\right)\mathbf{L}^\top.$$

Proof. The Jacobian of $\mathbf{z}$ with respect to $\mathbf{q}$ is $\mathbf{LD}$ by the vector chain rule. Therefore the desired expression for the gradient is

$$\frac{\partial \Psi}{\partial \mathbf{q}} = \frac{\partial \mathbf{q}^\top \mathbf{z}}{\partial \mathbf{q}} = \mathbf{z} + \mathbf{LDq}.$$

Applying the chain rule, we have

$$\frac{\partial \Psi}{\partial \mathbf{y}} = \mathbf{L}^{-1}\frac{\partial \Psi}{\partial \mathbf{q}} = \mathbf{L}^{-1}\mathbf{z} + \mathbf{D}\mathbf{q}. = \sum_k z_k \mathbf{L}^{-1}_{\cdot,k} + \sum_k q_k \mathbf{D}_{\cdot,k},$$

where $\mathbf{D}_{\cdot,k}$ and $\mathbf{L}^{-1}_{\cdot,k}$ represent the k^{th} columns of the respective matrices. Differentiating with respect to $\mathbf{y}$, we apply the vector product rule to each term to obtain

$$\begin{aligned}
\frac{\partial^2 \Psi}{\partial \mathbf{y}\partial \mathbf{y}} &= \left(\mathbf{L}^{-1}\frac{\partial \mathbf{z}}{\partial \mathbf{y}} + \sum_k z_k \frac{\partial}{\partial \mathbf{y}}\mathbf{L}^{-1}_{\cdot,k}\right) + \left(\mathbf{D}\frac{\partial \mathbf{q}}{\partial \mathbf{y}} + \sum_k q_k \frac{\partial}{\partial \mathbf{y}}\mathbf{D}_{\cdot,k}\right) \\
&= \left(\mathbf{L}^{-1}\mathbf{D} + \sum_k z_k \mathcal{K}^k\right) + \left(\mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} + \operatorname{diag}\left(\left[\frac{q_1}{b_1 y_1^2}, \ldots, \frac{q_n}{b_n y_n^2}\right]\right)\right) \\
&= \mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} + \mathbf{z}\cdot\mathcal{K} - \mathbf{D}\operatorname{diag}(\mathbf{g}).
\end{aligned}$$

In the second line, we have replaced the last term by a diagonal matrix. Note that because $\mathbf{D}$ is a diagonal matrix, the vector $\mathbf{D}_{\cdot,k}$ is all zero except for the k^{th} entry, $D_{k,k} = \frac{-1}{b_k y_k}$. The Jacobian with respect to $\mathbf{y}$ is the matrix of all zeros with (k,k) entry $\partial D_{k,k}/\partial y_k = \frac{1}{b_k y_k^2}$. The sum of these matrices weighted by the entries of $\mathbf{q}$ is then the diagonal matrix with entries $\frac{q_k}{b_k y_k^2} = \frac{y_k G_k(\mathbf{y})}{b_k y_k^2} = -G_k(\mathbf{y})\frac{-1}{b_k y_k}$, whence the expression in the last line. The rest of the derivation consists of simple substitution.

Applying the second chain rule of Proposition B.1, proved in Appendix B.3.1, we obtain the Hessian matrix with respect to $\mathbf{q}$:

$$\begin{aligned}
\frac{\partial^2 \Psi}{\partial \mathbf{q}\partial \mathbf{q}} &= \mathbf{L}\left(\frac{\partial^2 \Psi}{\partial \mathbf{y}\partial \mathbf{y}} - \mathbf{L}\frac{\partial \Psi}{\partial \mathbf{y}}\cdot\mathcal{K}\right)\mathbf{L}^{\top} \\
&= \mathbf{L}\left(\frac{\partial^2 \Psi}{\partial \mathbf{y}\partial \mathbf{y}} - \frac{\partial \Psi}{\partial \mathbf{q}}\cdot\mathcal{K}\right)\mathbf{L}^{\top} \\
&= \mathbf{L}\left(\frac{\partial^2 \Psi}{\partial \mathbf{y}\partial \mathbf{y}} - (\mathbf{z} + \mathbf{L}\mathbf{D}\mathbf{q})\cdot\mathcal{K}\right)\mathbf{L}^{\top} \\
&= \mathbf{L}\left(\mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} + \mathbf{z}\cdot\mathcal{K} - \mathbf{D}\operatorname{diag}(\mathbf{g}) - (\mathbf{z} + \mathbf{L}\mathbf{D}\mathbf{q})\cdot\mathcal{K}\right)\mathbf{L}^{\top} \\
&= \mathbf{L}\left(\mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^{\top} - \mathbf{L}\mathbf{D}\mathbf{q}\cdot\mathcal{K} - \mathbf{D}\operatorname{diag}(\mathbf{g})\right)\mathbf{L}^{\top} \\
&= \mathbf{D}\mathbf{L}^{\top} + \mathbf{L}\mathbf{D} - \mathbf{L}\left(\mathbf{L}\mathbf{D}\mathbf{q}\cdot\mathcal{K} + \mathbf{D}\operatorname{diag}(\mathbf{g})\right)\mathbf{L}^{\top}.
\end{aligned}$$

$\square$

C.1.3 The Sub-stochastic Matrix S

Lemma C.1. *The inverse M-matrix $\bar{\mathbf{S}} = \bar{\mathbf{L}}\,\mathrm{diag}\,(\bar{\mathbf{g}})$ satisfies $\bar{\mathbf{S}} \geq \mathbf{0}$ and $\bar{\mathbf{S}}\mathbf{e} = \frac{1}{\mu}\mathbf{e}$ (it is row-stochastic when scaled by μ). Moreover, for any diagonal matrix $\bar{\mathbf{B}} \in \mathbb{R}^{(n+1)\times(n+1)}$,*

$$\bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} \cdot \bar{\mathcal{K}} = -\bar{\mathbf{L}}\bar{\mathbf{D}}\bar{\mathbf{q}} \cdot \bar{\mathcal{K}} = \bar{\mathbf{G}} + (\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I}))\mathbf{e} \cdot \bar{\mathcal{K}}.$$

Proof. The inverse of $\bar{\mathbf{S}}$ is

$$\bar{\mathbf{S}}^{-1} = (\bar{\mathbf{L}}\,\mathrm{diag}\,(\bar{\mathbf{g}}))^{-1} = \mathrm{diag}\,(\bar{\mathbf{g}})^{-1}\,\bar{\mathbf{L}}^{-1}$$

where $\bar{\mathbf{L}}^{-1}$ is an M-matrix by Lemma 4.8. Then $\bar{\mathbf{S}}^{-1}$ is also an M-matrix because we have scaled by a positive diagonal matrix [see, for example, 43, Theorem 1.2.3]. Morevoer, from Equation (4.4) in the proof of Lemma 4.8,

$$\bar{\mathbf{S}}^{-1}\mathbf{e} = \mathrm{diag}\,(\bar{\mathbf{g}})^{-1}\,\bar{\mathbf{L}}^{-1}\mathbf{e} = \mu\mathbf{e}.$$

The inverse of an M-matrix has all entries non-negative [see, for example, 43, 72]. Thefore $\bar{\mathbf{S}}$ has non-negative entries and,

$$\bar{\mathbf{S}}\mathbf{e} = \bar{\mathbf{S}}\left(\frac{1}{\mu}\bar{\mathbf{S}}^{-1}\mathbf{e}\right) = \frac{1}{\mu}\mathbf{e},$$

so $\mu\bar{\mathbf{S}}$ is row stochastic.

From the last part of Lemma 4.12

$$(\bar{\mathbf{S}}\mathbf{e}) \cdot \bar{\mathcal{K}} = \frac{1}{\mu}(\bar{\mathbf{e}} \cdot \bar{\mathcal{K}}) = \bar{\mathbf{G}} \tag{C.2}$$

Then, because $\mathbf{D} = \bar{\mathbf{Y}}^{-1}\bar{\mathbf{B}}^{-1}$,

$$\begin{aligned}
-\bar{\mathbf{L}}\bar{\mathbf{D}}\bar{\mathbf{q}} \cdot \bar{\mathcal{K}} &= \bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} \cdot \bar{\mathcal{K}} \\
&= (\bar{\mathbf{S}}\mathbf{e} \cdot \bar{\mathcal{K}}) + (\bar{\mathbf{S}}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \bar{\mathcal{K}}) \\
&= \bar{\mathbf{G}} + (\bar{\mathbf{S}}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \bar{\mathcal{K}}).
\end{aligned}$$

This completes the proof. □

Lemma 4.15. *The inverse M-matrix $\mu\mathbf{S}$ is sub-stochastic. That is, its entries are non-negative and its rows sum to less than one. Moreover,*

$$\begin{aligned}
\mathbf{S}\mathbf{B}^{-1}\mathbf{e} \cdot \mathcal{K} = \mathbf{L}\mathbf{B}^{-1}\mathbf{g} \cdot \mathcal{K} &= -\mathbf{L}\mathbf{D}\mathbf{q} \cdot \mathcal{K} \\
&= \mathbf{G} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \mathcal{K} - \tilde{\mathbf{G}},
\end{aligned}$$

where $\tilde{\mathbf{G}}$ is the upper-left $n \times n$ sub-matrix of the tensor product

$$\frac{1}{\mu}\left(\frac{1}{S_{n+1,n+1}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right) \cdot \bar{\mathcal{K}}.$$

Proof. Define $\mathbf{S} = (\mathbf{L}\,\mathrm{diag}\,(\mathbf{g}))^{-1})$ analogously to $\bar{\mathbf{S}}$ of Lemma C.1. That is,

$$\mathbf{S}^{-1} = (\mathbf{L}\,\mathrm{diag}\,(\mathbf{g}))^{-1} = \mathrm{diag}\,(\mathbf{g})^{-1}\,\mathbf{L}^{-1} = \left(\mathbf{I} + \mathrm{diag}\,(\mathbf{g})^{-1}\,\mathbf{G}\mathbf{Y}\right),$$

Because $\mathbf{L}^{-1}$ and $\bar{\mathbf{L}}^{-1}$ are diagonally dominant M-matrices by Lemma 4.8, their diagonal elements are positive and their off-diagonal elements are non-positive. The same holds for $\mathbf{S}^{-1}$ and $\bar{\mathbf{S}}^{-1}$ because they result from a positive scaling of M-matrices and are therefore also M-matrices [43, Theorem 1.2.3]. We may write the componentwise inequality between two M-matrices

$$\bar{\mathbf{S}}^{-1} = \begin{bmatrix} \mathbf{S}^{-1} & \mathbf{v} \\ \mathbf{u}^{\top} & w \end{bmatrix} \leq \begin{bmatrix} \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{0}^{\top} & w \end{bmatrix},$$

where $\mathbf{u}, \mathbf{v} \leq 0$, and $w > 0$. Then, by a well-known result [see, for example, 43, Theorem 1.8], inverting both sides yields

$$\bar{\mathbf{S}} \geq \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0}^\top & \frac{1}{w} \end{bmatrix}.$$

We conclude that $\mu\mathbf{S}$ is a sub-stochastic matrix by Lemma C.1, and because $\mathbf{S}$ also has non-negative entries since it is also an inverse M-matrix.

We will now make the relationship between $\mathbf{S}$ and $\bar{\mathbf{S}}$ more precise. We define the new matrix

$$\hat{\mathbf{S}}^{-1} \triangleq \begin{bmatrix} \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & -w \end{bmatrix} = \bar{\mathbf{S}}^{-1} - \mathbf{e}_{n+1}\bar{\mathbf{u}}^\top - \bar{\mathbf{v}}\mathbf{e}_{n+1}^\top = \bar{\mathbf{S}}^{-1} - \begin{bmatrix} \mathbf{e}_{n+1} & \bar{\mathbf{v}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{u}}^\top \\ \mathbf{e}_{n+1}^\top \end{bmatrix},$$

where $\bar{\mathbf{u}}^\top$ and $\bar{\mathbf{v}}^\top$ are now the entire last row and column of $\bar{\mathbf{S}}^{-1}$. Then by applying the Sherman-Morrison-Woodbury formula [38, p. 51], we obtain,

$$\hat{\mathbf{S}} = \bar{\mathbf{S}} + \bar{\mathbf{S}} \begin{bmatrix} \mathbf{e}_{n+1} & \bar{\mathbf{v}} \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} \bar{\mathbf{u}}^\top \\ \mathbf{e}_{n+1}^\top \end{bmatrix} \bar{\mathbf{S}} = \bar{\mathbf{S}} + \begin{bmatrix} \bar{\mathbf{S}}\mathbf{e}_{n+1} & \mathbf{e}_{n+1} \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} \mathbf{e}_{n+1}^\top \\ \mathbf{e}_{n+1}^\top \bar{\mathbf{S}} \end{bmatrix},$$

where we have used the fact that $\mathbf{u}^\top$ and $\mathbf{v}$ are rows and columns of the inverse $\bar{\mathbf{S}}^{-1}$. The matrix $\mathbf{C} \in \mathbb{R}^{2\times 2}$ is

$$\begin{aligned}
\mathbf{C} &= \mathbf{I} - \begin{bmatrix} \bar{\mathbf{u}}^\top \\ \mathbf{e}_{n+1}^\top \end{bmatrix} \bar{\mathbf{S}} \begin{bmatrix} \mathbf{e}_{n+1} & \bar{\mathbf{v}} \end{bmatrix} = \mathbf{I} - \begin{bmatrix} \bar{\mathbf{u}}^\top \bar{\mathbf{S}}\mathbf{e}_{n+1} & \bar{\mathbf{u}}^\top \bar{\mathbf{S}}\bar{\mathbf{v}} \\ \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\mathbf{e}_{n+1} & \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\bar{\mathbf{v}} \end{bmatrix} \\
&= \mathbf{I} - \begin{bmatrix} \mathbf{e}_{n+1}^\top \mathbf{e}_{n+1} & \bar{\mathbf{u}}^\top \bar{\mathbf{S}}\bar{\mathbf{S}}^{-1}\bar{\mathbf{S}}\bar{\mathbf{v}} \\ \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\mathbf{e}_{n+1} & \mathbf{e}_{n+1}^\top \mathbf{e}_{n+1} \end{bmatrix} = \mathbf{I} - \begin{bmatrix} 1 & \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}^{-1}\mathbf{e}_{n+1} \\ \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\mathbf{e}_{n+1} & 1 \end{bmatrix} \\
&= \mathbf{I} - \begin{bmatrix} 1 & w \\ \hat{w} & 1 \end{bmatrix} = \begin{bmatrix} & -w \\ -\hat{w} & \end{bmatrix},
\end{aligned}$$

where we denote the bottom-right entry of $\bar{\mathbf{S}}$ by $\hat{w} = \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\mathbf{e}_{n+1}$. (Recall that w is the bottom-right entry of $\bar{\mathbf{S}}^{-1}$). Substituting $\mathbf{C}^{-1}$ inverse back into the expression for

$\hat{\mathbf{S}}$, we have

$$\hat{\mathbf{S}} = \bar{\mathbf{S}} + \begin{bmatrix} \bar{\mathbf{S}}\mathbf{e}_{n+1} & \mathbf{e}_{n+1} \end{bmatrix} \begin{bmatrix} & -\frac{1}{w} \\ -\frac{1}{\hat{w}} & \end{bmatrix} \begin{bmatrix} \mathbf{e}_{n+1}^\top \\ \mathbf{e}_{n+1}^\top \bar{\mathbf{S}} \end{bmatrix} = \bar{\mathbf{S}} - \frac{1}{\hat{w}} \bar{\mathbf{S}}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top \bar{\mathbf{S}} - \frac{1}{w}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top$$

Notice that the last term is the matrix of all zeros except for the bottom right entry $\frac{-1}{w}$. Moving it to the left side we have that

$$\begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} = \hat{\mathbf{S}} + \frac{1}{w}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top = \bar{\mathbf{S}} - \frac{1}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top \bar{\mathbf{S}} \tag{C.3}$$

We remark that this again verifies the sub-stochasticity of $\mu\mathbf{S}$, since $\mathbf{S}$ in the left-hand side is equal to the stochastic $\mu\bar{\mathbf{S}}$ minus a matrix with all positive entries, with the last row and column removed. The matrix $\mathbf{S}$ remains non-negative since it is an inverse M-matrix.

Continuing with the proof, define the extended matrix

$$\bar{\mathbf{B}} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

Right-multiplying eq. (C.3) by $\bar{\mathbf{B}}^{-1}\mathbf{e}$ on both sides yields,

$$\begin{aligned} \begin{bmatrix} \mathbf{S}\mathbf{B}^{-1}\mathbf{e} \\ 0 \end{bmatrix} &= \bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} - \frac{1}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} \\ &= \bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} - \frac{\phi}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}, \end{aligned}$$

where $\phi = \mathbf{e}_{n+1}^\top \bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e}$ is a positive scalar. Taking the tensor product with $\bar{\mathcal{K}}$ on both sides we have that, for the left-hand side,

$$\begin{bmatrix} \mathbf{S}\mathbf{B}^{-1}\mathbf{e} \\ 0 \end{bmatrix} \cdot \bar{\mathcal{K}} = \begin{bmatrix} \mathbf{S}\mathbf{B}^{-1}\mathbf{e} \cdot \mathcal{K} & \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2^\top & \tilde{x}_3 \end{bmatrix},$$

for some unimportant quantities $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_1$ and $\tilde{x}_3$. This follows because $\mathcal{K}$ is simply $\bar{\mathcal{K}}$

with the last entry in each coordinate removed. That is, the left-hand side is simply the quantity of interest, $\mathbf{SB}^{-1}\mathbf{e} \cdot \mathcal{K}$ with an additional row and column.

Also multiplying the right-hand side by $\bar{\mathcal{K}}$, and then applying Lemma C.1, we have that

$$\left(\bar{\mathbf{S}}\bar{\mathbf{B}}^{-1}\mathbf{e} - \frac{\phi}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right) \cdot \bar{\mathcal{K}} = \bar{\mathbf{G}} + (\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e}) \cdot \bar{\mathcal{K}} - \frac{\phi}{\hat{w}}\left(\bar{\mathbf{S}}\mathbf{e}_{n+1} \cdot \bar{\mathcal{K}}\right). \tag{C.4}$$

If we now right-multiply both sides of eq. (C.3) by $\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e}$ instead, we obtain that

$$\begin{aligned} \begin{bmatrix} \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \\ 0 \end{bmatrix} &= \bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e} - \frac{1}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\mathbf{e}_{n+1}^\top\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e} \\ &= \bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e} - \frac{\bar{\phi}}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}, \end{aligned}$$

where $\bar{\phi} = \mathbf{e}_{n+1}^\top\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e}$ is a scalar. We can now substitute

$$\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I})\mathbf{e} = \begin{bmatrix} \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \\ 0 \end{bmatrix} + \frac{\bar{\phi}}{\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}$$

back into (C.4), to obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{SB}^{-1}\mathbf{e} \cdot \mathcal{K} & \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2^\top & \tilde{x}_3 \end{bmatrix} &= \bar{\mathbf{G}} + \begin{bmatrix} \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \\ 0 \end{bmatrix} \cdot \bar{\mathcal{K}} + \left(\left(\frac{\bar{\phi}}{\hat{w}} - \frac{\phi}{\hat{w}}\right)\bar{\mathbf{S}}\mathbf{e}_{n+1}\right) \cdot \bar{\mathcal{K}} \\ &= \bar{\mathbf{G}} + \begin{bmatrix} \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \\ 0 \end{bmatrix} \cdot \bar{\mathcal{K}} - \left(\frac{1}{\mu\hat{w}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right) \cdot \bar{\mathcal{K}}, \end{aligned}$$

where we used the fact that

$$\bar{\phi} - \phi = \mathbf{e}_{n+1}^\top\bar{\mathbf{S}}(\bar{\mathbf{B}}^{-1} - \mathbf{I} - \bar{\mathbf{B}}^{-1})\mathbf{e} = -\mathbf{e}_{n+1}^\top\bar{\mathbf{S}}\mathbf{e} = -\mathbf{e}_{n+1}^\top\left(\frac{1}{\mu}\mathbf{e}\right) = -\frac{1}{\mu}.$$

The statement of the lemma follows by considering the preceding equation with the last row and column removed. $\square$

C.1.4 Negative-definiteness of the Hessian

Theorem 4.16. *The Hessian of* $\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda\mu G(\mathbf{q})$ *is*

$$\frac{\partial^2\Gamma}{\partial\mathbf{q}\partial\mathbf{q}} = \mathbf{L}\left(\mathbf{A}_{concave} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{A}}\right)\mathbf{L}^\top.$$

If the price sensitivity parameters $b'_1, b'_2, \ldots, b'_n$ *belong to the range* $(\frac{1}{2}, 1)$, *then the matrix* $\mathbf{A}_{concave}$ *is negative definite. The matrix* $\tilde{\mathbf{A}}$ *is the upper-left* $n \times n$ *submatrix of the tensor product*

$$\frac{1}{\mu}\left(\mathbf{e} + (1+\lambda\mu)\left(\frac{1}{S_{n+1,n+1}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right)\right)\cdot\bar{\mathcal{K}},$$

where the vector $\bar{\mathbf{S}}\mathbf{e}_{n+1}$ *denotes the last column of the matrix* $\bar{\mathbf{S}}$ *and* $S_{n+1,n+1}$ *is its last entry.*

Proof. The Hessians of Ψ and G are given by Lemmas 4.13 and 4.14. They are of the form $\mathbf{L}\mathbf{A}_\Psi\mathbf{L}^\top$ and $\mathbf{L}\mathbf{A}_G\mathbf{L}^\top$ for matrices $\mathbf{A}_\Psi$ and $\mathbf{A}_G$, respectively. Write the Hessian matrix of F as $\mathbf{A} = \mathbf{A}_\Psi - \lambda\mu\mathbf{A}_G$ so we can consider the terms resulting from Ψ and G separately.

Taking the expression from Lemma 4.13, and then substituting in the result of Lemma 4.15 with $\mathbf{B} = \mathbf{I}$, we have

$$\begin{aligned}\mathbf{A}_G &= \mathbf{G} - \mathbf{L}\mathbf{g}\cdot\mathcal{K}\\ &= \mathbf{G} - \left(\mathbf{G} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}}\right)\\ &= \tilde{\mathbf{G}}\end{aligned}$$

We will use that

$$\mathbf{L}^{-1}\mathbf{D}\mathbf{B} = -\mathbf{L}^{-1}\mathbf{Y} = -\mathbf{G} - \mathbf{Y}^{-1}\operatorname{diag}(\mathbf{g}).$$

Taking the expression from Lemma 4.14, we consider the other term of the Hessian,

$$\begin{aligned}
\mathbf{A}_\Psi &= \mathbf{L}^{-1}\mathbf{D} + \mathbf{D}\left(\mathbf{L}^{-1}\right)^\top - \mathbf{LDq}\cdot\mathcal{K} - \mathbf{D}\,\mathrm{diag}\,(\mathbf{g}) \\
&= \mathbf{L}^{-1}\mathbf{D}(\mathbf{I}-\mathbf{B}) + (\mathbf{I}-\mathbf{B})\mathbf{D}\left(\mathbf{L}^{-1}\right)^\top - 2\mathbf{G} \\
&\qquad\qquad - 2\mathbf{Y}^{-1}\,\mathrm{diag}\,(\mathbf{g}) - \mathbf{LDq}\cdot\mathcal{K} + \mathbf{B}^{-1}\mathbf{Y}^{-1}\,\mathrm{diag}\,(\mathbf{g}) \\
&= \mathbf{L}^{-1}\mathbf{D}(\mathbf{I}-\mathbf{B}) + (\mathbf{I}-\mathbf{B})\mathbf{D}\left(\mathbf{L}^{-1}\right)^\top - 2\mathbf{G} - \mathbf{LDq}\cdot\mathcal{K} + (\mathbf{B}^{-1}-2\mathbf{I})\mathbf{Y}^{-1}\,\mathrm{diag}\,(\mathbf{g}) \\
&= \mathbf{L}^{-1}\mathbf{D}(\mathbf{I}-\mathbf{B}) + (\mathbf{I}-\mathbf{B})\mathbf{D}\left(\mathbf{L}^{-1}\right)^\top + (\mathbf{B}^{-1}-2\mathbf{I})\mathbf{Y}^{-1}\,\mathrm{diag}\,(\mathbf{g}) + \mathbf{A}^*_\Psi,
\end{aligned}$$

where we define the matrix $\mathbf{A}^*_\Psi$, to which we again apply Lemma 4.15:

$$\begin{aligned}
\mathbf{A}^*_\Psi &= -2\mathbf{G} - \mathbf{LDq}\cdot\mathcal{K} \\
&= -2\mathbf{G} + \left(\mathbf{G} + \mathbf{S}(\mathbf{B}^{-1}-\mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}}\right) \\
&= \mathbf{S}(\mathbf{B}^{-1}-\mathbf{I})\mathbf{e}\cdot\mathcal{K} - \mathbf{G} - \tilde{\mathbf{G}},
\end{aligned}$$

We now show that the terms of $\mathbf{A}_\Psi$ are negative semi-definite, except for this last term $\mathbf{A}^*_\Psi$. Recall that the diagonal elements of $\mathbf{B}$ are in the range $(\frac{1}{2}, 1)$ by assumption. The first two terms of $\mathbf{A}_\Psi$ are M-matrices scaled by the negative diagonal matrix $\mathbf{D}(\mathbf{I}-\mathbf{B})$, so they are also negative M-matrices and therefore negtaive-definite. The third term is negative diagonal since the diagonal elements of $\mathbf{B}$ are less than 2. We can collect the negative semi-definite terms of $\mathbf{A}$ to obtain

$$\begin{aligned}
\mathbf{A} &= \mathbf{A}_\Psi - \lambda\mu\mathbf{A}_G \\
&= \mathbf{A}_{concave} + \mathbf{A}^*_\Psi - \lambda\mu\mathbf{A}_G \\
&= \mathbf{A}_{concave} + \left(\mathbf{S}(\mathbf{B}^{-1}-\mathbf{I})\mathbf{e}\cdot\mathcal{K} - \mathbf{G} - \tilde{\mathbf{G}}\right) - \lambda\mu\tilde{\mathbf{G}} \\
&= \mathbf{A}_{concave} + \mathbf{S}(\mathbf{B}^{-1}-\mathbf{I})\mathbf{e}\cdot\mathcal{K} - \mathbf{G} - (1+\lambda\mu)\,\tilde{\mathbf{G}}
\end{aligned}$$

for a negative-definite matrix $\mathbf{A}_{concave}$. Using Lemmas 4.12 and 4.15, we can expand the terms in $\mathbf{G}$ and $\tilde{\mathbf{G}}$ to obtain

$$\mathbf{A} = \mathbf{A}_{concave} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \mathcal{K} - \frac{1}{\mu}\left(\mathbf{e} + (1 + \lambda\mu)\left(\frac{1}{S_{n+1,n+1}}\bar{\mathbf{S}}\mathbf{e}_{n+1}\right)\right) \cdot \bar{\mathcal{K}},$$

where it is understood that we use $\bar{\mathcal{K}}$ to denote only the upper-left $n \times n$ sub-matrices, without the last row and columns (but including the term $\bar{\mathcal{K}}^{n+1}$ in the tensor product). This completes the proof. $\square$

Appendix D

First-order Methods for Pricing

D.1 Uniqueness of the Solution

Lemma 5.3. *The Jacobian of $F'(\mathbf{z})$ is*

$$\frac{\partial F'}{\partial \mathbf{z}} = \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{B}\mathbf{L}^{\top} - \left(\tilde{\mathbf{L}}^{-1}\right)^{\top}\mathbf{L}^{\top} - \tilde{\mathbf{E}},$$

with the last term

$$\tilde{\mathbf{E}} = \mathbf{B}\mathbf{Y}\left(\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e} \cdot \mathcal{K} - (1+\lambda\mu)\tilde{\mathbf{G}}\right)\mathbf{L}^{\top},$$

and where the matrix

$$\left(\tilde{\mathbf{L}}^{-1}\right)^{\top} = (\mathrm{diag}\,(\mathbf{g}) + \mathbf{B}\mathbf{Y}\mathbf{G})$$

is equal to $\left(\mathbf{L}^{-1}\right)^{\top}$ *when* $\mathbf{B} = \mathbf{I}$.

Proof. Lemmas 4.13 and 4.14 give the Hessian of F with respect to $\mathbf{q}$. The following chain rule application, yields the Jacobian of $F(\mathbf{z})$ with respect to $\mathbf{z}$.

$$\begin{aligned}\frac{\partial F'}{\partial \mathbf{z}} &= \frac{\partial}{\partial \mathbf{z}}\left(\frac{\partial F}{\partial \mathbf{q}}\right)\\ &= \frac{\partial \mathbf{q}}{\partial \mathbf{z}}\left(\frac{\partial}{\partial \mathbf{q}}\left(\frac{\partial F}{\partial \mathbf{q}}\right)\right)\\ &= \frac{\partial \mathbf{q}}{\partial \mathbf{z}}\left(\frac{\partial^2 F}{\partial \mathbf{q}\partial \mathbf{q}}\right)\\ &= \mathbf{D}^{-1}\mathbf{L}^{-1}\left(\frac{\partial^2 F}{\partial \mathbf{q}\partial \mathbf{q}}\right)\end{aligned}$$

In the last line we have substituted the value of

$$\frac{\partial \mathbf{q}}{\partial \mathbf{z}} = \mathbf{L}\mathbf{D}^{-1} = \mathbf{D}^{-1}\mathbf{L}^{-1}$$

discussed in Section 4.5.3. This is just the Jacobian of the demands with respect to the prices.

We re-use the same matrices $\mathbf{A}_G$ and $\mathbf{A}_\Psi$ defined in the proof of Theorem 4.16 to decompose the Hessian of F into the parts arising from the two terms, analyzed in Lemmas 4.13 and 4.14, respectively.

$$\begin{aligned}\frac{\partial F'}{\partial \mathbf{z}} &= \mathbf{D}^{-1}\mathbf{L}^{-1}\left(\mathbf{L}\left(\mathbf{A}_\Psi - \lambda\mu\mathbf{A}_G\right)\mathbf{L}^\top\right)\\ &= \mathbf{D}^{-1}\left(\mathbf{A}_\Psi - \lambda\mu\mathbf{A}_G\right)\mathbf{L}^\top\\ &= \mathbf{D}^{-1}\mathbf{A}_\Psi\mathbf{L}^\top - \lambda\mu\mathbf{D}^{-1}\mathbf{A}_G\mathbf{L}^\top.\end{aligned}$$

The second term of the last line is simply $\mathbf{D}^{-1}\mathbf{A}_G\mathbf{L}^\top = \mathbf{D}^{-1}\tilde{\mathbf{G}}\mathbf{L}^\top$. Expanding the first term, we have that

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{A}_\Psi\mathbf{L}^\top &= \mathbf{D}^{-1}\left(\mathbf{D}\left(\mathbf{L}^{-1}\right)^\top + \mathbf{L}^{-1}\mathbf{D} - \mathbf{D}\,\mathrm{diag}\left(\mathbf{g}\right) - \mathbf{L}\mathbf{D}\mathbf{q}\cdot\mathcal{K}\right)\mathbf{L}^\top\\ &= \mathbf{I} + \left(\mathbf{D}^{-1}\mathbf{L}^{-1}\mathbf{D} - \mathrm{diag}\left(\mathbf{g}\right)\right)\mathbf{L}^\top - \mathbf{D}^{-1}\left(\mathbf{L}\mathbf{D}\mathbf{q}\cdot\mathcal{K}\right)\mathbf{L}^\top.\end{aligned}$$

But from the definition of $\mathbf{L}^{-1}$ in Section 4.5.1 and of $\mathbf{D}$ in Section 4.5.3,

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{L}^{-1}\mathbf{D} &= \mathbf{BY}\left(\operatorname{diag}(\mathbf{g}) + \mathbf{GY}\right)\mathbf{Y}^{-1}\mathbf{B}^{-1} \\ &= \mathbf{B}\left(\operatorname{diag}(\mathbf{g}) + \mathbf{YG}\right)\mathbf{B}^{-1} \\ &= \mathbf{B}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{B}^{-1}.\end{aligned}$$

Then, substituting back,

$$\mathbf{D}^{-1}\mathbf{A}_{\Psi}\mathbf{L}^{\top} = \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{BL}^{\top} - \operatorname{diag}(\mathbf{g})\,\mathbf{L}^{\top} - \mathbf{D}^{-1}\left(\mathbf{LDq}\cdot\mathcal{K}\right)\mathbf{L}^{\top}.$$

Applying Lemma 4.15 yields

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{A}_{\Psi}\mathbf{L}^{\top} &= \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{BL}^{\top} - \operatorname{diag}(\mathbf{g})\,\mathbf{L}^{\top} \\ &\qquad + \mathbf{D}^{-1}\left(\mathbf{G} + \mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}}\right)\mathbf{L}^{\top} \\ &= \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{BL}^{\top} - \left(\operatorname{diag}(\mathbf{g}) + \mathbf{BYG}\right)\mathbf{L}^{\top} \\ &\qquad + \mathbf{D}^{-1}\left(\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}}\right)\mathbf{L}^{\top} \\ &= \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{BL}^{\top} - \left(\tilde{\mathbf{L}}^{-1}\right)^{\top}\mathbf{L}^{\top} + \mathbf{D}^{-1}\left(\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - \tilde{\mathbf{G}}\right)\mathbf{L}^{\top}.\end{aligned}$$

Adding back the first term from the beginning of the proof, the Jacobian matrix is

$$\begin{aligned}\frac{\partial F'}{\partial \mathbf{z}} = \mathbf{I} + \mathbf{B}^{-1}\left(\mathbf{L}^{-1}\right)^{\top}\mathbf{BL}^{\top} - \left(\tilde{\mathbf{L}}^{-1}\right)^{\top}\mathbf{L}^{\top} + \\ \mathbf{D}^{-1}\left(\mathbf{S}(\mathbf{B}^{-1} - \mathbf{I})\mathbf{e}\cdot\mathcal{K} - (1 + \lambda\mu)\tilde{\mathbf{G}}\right)\mathbf{L}^{\top}.\end{aligned}$$

This completes the proof. □

D.2 FOCs of the Original Profit Function

Proposition 5.2. *Under a GEV demand model, setting* $\lambda = \Pi(\mathbf{z}) = \mathbf{p}^\top \mathbf{z}$*, the FOCs of the parametric reformulation in* (5.4) *are equivalent to the FOCs*

$$\frac{\partial \Pi}{\partial \mathbf{z}} = -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \mathbf{w}\right) = \mathbf{0}$$

of the unconstrained pricing problem (5.1).

Proof. Applying the chain rule to $\Pi(\mathbf{z}) = \mathbf{p}^\top \mathbf{z}$, and then using the expression for $\mathbf{J}_\mathbf{z}^{-1}$ derived from Proposition 4.17, we have

$$\begin{aligned}
\frac{\partial \Pi}{\partial \mathbf{z}} &= \mathbf{J}_\mathbf{z}^{-1}\mathbf{z} + \mathbf{p} \\
&= \mathbf{D}^{-1}\mathbf{J}^{-1}\mathbf{z} + \mathbf{p} \\
&= -\mathbf{BYJ}^{-1}\mathbf{z} + \mathbf{p} \\
&= -\frac{1}{\mu G(\mathbf{y})}\mathbf{BYL}^{-1}\mathbf{z} + \mathbf{Bpp}^\top \mathbf{z} + \mathbf{p}.
\end{aligned}$$

Recall that $\mathbf{J}^{-1}$ is simply $\bar{\mathbf{J}}^{-1}$ without the last row and column. Then, from the definition of $\mathbf{S}^{-1}$ in (5.3), and from the definition of the demand $p_i = \frac{y_i G_i(\mathbf{u})}{\mu G(\mathbf{y}}$,

$$\begin{aligned}
\frac{\partial \Pi}{\partial \mathbf{z}} &= -\frac{1}{\mu G(\mathbf{y})}\mathbf{BY}\,\mathrm{diag}\left(\mathbf{g}\right)\mathbf{S}^{-1}\mathbf{z} + \mathbf{Bpp}^\top \mathbf{z} + \mathbf{p} \\
&= -\mathbf{BPS}^{-1}\mathbf{z} + \mathbf{BPp}^\top \mathbf{z} + \mathbf{p} \\
&= -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \mathbf{ep}^\top \mathbf{z} - \mathbf{B}^{-1}\mathbf{e}\right) \\
&= -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \lambda\mathbf{e} - \mathbf{B}^{-1}\mathbf{e}\right) \\
&= -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \left(\lambda\mathbf{I} + \mathbf{B}^{-1}\right)\mathbf{e}\right) \\
&= -\mathbf{BP}\left(\mathbf{S}^{-1}\mathbf{z} - \mathbf{w}\right)
\end{aligned}$$

Because the matrix **BP** has a strictly positive diagonal, this system of equations is exactly equivalent to (5.4) in Lemma 5.1, as claimed. □

D.3 Derivations for the CNL and MMNL Model

Lemma 5.9. *For the cross-nested logit (CNL) model,*

$$\bar{\mathbf{S}}^{-1} = \operatorname{diag}(\mathbf{NWe}) - \mathbf{NWP}_{|}^{\top} + \mu \mathbf{NP}_{|}^{\top},$$

where $\mathbf{W} = \operatorname{diag}(\mu_1, \ldots, \mu_M)$ *is a diagonal matrix,* $\mathbf{P}_{|} \in \mathbb{R}^{n+1 \times m}$ *is the matrix of conditional choice probabilities with* $[\mathbf{P}_{|}]_{im} = p_{i|m}$ *in the* (i, m) *component, and* $\mathbf{N} = \bar{\mathbf{P}}^{-1}\mathbf{P}_{|}\mathbf{Q}$. [1]

Moreover, the Jacobian matrix of the choice probabilities with respect to prices is

$$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = -\bar{\mathbf{B}}\left(\mathbf{PS}^{-1} - \mathbf{pp}^{\top}\right) = -\bar{\mathbf{B}}\left(\operatorname{diag}(\mathbf{Ae}) - \mathbf{A}\right) \tag{D.1}$$

where $\mathbf{A} \in \mathbb{R}^{n+1,n+1}$ *is the symmetric matrix*

$$\mathbf{A} = \begin{bmatrix} \mathbf{P}_{|} & \bar{\mathbf{p}} \end{bmatrix} \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{|}^{\top} \\ \bar{\mathbf{p}}^{\top} \end{bmatrix}.$$

Proof. The second derivatives of the generating function are, for $i \neq j$,

$$\begin{aligned} G_{ij}(\mathbf{y}) = \frac{\partial^2 G(y)}{\partial y_i \partial y_j} &= \mu \sum_{m=1}^{M} (\mu - \mu_m)(\alpha_{im}\alpha_{jm})^{\frac{\mu_m}{\mu}} (y_i y_j)^{\mu_m - 1} \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 2}, \\ G_{ii}(\mathbf{y}) = \frac{\partial^2 G(y)}{\partial y_i^2} &= \mu \sum_{m=1}^{M} (\mu - \mu_m) \left(\alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 1} \right)^2 \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 2} + \\ &\quad \mu \sum_{m=1}^{M} (\mu_m - 1) \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m - 2} \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1} \end{aligned}$$

We first find an expression for the matrix $\frac{1}{G(\mathbf{y})}\mathbf{YGY}$. Multiplying the preceding equations appropriately, we have

[1] For the NL models with no cross-nesting, $\mathbf{N}$ was the incidence matrix of products to nests. It is no longer constant for CNL models, although the equation given here obviously holds in the special case of the NL model.

$$
\begin{aligned}
\frac{y_i y_j G_{ij}(\mathbf{y})}{G(\mathbf{y})} &= \mu \frac{1}{G(\mathbf{y})} \sum_{m=1}^{M} (\mu - \mu_m)(\alpha_{im}\alpha_{jm})^{\frac{\mu_m}{\mu}} (y_i y_j)^{\mu_m} \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m}-2} \\
&= -\mu \sum_{m=1}^{M} (\mu_m - \mu) p_{i|m} p_{j|m} q_m, \\
\frac{y_i^2 G_{ii}(\mathbf{y})}{G(\mathbf{y})} &= \mu \frac{1}{G(\mathbf{y})} \sum_{m=1}^{M} (\mu - \mu_m) \left(\alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m} \right)^2 \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m}-2} + \\
&\qquad \mu \frac{1}{G(\mathbf{y})} \sum_{m=1}^{M} (\mu_m - 1) \alpha_{im}^{\frac{\mu_m}{\mu}} y_i^{\mu_m} \left(\sum_{k=1}^{n} \alpha_{km}^{\frac{\mu_m}{\mu}} y_k^{\mu_m} \right)^{\frac{\mu}{\mu_m}-1} \\
&= -\mu \sum_{m=1}^{M} (\mu_m - \mu) p_{i|m}^2 q_m + \mu \sum_{m=1}^{M} (\mu_m - 1) p_{i|m} q_m
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{1}{\mu G(\mathbf{y})} \mathbf{YGY} &= \sum_{m=1}^{M} (\mu_m - 1) q_m \operatorname{diag}\left(\mathbf{p}_{|\mathbf{m}}\right) - \sum_{m=1}^{M} (\mu_m - \mu) q_m \mathbf{p}_{|m} \mathbf{p}_{|m}^{\top} \\
&= \sum_{m=1}^{M} \mu_m q_m \operatorname{diag}\left(\mathbf{p}_{|\mathbf{m}}\right) - \mathbf{P} - \sum_{m=1}^{M} (\mu_m - \mu) q_m \mathbf{p}_{|m} \mathbf{p}_{|m}^{\top} \\
&= \operatorname{diag}\left(\mathbf{P}_{|}(\mathbf{WQ})\mathbf{e}\right) - \mathbf{P}_{|}(\mathbf{W} - \mu \mathbf{I})\mathbf{Q})\mathbf{P}_{|}{}^{\top} - \mathbf{P}.
\end{aligned}
$$

Now ,

$$
\begin{aligned}
\mathbf{S}^{-1} &= \operatorname{diag}(\mathbf{g})^{-1} \mathbf{L}^{-1} \\
&= \frac{1}{\mu G} \mathbf{P}^{-1} \mathbf{Y} \mathbf{L}^{-1} \\
&= \mathbf{P}^{-1} \frac{1}{\mu G} (\operatorname{diag}(\mathbf{Yg}) + \mathbf{YGY}) \\
&= \mathbf{P}^{-1} (\mathbf{P} + \frac{1}{\mu G(\mathbf{y})} \mathbf{YGY}) \\
&= \mathbf{P}^{-1} (\operatorname{diag}\left(\mathbf{P}_{|}(\mathbf{WQ})\mathbf{e}\right) - \mathbf{P}_{|}(\mathbf{W} - \mu \mathbf{I})\mathbf{Q})\mathbf{P}_{|}{}^{\top}) \\
&= \operatorname{diag}\left(\mathbf{P}^{-1}\mathbf{P}_{|}(\mathbf{WQ})\mathbf{e}\right) - \mathbf{P}^{-1}\mathbf{P}_{|}\mathbf{WQP}_{|}{}^{\top} + \mu \mathbf{P}^{-1}\mathbf{P}_{|}\mathbf{QP}_{|}{}^{\top}
\end{aligned}
$$

Substituting $\mathbf{P}^{-1}\mathbf{P}_{|}\mathbf{Q} = \mathbf{N}$, we we have

$$\mathbf{S}^{-1} = \operatorname{diag}(\mathbf{NWe}) - \mathbf{NWP}_{|}^{\top} + \mu \mathbf{NP}_{|}^{\top}$$

The expression for the Jacobian can be obtained straighforwardly from Proposition 4.17 and the definitions of $\mathbf{S}^{-1}$ and $\mathbf{D}$. □

Lemma 5.10. *Under mixed logit (MMNL) models, the Jacobian of the choice probabilities with respect to prices is given by*

$$\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1} = -\left(\operatorname{diag}\left(\mathbf{A}^{MIX}\mathbf{e}\right) - \mathbf{A}^{MIX}\right)$$

$$\mathbf{A}^{MIX} = \mathbf{P}_{|}\mathbf{W}^{MIX}\mathbf{P}_{|}^{\top}$$

where $\mathbf{W}^{MIX} = \operatorname{diag}\left(b_{\cdot|1}\gamma_1, \ldots, b_{\cdot|M}\gamma_M\right)$ *is a diagonal matrix with the price sensitivity parameters for each nest scaled by the nest sizes, and* $\mathbf{P}_{|} \in \mathbb{R}^{n+1\times m}$ *is the matrix of choice probabilities* $p_{i|m}$.

Proof. The MNL model is a GEV model with $\mu = 1$, $G(\mathbf{y}) = sum_i y_i$. Then $\mathbf{g} = \mathbf{e}$ and $\mathbf{G} = \mathbf{0}$. It is easily verified that

$$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = \bar{\mathbf{D}}^{-1}\bar{\mathbf{J}}^{-1} = -\bar{\mathbf{B}}(\bar{\mathbf{P}} - \bar{\mathbf{p}}\bar{\mathbf{p}}^{\top}).$$

In fact, this is a special case of Lemma 5.9. For the MNL, using (5.3), $\mathbf{S}^{-1} = \left(\mathbf{I} + \operatorname{diag}(\mathbf{g})^{-1}\mathbf{GY}\right) = (\mathbf{I} + \mathbf{I0Y}) = \mathbf{I}$. Substituting into the expression for $\bar{\mathbf{J}}_{\mathbf{z}}^{-1}$ of Lemma 5.9, we obtained the above expression. But since the price sensitivities are all equal within a nest, we have

$$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = -b_{\cdot|m}\left(\bar{\mathbf{P}} - \bar{\mathbf{p}}\bar{\mathbf{p}}^{\top}\right)$$

Then $\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1}$ is simply the sum of M such Jacobians, weighted by the segment sizes γ_m as well as by the price sensitivities $b_{\cdot|m}$. □

Theorem 5.11. *A MMNL model with a constant price sensitivity parameter for each nest can be approximated locally, at a give price vector* $\mathbf{z}$*, by a CNL model the same choice probabilities* $\bar{\mathbf{p}}$*, the same conditional choice probabilities* $\mathbf{P}_|$ *and an arbitrarily close Jacobian matrix* $\mathbf{J}_{\mathbf{z}}^{-1}$ *of the choice probabilities with respect to the prices.*

Proof. We relax the assumption that y_{n+1} is fixed without loss of generality, as discussed earlier, because fixing it simply removes the last row from the Jacobian matrix of the demand model. Indeed, given any GEV model over $n+1$ products where x_{n+1} is fixed (i.e. there is no price to change for the outside alternative), the parameters $d_1, \ldots, d_{n+1}$ can be shifted by a constant such that $y_{n+1} = 1$. This does not affect the demand model, since it corresponds to a scaling of the vector $\bar{\mathbf{y}}$.

Suppose we are given the matrix of conditional choice probabilities $\mathbf{P}_|$ and the nest probabilities $\mathbf{Q} = \operatorname{diag}([\gamma_1, \ldots, \gamma_M])$ for a MMNL model with at fixed price vector $\mathbf{z}$. We construct a CNL approximation. First, choose the parameter $0 < \gamma_{M+1} < 1$. We will show that as γ_{M+1} approaches zero, the approximation of the Jacobian becomes exact, but the nest scale parameters $\mu_1, \ldots \mu_M$ become large. First, fix $\mu = 1$ and select price sensitivity parameters $b_1 = b_2 = \ldots = b_n = b_{n+1} = \frac{\gamma_{M+1}}{1+\gamma_{M+1}}$. Then let, for each $m = 1, \ldots, M$,

$$\mu_m = \mu + \frac{b_{\cdot|m}\gamma_m}{\gamma_{M+1}} = \mu + \frac{b_{\cdot|m}Q_m}{\gamma_{M+1}}.$$

Clearly, $\lim_{\gamma_{M+1}\to 0}(\gamma_{M+1}\mu_m) = b_{\cdot|m}\gamma_m$, and therefore $\lim_{\gamma_{M+1}\to 0}\gamma_{M+1}\mathbf{W} = \mathbf{W}_{MIX}$, where we refer to the quantities in the expression of the Jacobians in Lemma 5.9 and Lemma 5.10. We can write,

$$\mathbf{A} = \mathbf{P}_|\mathbf{W}\mathbf{P}_|^{\top} + \bar{\mathbf{p}}\bar{\mathbf{p}}^{\top},$$

and therefore,

$$\begin{aligned}\lim_{\gamma_{M+1}\to 0}\gamma_{M+1}\mathbf{A} &= \lim_{\gamma_{M+1}\to 0}\gamma_{M+1}\mathbf{P}_|\mathbf{W}\mathbf{P}_|^{\top} + \lim_{\gamma_{M+1}\to 0}\gamma_{M+1}\bar{\mathbf{p}}\bar{\mathbf{p}}^{\top} \\ &= \mathbf{P}_|\mathbf{W}_{MIX}\mathbf{P}_|^{\top}.\end{aligned}$$

We also have, by the choice of the b_i, that

$$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = -\frac{\gamma_{M+1}}{1+\gamma_{M+1}}\mathbf{I}(\operatorname{diag}(\mathbf{A}\mathbf{e}) - \mathbf{A}).$$

Then, in the limit,

$$\begin{aligned}
\lim_{\gamma_{M+1}\to 0} \bar{\mathbf{J}}_{\mathbf{z}}^{-1} &= \lim_{\gamma_{M+1}\to 0} -\frac{1}{1+\gamma_{M+1}}(\operatorname{diag}(\mathbf{A}_{MIX}\mathbf{e}) - \mathbf{A}_{MIX}) \\
&= -\left(\lim_{\gamma_{M+1}\to 0} \frac{1}{1+\gamma_{M+1}}\right)(\operatorname{diag}(\mathbf{A}_{MIX}\mathbf{e}) - \mathbf{A}_{MIX}) \\
&= \bar{\mathbf{J}}_{\mathbf{MIX}}^{-1}.
\end{aligned}$$

This shows that the Jacobian of the CNL model can be made arbitrarily close to that of the MMNL model, by choosing a sufficiently small value of γ_{n+1}.

We now choose the remaining model parameters that yield the correct choice probabilities $\mathbf{P}_|$, $\mathbf{Q}$, and $\mathbf{p} = \bar{\mathbf{P}}\mathbf{e} = \mathbf{P}_|\mathbf{Q}\mathbf{e}$. Notice that the values of $\mathbf{P}_|$ and $\mathbf{Q}$ determines the value of $\mathbf{p}$, so we need only show that the former two quantities are accurate. Let $\mathbf{N} = \mathbf{Y}^{-1}\mathbf{P}_| \operatorname{diag}(\gamma)$. We define the nesting parameters as

$$\alpha_{im} = p_{i|m}^{\frac{\mu}{\mu_m}} \gamma_m y_i^{-1}, \qquad \forall i, \forall m.$$

Solving for the conditional probabilities, and substituting $Q_m = \gamma_m$ yields.

$$(\alpha_{im} y_i)^{\frac{\mu_m}{\mu}} = p_{i|m} Q_m^{\frac{\mu_m}{\mu}}, \qquad \forall i, \forall m. \tag{D.2}$$

Summing over i, we have

$$\sum_i (\alpha_{im} y_i)^{\frac{\mu_m}{\mu}}) = Q_m^{\frac{\mu_m}{\mu}}, \qquad \forall m. \tag{D.3}$$

But because both probabilities in (5.11) are normalized, (D.2) and (D.3) are sufficient to ensure that the equalities hold. (Recall that we set $\mu = 1$. For (D.2), the factor $Q_m^{\frac{\mu_m}{\mu}}$ scales all the terms for the conditional probabilities in nest m equally.)

We must still scale the α_{im} so that the condition $\sum_m \alpha_{im} = 1, \forall i$ is satisfied.

Because α_{im} always occurs in the products of the form $(\alpha_{im} y_i)$ above, scaling the terms of this sum for product i amounts to scaling y_i. This is the same as shifting the parameter d_i, which is still free. This completes the proof.

We remark that different authors have employed different but equivalent normalizations of the CNL parameters [11]. We also point out that, although we chose all the price sensitivity parameters to be equal, their relative values are arbitrary. If we allowed the MMNL model to have different price-sensitibity parameters for the products in each nest, choosing different CNL parameters $b_1, \ldots, b_{n+1}$ could still allow a good approximation of the Jacobian matrix even though it would no longer be exact in the limit. $\square$

Appendix E

Notation Summary

The following pages provide a summary of the notation used in Chapters 3, 4 and 5. Notation specific to one proof or one section is excluded. Notation for GA models is defined in the body of Chapter 2. We begin with some basic definitions.

n	Number of products.
$\mathbf{I}$	Identity matrix.
$\mathbf{e}$	Vector of all ones.
$\mathbf{0}$	Vector of all zeros.
$\mathbf{e}_i$	Vector of all zeros except for a one in position i.
$\mathrm{diag}\,(\mathbf{x})$	Diagonal matrix with $\mathbf{x}$ on the diagonal.
$\Delta_{n+1} \subset \mathbb{R}_+^{n+1}$	Simplex of probability distributions.
$[x_{ij}]_{ij}$	The matrix with entries x_{ij}.
$[x_{ijk}]_{ijk}$	The tensor with entries x_{ij}.
$[\mathbf{X}]_{ij}$	The (i, j) component of the matrix $\mathbf{X}$.

E.1 Notation Common to all GEV Models

The following quantities are defined for all GEV models.

$\mathbf{p} = [p_1, \ldots, p_n]^\top$	Demands for the n products (choice probabilities).
$\bar{\mathbf{p}} = [p_1, \ldots, p_{n+1}]^\top$	Demands, including no-purchase probability.
$\mathbf{q} = [q_1, \ldots, q_n]^\top$	Unnormalized demands.
$\bar{\mathbf{q}} = [q_1, \ldots, q_{n+1}]^\top$	Unnormalized demands, including lost demand.
$\mathbf{x} = [x_1, \ldots, x_n]^\top$	Prices of the n products.
$z_1, \ldots, z_n$	Adjusted prices accountomg for the a_i and c_i.
$a_1, \ldots, a_n$	Profit margins (problem parameters).
$c_1, \ldots, c_n$	Production costs (problem parameters).
$b_1, \ldots, b_n$	Price sensitivities (demand model parameters).
$\mathbf{B} = \operatorname{diag}([b'_1, \ldots, b'_n])$	Adjusted price sensitivities (when using $\mathbf{z}$).
$d_1, \ldots, d_n$	Quality parameters (demand model parameters).
$d'_1, \ldots, d'_n$	Adjusted quality parameters (when using $\mathbf{z}$).
$\Pi(\mathbf{x}) = \Pi(\mathbf{z})$	Profit, in terms of the (adjusted) prices.
$\Pi(\mathbf{p}) = \Pi(\mathbf{q})$	Profit, in terms of the (unnormalized) demands.
$\mathbf{y} = [y_1, \ldots, y_n]^\top$	Product *attractions* under all GEV models.
$\bar{\mathbf{y}} = [y_1, \ldots, y_{n+1} = 1]^\top$	Product *attractions*, with no-purchase attraction.
$\mathbf{Y}, \bar{\mathbf{Y}}$	Matrices with attractions $\mathbf{y}$, $\bar{\mathbf{y}}$ on the diagonal.
μ	Scale parameter under all GEV models. Usually $\mu = 1$.
$G(\mathbf{y}) = G(\bar{\mathbf{y}})$	GEV generating function under all GEV models.
$\mathbf{g} = [G_i(\mathbf{y}), \ldots, G_n(\mathbf{y})]^\top$	Gradient of $G(\mathbf{y})$.
$\mathbf{G} = [G_{ij}(\mathbf{y})]_{ij}$	Hessian of $G(\mathbf{y})$.
$\mathcal{G} = [G_{ijk}(\mathbf{y})]_{ijk}$	Tensor of third partial derivative of $G(\mathbf{y})$.

E.2 Notation for NL, CNL and MMNL models.

The following quantities are defined for NL models (Chapter 3). Some of them carry over the to the more general CNL model, or even to the MMNL model (Chapter 4 and Section 5.7).

M	Number of nests under NL, CNL and MMNL models.
$\mu_1, \ldots \mu_M$	Nest scale parameters under NL and CNL models.
$\alpha_{im} \in [0, 1]$	Nesting structure parameters under NL and CNL models. For NL models, $\alpha_{im} \in \{0, 1\}$.
$\mathbf{M} = [\mathbf{I} \quad \mathbf{e}^\top] \in \mathbb{R}^{n \times (n+1)}$	Jacobian of $\bar{\mathbf{p}}$ w.r.t. $\mathbf{p}$ under NL models.
$\mathrm{N} = [\alpha_{im}] \in \{0, 1\}^{n+1, M}$	Incidence matrix of products to nests (NL only).
m_i	Nest corresponding to product i in NL models.
$\mathbf{Q_m} = [Q_1, \ldots, Q_M]^\top$	Nest probabilities under NL, CNL and MMNL models.
$\mathbf{p}_{\mathbf{\vert m}} = [p_{1\vert 1}, \ldots, p_{n\vert M}]^\top$	Vector containing the $(n+1)M$ conditional choice probabilities within each nest under NL, CNL and MMNL models.
$\Pi(\mathbf{p}_{\mathbf{\vert m}}, \mathbf{Q_m}) = \Pi(\mathbf{p})$	Profit, in terms of the conditional choice probabilities.
$\mathbf{J_z} = \left[\frac{\partial z_j}{\partial p_i}\right]_{ij}$	Jacobian matrix of $\mathbf{z}$ w.r.t. $\mathbf{p}$.
$\mathbf{J_z}^{-1} = \left[\frac{\partial p_j}{\partial z_i}\right]_{ij}$	Jacobian matrix of $\mathbf{p}$ w.r.t. $\mathbf{x}$.
$\mathbf{H} = \left[\frac{\partial^2 \Pi}{\partial p_i \partial p_j}\right]_{ij}$	Hessian matrix of $\Pi(\mathbf{p})$ under NL demand models.

The following matrices are mainly used in the analysis of NL and CNL models in Chapter 3 and Section 5.7.

$\mathbf{P} = \operatorname{diag}(\mathbf{p})$	Matrix with choice probabilities on the diagonal.
$\bar{\mathbf{P}} = \operatorname{diag}(\bar{\mathbf{p}})$	Matrix with choice probabilities on the diagonal.
$\mathbf{Q} = \operatorname{diag}(\mathbf{Q_m})$	Matrix with nest probabilities on the diagonal.
$\bar{\mathbf{U}} \in \mathbb{R}^{(n+1)\times(n+1)}$	Matrix with parameters $\mu_{m_i}^{-1}$ on the diagonal.
$\mathbf{V} \in \mathbb{R}^{M\times M}$	Matrix with parameters $1 - \mu_m^{-1}$ on the diagonal.
$\bar{\mathbf{B}} \in \mathbb{R}^{(n+1)\times(n+1)}$	Matrix $\mathbf{B}$ augmented with an additional diagonal component b_{n+1} for notational convenience.[1]

The following matrices are used in Appendix B, Chapter 4 and Chapter 5.

$\mathbf{D} = -\mathbf{B}^{-1}\mathbf{Y}^{-1} = \left[\frac{\partial z_j}{\partial y_i}\right]_{ij}$	Jacobian matrix of $\mathbf{z}$ w.r.t. $\mathbf{y}$.
$\mathbf{D}^{-1} = -\mathbf{B}\mathbf{Y} = \left[\frac{\partial y_j}{\partial z_i}\right]_{ij}$	Jacobian matrix of $\mathbf{y}$ w.r.t. $\mathbf{z}$.
$\mathbf{J} = \left[\frac{\partial y_j}{\partial p_i}\right]_{ij}$	Jacobian matrix of $\mathbf{y}$ w.r.t. $\mathbf{p}$.
$\mathbf{J}^{-1} = \left[\frac{\partial p_j}{\partial y_i}\right]_{ij}$	Jacobian matrix of $\mathbf{p}$ w.r.t. $\mathbf{y}$.
$\mathbf{J}_{\bar{\mathbf{p}}\mathbf{q}} \in \mathbb{R}^{(n+1+M)\times(n+1)}$	Jacobian matrix of $\mathbf{y}$ w.r.t. $(\mathbf{p}, \mathbf{Q_m})$.

[1] This avoids cumbersome definitions while making clear the relationship with $\mathbf{B}$ in a consistent manner. In Chapter 3, we set the last component to $b_{n+1} = -p_{n+1}/(\mathbf{e}^\top\mathbf{B}^{-1}\mathbf{p})$. Within the proof of Lemma 4.15, we let $b_{n+1} = 1$. In Section 5.7, we let $b_{n+1} = 0$.

E.3 Notation: GEV Model Reformulation

The following quantities define the reformulation in Chapters 4 and 5.

λ	Parameter in the reformulation. Dual variable for the simplex constraint when maximizing $\Pi(\mathbf{p})$ or $\Pi(\mathbf{p}_{\|\mathbf{m}}, \mathbf{Q}_{\mathbf{m}})$. Equal to Π at optimality.
$\Psi(\mathbf{q})$	Numerator of the profit in terms of $\mathbf{q}$.
$\Gamma(\mathbf{q}) = \Psi(\mathbf{q}) - \lambda\mu G(\mathbf{y})$	Objective of the parametric reformulation.
$F(\mathbf{z}) = \Psi(\mathbf{z}) - \lambda G(\mathbf{y})$	Objective of the parametric reformulation, in terms of $\mathbf{z}$, absorbing the constant μ into the parameter λ for convenience (Chapter 5 only).

The following matrices are used in the analysis of Chapters 4 and 5.

$\mathbf{L} = \left[\frac{\partial y_j}{\partial q_i}\right]_{ij}$	Jacobian matrix of $\mathbf{y}$ w.r.t. $\mathbf{q}$.
$\mathbf{L}^{-1} = \left[\frac{\partial q_j}{\partial y_i}\right]_{ij}$	Jacobian matrix of $\mathbf{q}$ w.r.t. $\mathbf{y}$.
$\mathcal{K}^k = \left[\frac{\partial^2 q_k}{\partial y_i \partial y_j}\right]_{ij}$	Hessian matrix of q_k w.r.t. $\mathbf{y}, k = 1, \ldots, n+1$.
$\mathcal{K} = \left[\frac{\partial^2 q_k}{\partial y_i \partial y_j}\right]_{ijk}$	Tensor of second partial derivatives of $\bar{\mathbf{q}}$ w.r.t. $\mathbf{y}$.
$\bar{\mathbf{L}}^{-1} = \left[\frac{\partial q_j}{\partial y_i}\right]_{ij}$	Jacobian matrix of $\bar{\mathbf{q}}$ w.r.t. $\bar{\mathbf{y}}$. Singular matrix.
$\bar{\mathcal{K}}^k = \left[\frac{\partial^2 q_k}{\partial y_i \partial y_j}\right]_{ij}$	Hessian matrix of q_k w.r.t. $\bar{\mathbf{y}}, k = 1, \ldots, n+1$.
$\bar{\mathcal{K}} = \left[\frac{\partial^2 q_k}{\partial y_i \partial y_j}\right]_{ijk}$	Tensor of second partial derivatives of $\bar{\mathbf{q}}$ w.r.t. $\bar{\mathbf{y}}$.
$\mu\mathbf{S} = \mu\mathbf{L}\,\mathrm{diag}\,(\mathbf{g})$	A row sub-stochastic matrix of interest.
$\mathbf{S}^{-1} = \mu\,\mathrm{diag}\,(\mathbf{g})^{-1}\,\mathbf{L}^{-1}$	An M-matrix of interest.
$\mu\bar{\mathbf{S}} = \mu\bar{\mathbf{L}}\,\mathrm{diag}\,(\bar{\mathbf{g}})$	A row stochastic matrix of interest.
$\bar{\mathbf{S}}^{-1} = \mathrm{diag}\,(\bar{\mathbf{g}})^{-1}\,\bar{\mathbf{L}}^{-1}$	An M-matrix of interest.

E.4 Notation: Fixed-point Iterations

The following quantities are used to define the steps of the algorithm in Chapter 5.

$0 < \alpha \leq 1$	Step size parameter.
L	Number of steps between iteration matrix updates.
t	Step number.
k	Number of the most recent step in which the iteration matrix was updated.
$\mathbf{w} = (\mathbf{B}^{-1} + \lambda\mathbf{I})\mathbf{e}$	A vector that varies only in λ.
$\mathbf{S}^{-1} = \mathbf{E} + \mathbf{F}$	Splitting of $\mathbf{S}^{-1}$ into its diagonal entries $\mathbf{E}$ and its off-diagonal entries $\mathbf{F}$, for the linear Jacobi method.
$\mathbf{S_k}^{-1} = \mathbf{E}_k + \mathbf{F}_k$	Splitting of $\mathbf{S}_k^{-1}$, for the linear Jacobi method.
$\mathbf{M}_k$	Jacobi-type iteration matrix. Depends nonlinearly on z^k. Substochastic.
$T(\mathbf{z}_t) = \mathbf{M}_k\mathbf{z}_t + \alpha\mathbf{E}_k^{-1}\mathbf{w}$	Nonlinear operator for the Jacobi-type iteration.
$\mathbf{R}_k$	Richardson-type iteration matrix. Depends nonlinearly on z^k. Substochastic.
$T(\mathbf{z}_t) = \mathbf{R}_k\mathbf{z}_t + \alpha\mathbf{w}$	Nonlinear operator for the Richardson-type iteration.

E.5 Additional Notation for MMNL Models

The following quantities are used to define MMNL models and explore their relationship with CNL models in Section 5.7.

$\gamma_1, \ldots, \gamma_M$	Constant nest probability paramaters (weights).
$b_{\cdot\|1}, \ldots, b_{\cdot\|M}$	Price sensitivity parameters for each nest.
$\mathbf{P}_{\|} \in \mathbb{R}^{(n+1)\times M}$	Matrix of conditional choice probabilities $p_{i\|m}$.
$\bar{\mathbf{J}}_{\mathbf{z}}^{-1} = \left[\frac{\partial p_j}{\partial z_i}\right]_{ij}$	Jacobian matrix of $\bar{\mathbf{p}}$ with respect to $\mathbf{z}$ for CNL models. An $(n+1) \times (n+1)$ matrix for convenience. The last row is assumed to contain all zeros.
$\bar{\mathbf{J}}_{\mathbf{MIX}}^{-1} = \left[\frac{\partial p_j}{\partial z_i}\right]_{ij}$	Jacobian matrix of $\bar{\mathbf{p}}$ with respect to $\mathbf{z}$ for MMNL models. An $(n+1) \times (n+1)$ matrix for convenience. The last row is assumed to contain all zeros.

Bibliography

[1] D. Adelman. Dynamic bid prices in revenue management. *Operations Research*, 55(4):647–661, 2007.

[2] S. Damla Ahipasaoglu, Peng Sun, and Michael J. Todd. Linear convergence of a modified Frank-Wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimization Methods Software*, 23(1):5–19, February 2008.

[3] Yaln Akay, Harihara Prasad Natarajan, and Susan H. Xu. Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science*, 56 (8):1345–1361, 2010.

[4] G. Allon, A. Federgruen, and M. Pierson. Price competition under multinomial logit demand function with random coefficients, 2010.

[5] G. Aydin and E. L. Porteus. Joint inventory and pricing decisions for an assortment. *Operations Research*, 56(5):1247–1255, 2008.

[6] Roberto Bagnara. A unified proof for the convergence of Jacobi and Gauss-Seidel methods. *SIAM Review*, 37, 1995.

[7] Moshe E. Ben-Akiva and Steven R. Lerman. *Discrete choice analysis : theory and application to travel demand.* MIT Press, Cambridge, Mass. :, 1985. ISBN 0262022176.

[8] Steven T. Berry, Amit Gandhi, and Philip Haile. Connected substitutes and invertibility of demand. Working Paper 17193, National Bureau of Economic Research, July 2011.

[9] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 2nd edition, September 1999. ISBN 1886529000.

[10] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, 2007.

[11] Michel Bierlaire. A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144:287–300, 2006.

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

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

[14] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, March 2004.

[15] R.P. Brent. *Algorithms for Minimization without Derivatives.* Prentice-Hall, Englewood Cliffs, NJ, 1973.

[16] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM J. on Optimization*, 18(3):853–877, 2007.

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

[18] Lj. B. Ćirić. A generalization of Banach's contraction principle. *Proceedings of the American Mathematical Society*, 45(2):pp. 267–273, 1974.

[19] Andrew Daly and Michel Bierlaire. A General and Operational Representation of Generalised Extreme Value Models. *Transportation Research Part B: Methodological*, 40(4):285–305, 2006.

[20] James M. Davis, Guillermo Gallego, and Huseyin Topaloglu. Assortment optimization under variants of the nested logit model. 2012. Working Paper.

[21] D. den Hertog, F. Jarre, C. Roos, and T. Terlaky. A sufficient condition for self-concordance, with application to some classes of structured convex programming problems. *Mathematical Programming*, 69(1):75–88, 1995.

[22] Werner Dinkelbach. On nonlinear fractional programming. *Management Science*, 13(7):pp. 492–498, 1967.

[23] L. Dong, P. Kouvelis, and Z. Tian. Dynamic pricing and inventory control of substitute products. *Manufacturing and Service Operations Management*, 2008.

[24] Didier El Baz. M-functions and parallel asynchronous algorithms. *SIAM J. Numer. Anal.*, 27(1):136–140, February 1990.

[25] Abed Elhashash and Daniel B. Szyld. Generalizations of m-matrices which may not have a nonnegative inverse. *Linear Algebra and its Applications*, 429(10): 2435 – 2450, 2008. ¡ce:title¿Special Issue in honor of Richard S. Varga¡/ce:title¿.

[26] W. Elmaghraby and P. Keskinocak. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, 49(10):1287–1309, 2003.

[27] R. Fourer, D. M. Gay, and B. W. Kernighan. A modeling language for mathematical programming. *Management Science*, 36(5):519–554, 1990.

[28] Andreas Frommer. Generalized nonlinear diagonal dominance and applications to asynchronous iterative methods. *J. Comput. Appl. Math.*, 38(1-3):105–124, December 1991.

[29] T. Fujisawa and E. Kuh. Some results on existence and uniqueness of solutions of nonlinear networks. *Circuit Theory, IEEE Transactions on*, 18(5):501 – 506, sep 1971. ISSN 0018-9324. doi: 10.1109/TCT.1971.1083336.

[30] David Gale and Hukukane Nikaido. The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159:81–93, 1965. 10.1007/BF01360282.

[31] G. Gallego and C. Stefanescu. Upgrades, upsells and pricing in revenue management. 2009. Under review.

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

[33] G. Gallego, G. Iyengar, R. Phillips, and A. Dubey. Managing flexible products on a network. Technical report, CORS Technical Report TR-2004-01, 2004.

[34] G. Gallego, W. T. Huh, W. Kang, and R. Phillips. Price competition with the attraction demand model: Existence of unique equilibrium and its stability. *Manufacturing & Service Operations Management*, 8(4):359–375, 2006.

[35] Guillermo Gallego and Huseyin Topaloglu. Assortment optimization under variants of the nested logit model. 2012. Working Paper.

[36] Guillermo Gallego and Ruxian Wang. Multi-product price optimization and competition under the nested logit model with product-differentiated price sensitivities. 2012. Working Paper.

[37] T.-B. Gan, T.-Z. Huang, and J. Gao. A note on generalized nonlinear diagonal dominance. *Journal of Mathematical Analysis and Applications*, 313:581–586, January 2006.

[38] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

[39] Jochen Gorski, Frank Pfeuffer, and Kathrin Klamroth. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research*, 66:373–407, 2007. 10.1007/s00186-007-0161-1.

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

[41] Harald Hruschka. Market share analysis using semi-parametric attraction models. *European Journal of Operational Research*, 138(1):212 – 225, 2002.

[42] Ilse C. F. Ipsen and Teresa M. Selee. Ergodicity coefficients defined by vector norms. *SIAM J. Matrix Analysis Applications*, 32(1):153–200, 2011.

[43] Charles R. Johnson and Ronald L. Smith. Inverse m-matrices, ii. *Linear Algebra and its Applications*, 435(5):953 – 983, 2011.

[44] P. E. Kloeden and A. M. Rubinov. A generalization of the perron-frobenius theorem. *Nonlinear Anal.*, 41(1-2):97–115, July 2000. ISSN 0362-546X.

[45] Elon Kohlberg and John W. Pratt. The contraction mapping approach to the perron-frobenius theory: Why hilbert's metric? *Mathematics of Operations Research*, 7(2):198–210, 1982.

[46] Ulrich Krause. Perron's stability theorem for non-linear mappings. *Journal of Mathematical Economics*, 15(3):275 – 282, 1986.

[47] J. Lasserre and J. Hiriart-Urruty. Mathematical properties of optimization problems defined by positively homogeneous functions. *Journal of Optimization Theory and Applications*, 112:31–52, 2002. 10.1023/A:1013088311288.

[48] B. Lemmens and R. Nussbaum. *Nonlinear Perron-Frobenius Theory*. Cambridge Tracts in Mathematics. Cambridge University Press, 2012. ISBN 9780521898812. URL `http://books.google.com/books?id=EYud2hfi_c4C`.

[49] B. Li. Generalizations of diagonal dominance in matrix theory, 1997.

[50] B. Li and M. Tsatsomeros. Doubly diagonally dominant matrices. *Linear Algebra and its Applications*, (261):221–235, 1997.

[51] Hongmin Li and Woonghee Tim Huh. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management*, 13(4):549–563, Fall 2011.

[52] Q. Liu and G. van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing Service Operations Management*, 10(2):288–310, 2008.

[53] Bacel Maddah and Ebru K. Bish. Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics (NRL)*, 54(3):315–330, 2007. ISSN 1520-6750.

[54] Andreu Mas-Colell. Homeomorphisms of compact, convex sets and the jacobian matrix. *General equilibrium, Growth and Trade.*, 1979. J. Green and J. Scheinlsman (Eds.).

[55] Andreu Mas-Colell. Two propositions on the global univalence of systems of cost function. *SIAM Journal on Mathematical Analysis*, 10(6):1105–5, 1979.

[56] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, June 1995. ISBN 0195073401.

[57] D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142., 1974. Zarembka, P. (Ed.).

[58] D. Mcfadden and K. Train. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

[59] J. Meissner and A. Strauss. Choice-based network revenue management under weak market segmentation. Working paper (available at http://www.meiss.com), Lancaster University Management School, 2009.

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

[61] Shashi Kant Mishra and Giorgio Giorgi. *Invexity and optimization.* Nonconvex Optimization and Its Applications. Springer, Dordrecht, 2008.

[62] Jorge J. More. Nonlinear generalizations of matrix diagonal dominance with application to gauss-seidel iterations. *SIAM Journal on Numerical Analysis*, 9 (2):pp. 357–378, 1972.

[63] M. Nakanishi and L. G. Cooper. Simplified estimation procedures for mci models. *Marketing Science*, 1(3):314–322, 1982.

[64] Y. E. Nesterov and A. S. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Studies in Applied Mathematics*. SIAM, Philadelphia,PA, 1994.

[65] P. Rusmevichientong, Z.-J. Max Shen, and D. B. Shmoys. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. Submitted for publication.

[66] Siegfried Schaible. Parameter-free convex equivalent and dual programs of fractional programming problems. *Mathematical Methods of Operations Research*, 18:187–196, 1974.

[67] Siegfried Schaible. Fractional programming. i. duality. *Management Science*, 22 (8):pp. 858–867, 1976.

[68] Siegfried Schaible and Jianming Shi. Recent developments in fractional programming: Single ratio and max-min case, 1998.

[69] Cornelia Schön. On the product line selection problem under attraction choice models of consumer behavior. *European Journal of Operational Research*, 206 (1):260 – 264, 2010.

[70] E. Seneta. Coefficients of ergodicity: Structure and applications. *Advances in Applied Probability*, 11(3):pp. 576–590, 1979.

[71] D. F. Shanno and R. J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: orderings and higher-order methods. *Mathematical Programming*, 87(2):303–316, 2000.

[72] R. L. Smith. M-matrices whose inverses are stochastic. *SIAM Journal on Algebraic and Discrete Methods*, 2(3):259–265, 1981.

[73] J.-S. Song and Z. Xue. Demand management and inventory control for substitutable products. 2007. Working paper.

[74] K. T. Talluri and G. J. van Ryzin. An analysis of bid-price controls for network revenue management. *Management Science*, 44(11-Part-1):1577–1593, 1998.

[75] K. T. Talluri and G. J. van Ryzin. *The Theory and Practice of Revenue Management.* Springer, 2004.

[76] K. T. Talluri and G. J. van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Manage. Sci.*, 50(1):15–33, 2004.

[77] K. E. Train. *Discrete choice methods with simulation.* Cambridge University Press, 2003.

[78] G. L. Urban. A mathematical modeling approach to product line decisions. *Journal of Marketing Research*, 6(1):40–47, 1969.

[79] R. J. Vanderbei and D. F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13(1): 231–252, 1999.

[80] A. Vladimirov, V.I. Oseledets, A. Rybko, K.M. Khanin, and D.V. Khmelev. A Nonlinear Generalization of Perron's Theorem. *Doklady Mathematics*, 67(2): 221–225, 2003.

[81] D. Zhang and D. Adelman. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science*, 43 (3):381–394, 2009.